

# Probabilità e Processi Stocastici (455AA)

## Lezione 20

Dario Trevisan – <https://web.dm.unipi.it/trevisan>

1/12/2025

# Modelli ARIMA

# Richiami sui processi a stati continui

Dato un processo  $(X_t)_{t \in \mathcal{T}}$  a stati continui  $(\mathbb{R})$ , abbiamo introdotto

- ▶ La funzione di media  $t \mapsto \mathbb{E}[X_t]$ ,

# Richiami sui processi a stati continui

Dato un processo  $(X_t)_{t \in \mathcal{T}}$  a stati continui  $(\mathbb{R})$ , abbiamo introdotto

- ▶ La funzione di media  $t \mapsto \mathbb{E}[X_t]$ ,
- ▶ La funzione di autocovarianza  $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$ ,

# Richiami sui processi a stati continui

Dato un processo  $(X_t)_{t \in \mathcal{T}}$  a stati continui  $(\mathbb{R})$ , abbiamo introdotto

- ▶ La funzione di media  $t \mapsto \mathbb{E}[X_t]$ ,
- ▶ La funzione di autocovarianza  $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$ ,
- ▶ La funzione di autocorrelazione  $(s, t) \mapsto \text{ACF}(s, t) = \rho_{X_s, X_t}$

# Richiami sui processi a stati continui

Dato un processo  $(X_t)_{t \in \mathcal{T}}$  a stati continui  $(\mathbb{R})$ , abbiamo introdotto

- ▶ La funzione di media  $t \mapsto \mathbb{E}[X_t]$ ,
- ▶ La funzione di autocovarianza  $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$ ,
- ▶ La funzione di autocorrelazione  $(s, t) \mapsto \text{ACF}(s, t) = \rho_{X_s, X_t}$
- ▶ Il concetto di *stazionarietà in senso lato*: media costante e  $C(s, t) = C(0, |t - s|)$

# Richiami sui processi a stati continui

Dato un processo  $(X_t)_{t \in \mathcal{T}}$  a stati continui  $(\mathbb{R})$ , abbiamo introdotto

- ▶ La funzione di media  $t \mapsto \mathbb{E}[X_t]$ ,
- ▶ La funzione di autocovarianza  $(s, t) \mapsto C(s, t) = \text{Cov}(X_s, X_t)$ ,
- ▶ La funzione di autocorrelazione  $(s, t) \mapsto \text{ACF}(s, t) = \rho_{X_s, X_t}$
- ▶ Il concetto di *stazionarietà in senso lato*: media costante e  $C(s, t) = C(0, |t - s|)$
- ▶ La classe dei *processi gaussiani*

Introduciamo una famiglia generale di processi, detti ARIMA:  
**A**uto**R**egressive **I**ntegrated **M**oving **A**verage.

- ▶ Studiamo separatamente i tre “ingredienti” di un processo ARIMA:



Introduciamo una famiglia generale di processi, detti ARIMA:  
**A**uto**R**egressive **I**ntegrated **M**oving **A**verage.

- ▶ Studiamo separatamente i tre “ingredienti” di un processo ARIMA:
1. la componente autoregressiva (AR)

Introduciamo una famiglia generale di processi, detti ARIMA:  
**A**uto**R**egressive **I**ntegrated **M**oving **A**verage.

- ▶ Studiamo separatamente i tre “ingredienti” di un processo ARIMA:
  1. la componente autoregressiva (AR)
  2. quella a media mobile (MA)

Introduciamo una famiglia generale di processi, detti ARIMA:  
**A**uto**R**egressive **I**ntegrated **M**oving **A**verage.

- ▶ Studiamo separatamente i tre “ingredienti” di un processo ARIMA:
  1. la componente autoregressiva (AR)
  2. quella a media mobile (MA)
  3. il procedimento di integrazione (I) a tempi discreti.

# Introduzione

Introduciamo una famiglia generale di processi, detti ARIMA:  
**A**uto**R**egressive **I**ntegrated **M**oving **A**verage.

- ▶ Studiamo separatamente i tre “ingredienti” di un processo ARIMA:
  1. la componente autoregressiva (AR)
  2. quella a media mobile (MA)
  3. il procedimento di integrazione (I) a tempi discreti.
- ▶ Supponiamo che  $\mathcal{T} = \{0, 1, \dots, n\}$  oppure  $\mathcal{T} = \mathbb{N}$  o anche  $\mathcal{T} = \mathbb{Z}$ , e che  $(W_t)_{t \in \mathcal{T}}$  sia un rumore bianco gaussiano di intensità  $\sigma^2$ .

# Operatore di Lag

Introduciamo l'operatore  $L$  che trasforma

$$(X_t)_{t \in \mathcal{T}} \mapsto ((LX)_t)_{t \in \mathcal{T}} = (X_{t-1})_{t \in \mathcal{T}}$$

- Scriviamo semplicemente  $LX_t$  invece di  $(LX)_t$ .

# Operatore di Lag

Introduciamo l'operatore  $L$  che trasforma

$$(X_t)_{t \in \mathcal{T}} \mapsto ((LX)_t)_{t \in \mathcal{T}} = (X_{t-1})_{t \in \mathcal{T}}$$

- ▶ Scriviamo semplicemente  $LX_t$  invece di  $(LX)_t$ .
- ▶  $L$  è lineare:

$$L(X + Y)_t = X_{t-1} + Y_{t-1} = LX_t + LY_t, \quad L(cX)_t = cLX_t.$$

# Operatore di Lag

Introduciamo l'operatore  $L$  che trasforma

$$(X_t)_{t \in \mathcal{T}} \mapsto ((LX)_t)_{t \in \mathcal{T}} = (X_{t-1})_{t \in \mathcal{T}}$$

► Scriviamo semplicemente  $LX_t$  invece di  $(LX)_t$ .

►  $L$  è lineare:

$$L(X + Y)_t = X_{t-1} + Y_{t-1} = LX_t + LY_t, \quad L(cX)_t = cLX_t.$$

► Componendo  $L$  con se stesso si ottengono ritardi di ordine superiore:  $L^2X_t = LLX_t = X_{t-2}$ ,  $L^3X_t = X_{t-3}$ , ecc.

# Operatore di Lag

Introduciamo l'operatore  $L$  che trasforma

$$(X_t)_{t \in \mathcal{T}} \mapsto ((LX)_t)_{t \in \mathcal{T}} = (X_{t-1})_{t \in \mathcal{T}}$$

- ▶ Scriviamo semplicemente  $LX_t$  invece di  $(LX)_t$ .
- ▶  $L$  è lineare:

$$L(X + Y)_t = X_{t-1} + Y_{t-1} = LX_t + LY_t, \quad L(cX)_t = cLX_t.$$

- ▶ Componendo  $L$  con se stesso si ottengono ritardi di ordine superiore:  $L^2X_t = LLX_t = X_{t-2}$ ,  $L^3X_t = X_{t-3}$ , ecc.
- ▶ Espressioni del tipo

$$a_0X_t + a_1X_{t-1} + \dots + a_kX_{t-k} = a_0X_t + a_1LX_t + a_2L^2X_t + \dots + a_kL^kX_t$$

si possono abbreviare come

$$p(L)X_t = (a_0 + a_1L + a_2L^2 + \dots + a_kL^k)X_t.$$



I modelli autoregressivi generalizzano l'equazione lineare con smorzamento.

- ▶ L'equazione

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

è una regressione lineare semplice di  $X_t$  rispetto a  $X_{t-1}$ .

I modelli autoregressivi generalizzano l'equazione lineare con smorzamento.

- ▶ L'equazione

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

è una regressione lineare semplice di  $X_t$  rispetto a  $X_{t-1}$ .

- ▶ Si estende ad una regressione lineare multipla su  $p \geq 1$  istanti precedenti.

I modelli autoregressivi generalizzano l'equazione lineare con smorzamento.

- ▶ L'equazione

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

è una regressione lineare semplice di  $X_t$  rispetto a  $X_{t-1}$ .

- ▶ Si estende ad una regressione lineare multipla su  $p \geq 1$  istanti precedenti.
- ▶ Dato  $p \geq 0$ , un processo  $(X_t)_{t \in \mathcal{T}}$  è detto  $\text{AR}(p)$  (autoregressivo di ordine  $p$ ) se esistono parametri  $\alpha_1, \alpha_2, \dots, \alpha_p \in \mathbb{R}$  tali che, per ogni  $t \in \mathcal{T}$  (tale che  $t - p \in \mathcal{T}$ ) si abbia

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + W_t.$$

# Notazione compatta

Possiamo scrivere la ricorsione

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + W_t.$$

in modo compatto:

$$p(L)X_t = W_t,$$

dove  $p(L)$  è il polinomio formale nella variabile  $L$  dato da

$$p(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p = 1 - \sum_{i=1}^p \alpha_i L^i.$$

La media mobile su una finestra temporale sinistra di ampiezza  $q \geq 1$ , trasforma un processo  $(Z_t)_{t \in \mathcal{T}}$  con le medie

$$\bar{Z}_t = \frac{1}{q} \sum_{i=0}^{q-1} Z_{t-i}.$$

- È caso particolare di *convoluzione*  $Z * g$  tra il processo e il filtro

$$g(t) = \begin{cases} \frac{1}{q} & \text{se } i = 0, 1, \dots, (q-1) \\ 0 & \text{altrimenti.} \end{cases}$$

# Convoluzione e stazionarietà

Data una qualsiasi sia  $g$  (nota e fissata), se il processo  $Z$  è stazionario (in senso lato o anche in senso stretto), anche  $Z * g$  lo è (nello stesso senso).

► Infatti, la funzione di media è

$$\mathbb{E}[(Z * g)_t] = \mathbb{E}\left[\sum_i Z_{t-i}g(i)\right] = \sum_i \mathbb{E}[Z_{t-i}]g(i) = m \sum_i g(i),$$

avendo indicato con  $m = \mathbb{E}[Z_s]$ .

# Convoluzione e stazionarietà

Data una qualsiasi sia  $g$  (nota e fissata), se il processo  $Z$  è stazionario (in senso lato o anche in senso stretto), anche  $Z * g$  lo è (nello stesso senso).

- Infatti, la funzione di media è

$$\mathbb{E}[(Z * g)_t] = \mathbb{E}\left[\sum_i Z_{t-i}g(i)\right] = \sum_i \mathbb{E}[Z_{t-i}]g(i) = m \sum_i g(i),$$

avendo indicato con  $m = \mathbb{E}[Z_s]$ .

- La funzione di autocovarianza è, usando la bilinearità,

$$\begin{aligned} C(s, t) &= \text{Cov}((Z * g)_s, (Z * g)_t) = \sum_i \sum_j g(i)g(j) \text{Cov}(Z_{s-i}, Z_{t-j}) \\ &= \sum_i \sum_j g(i)g(j) C((t-s) + (i-j)) \end{aligned}$$

che dipende da  $t - s$  solamente.

# Modelli MA

Dato  $q \geq 0$ , un processo  $(X_t)_{t \in \mathcal{T}}$  è detto  $\text{MA}(q)$  (a media mobile di ordine  $q$ ) se esistono parametri  $\beta_1, \beta_2, \dots, \beta_q \in \mathbb{R}$  tali che, per ogni  $t \in \mathcal{T}$  (tale che  $t - q \in \mathcal{T}$ ) si abbia

$$X_t = W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2} + \dots + \beta_q W_{t-q}.$$

- Un processo a media mobile  $\text{MA}(q)$  è semplicemente del tipo  $W * g$ , dove  $g$  è dato dai coefficienti  $1, \beta_1, \beta_2, \dots, \beta_q$  (e nullo altrove). In particolare,  $X$  è **gaussiano e stazionario**.



Dato  $q \geq 0$ , un processo  $(X_t)_{t \in \mathcal{T}}$  è detto  $\text{MA}(q)$  (a media mobile di ordine  $q$ ) se esistono parametri  $\beta_1, \beta_2, \dots, \beta_q \in \mathbb{R}$  tali che, per ogni  $t \in \mathcal{T}$  (tale che  $t - q \in \mathcal{T}$ ) si abbia

$$X_t = W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2} + \dots + \beta_q W_{t-q}.$$

- ▶ Un processo a media mobile  $\text{MA}(q)$  è semplicemente del tipo  $W * g$ , dove  $g$  è dato dai coefficienti  $1, \beta_1, \beta_2, \dots, \beta_q$  (e nullo altrove). In particolare,  $X$  è **gaussiano e stazionario**.
- ▶ Con il polinomio dell'operatore ritardo riscriviamo

$$X_t = q(L)W_t,$$

dove

$$q(L) = 1 + \beta_1 L + \dots + \beta_q L^q = 1 + \sum_{j=1}^q \beta_q L^j.$$

# Integrazione discreta

Consideriamo l'operazione di derivazione discreta: la derivata di un processo  $(X_t)_{t \in \mathcal{T}}$  a tempi discreti è

$$X_t - X_{t-1} = (1 - L)X_t,$$

per  $t \geq 1$ .

► Iterando, si trova

$$(1 - L)^d X_t = \sum_{i=0}^d \binom{d}{i} (-1)^i L^i X_t.$$

# Integrazione discreta

Consideriamo l'operazione di derivazione discreta: la derivata di un processo  $(X_t)_{t \in \mathcal{T}}$  a tempi discreti è

$$X_t - X_{t-1} = (1 - L)X_t,$$

per  $t \geq 1$ .

- ▶ Iterando, si trova

$$(1 - L)^d X_t = \sum_{i=0}^d \binom{d}{i} (-1)^i L^i X_t.$$

- ▶ La formula è una convoluzione  $X * g$ : se  $X$  è stazionario, lo è anche ogni derivata discreta di qualsiasi ordine  $d$ .

L'integrazione discreta è l'inversa della derivata discreta:  $X$  è l'integrale discreto di  $Y$  se vale  $(1 - L)X = Y$ , e similmente per gli integrali iterati.

- ▶ la passeggiata aleatoria gaussiana,  $S_t = S_{t-1} + W_t$ , ossia  $(1 - L)S_t = W_t$  è l'integrale di  $W_t$ .

L'integrazione discreta è l'inversa della derivata discreta:  $X$  è l'integrale discreto di  $Y$  se vale  $(1 - L)X = Y$ , e similmente per gli integrali iterati.

- ▶ la passeggiata aleatoria gaussiana,  $S_t = S_{t-1} + W_t$ , ossia  $(1 - L)S_t = W_t$  è l'integrale di  $W_t$ .
- ▶ l'integrazione discreta non mantiene la stazionarietà di un processo.

# Definizione generale ARIMA

Dati  $p, d, q \geq 0$ , un processo  $(X_t)_{t \in \mathcal{T}}$  è detto  $\text{ARIMA}(p, d, q)$  se esistono parametri  $(\alpha_i)_{i=1}^p, (\beta_j)_{j=1}^q$  reali tali che, per ogni  $t \in \mathcal{T}$  (tale che  $t - d - p$  e  $t - q \in \mathcal{T}$ ), posto

$$Y_t = (1 - L)^d X_t$$

valga

$$Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + W_t + \sum_{j=1}^q \beta_j W_{t-j}.$$

► Usando i polinomi

$$p(L) = 1 - \sum_{i=1}^p \alpha_i L^i, \quad \text{e} \quad q(L) = 1 + \sum_{j=1}^q \beta_j L^j$$

si può scrivere in forma compatta:

$$p(L)(1 - L)^d X_t = q(L)W_t.$$

- ▶ il rumore bianco gaussiano è  $\text{ARIMA}(0, 0, 0)$ ,

- ▶ il rumore bianco gaussiano è  $\text{ARIMA}(0, 0, 0)$ ,
- ▶ la passeggiata aleatoria è  $\text{ARIMA}(0, 1, 0)$



- ▶ il rumore bianco gaussiano è  $\text{ARIMA}(0, 0, 0)$ ,
- ▶ la passeggiata aleatoria è  $\text{ARIMA}(0, 1, 0)$
- ▶ l'equazione lineare con smorzamento è  $\text{ARIMA}(1, 0, 0)$ .

# Modelli ARIMA: proprietà

Discutiamo tre proprietà fondamentali dei modelli ARIMA:

1. condizioni sulla stazionarietà,

# Modelli ARIMA: proprietà

Discutiamo tre proprietà fondamentali dei modelli ARIMA:

1. condizioni sulla stazionarietà,
2. una equazione ricorsiva per la funzione di autocovarianza (nel caso stazionario)

# Modelli ARIMA: proprietà

Discutiamo tre proprietà fondamentali dei modelli ARIMA:

1. condizioni sulla stazionarietà,
2. una equazione ricorsiva per la funzione di autocovarianza (nel caso stazionario)
3. come stimare i parametri sulla base delle osservazioni.

# Modelli ARIMA: proprietà

Discutiamo tre proprietà fondamentali dei modelli ARIMA:

1. condizioni sulla stazionarietà,
  2. una equazione ricorsiva per la funzione di autocovarianza (nel caso stazionario)
  3. come stimare i parametri sulla base delle osservazioni.
- Consideriamo un processo  $(X_t)_{t \in \mathcal{T}}$  ARIMA( $p, d, q$ ) con parametri  $(\alpha_i)_{i=1}^p$  e  $(\beta_j)_{j=1}^q$ .

# Stazionarietà

Nel caso ARIMA(1, 0, 0), la stazionarietà può avvenire se e solo se  $|\alpha_1| < 1$ .

- ▶ I processi a media mobile ARIMA(0, 0,  $q$ ) possono sempre essere stazionari:

$$X_t = q(L)W_t = W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2} + \dots + \beta_q W_{t-q}$$

# Stazionarietà

Nel caso ARIMA(1, 0, 0), la stazionarietà può avvenire se e solo se  $|\alpha_1| < 1$ .

- ▶ I processi a media mobile ARIMA(0, 0,  $q$ ) possono sempre essere stazionari:

$$X_t = q(L)W_t = W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2} + \dots + \beta_q W_{t-q}$$

- ▶ Nel caso generale, l'idea è “risolvere” l'equazione del modello

$$p(L)(1-L)^d X_t = q(L)W_t, \quad \text{da cui} \quad X_t = \frac{q(L)}{p(L)(1-L)^d} W_t,$$

# Stazionarietà

Nel caso ARIMA(1, 0, 0), la stazionarietà può avvenire se e solo se  $|\alpha_1| < 1$ .

- ▶ I processi a media mobile ARIMA(0, 0,  $q$ ) possono sempre essere stazionari:

$$X_t = q(L)W_t = W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2} + \dots + \beta_q W_{t-q}$$

- ▶ Nel caso generale, l'idea è “risolvere” l'equazione del modello

$$p(L)(1-L)^d X_t = q(L)W_t, \quad \text{da cui} \quad X_t = \frac{q(L)}{p(L)(1-L)^d} W_t,$$

- ▶ Per dare senso alla scrittura, sviluppiamo in serie

$$\frac{q(z)}{p(z)(1-z)^d} = \sum_{k=0}^{\infty} b_k z^k, \quad \text{così} \quad X_t = \sum_{k=0}^{\infty} b_k L^k W_t.$$



## Criterio di stazionarietà

Il problema sta nella convergenza della serie, che dipende dalla crescita dei coefficienti  $b_k$  al tendere di  $k \rightarrow \infty$  e in ultima analisi agli zeri (complessi) del denominatore  $p(z)(1 - z)^d$ .

► Dati  $(p, d, q)$  e coefficienti  $(\alpha_i)_{i=1}^p, (\beta_j)_{j=1}^q$ , posto

$$p(z) = 1 - \sum_{i=1}^p \alpha_i z^i,$$

allora esiste un modello ARIMA( $p, d, q$ ) stazionario con tali coefficienti se e solo se  $d = 0$  e tutte le radici complesse di  $p(z)$  hanno modulo  $|z| > 1$ , ossia

se  $z \in \mathbb{C}$  è tale che  $p(z) = 0$ , allora  $|z| > 1$ .

# Criterio di stazionarietà

Il problema sta nella convergenza della serie, che dipende dalla crescita dei coefficienti  $b_k$  al tendere di  $k \rightarrow \infty$  e in ultima analisi agli zeri (complessi) del denominatore  $p(z)(1 - z)^d$ .

- Dati  $(p, d, q)$  e coefficienti  $(\alpha_i)_{i=1}^p, (\beta_j)_{j=1}^q$ , posto

$$p(z) = 1 - \sum_{i=1}^p \alpha_i z^i,$$

allora esiste un modello ARIMA( $p, d, q$ ) stazionario con tali coefficienti se e solo se  $d = 0$  e tutte le radici complesse di  $p(z)$  hanno modulo  $|z| > 1$ , ossia

se  $z \in \mathbb{C}$  è tale che  $p(z) = 0$ , allora  $|z| > 1$ .

- Nel caso dell'equazione lineare con smorzamento si trova  $|\alpha| < 1$ .

# Autocovarianza

I processi ARIMA hanno media nulla e l'equazione permette anche di ottenere una formula ricorsiva (le *equazioni di Yule-Walker*) per la funzione di autocovarianza:

$$C(0, t) = \sum_{i=1}^p \alpha_i C(0, t - i), \quad \text{se } t > q.$$

► Nel caso ARIMA(1, 0, 0) si ha

$$C(0, t) = \alpha_1 C(0, t - 1) = \dots = \alpha_1^t C(0, 0).$$

# Autocovarianza

I processi ARIMA hanno media nulla e l'equazione permette anche di ottenere una formula ricorsiva (le *equazioni di Yule-Walker*) per la funzione di autocovarianza:

$$C(0, t) = \sum_{i=1}^p \alpha_i C(0, t - i), \quad \text{se } t > q.$$

- ▶ Nel caso ARIMA(1, 0, 0) si ha

$$C(0, t) = \alpha_1 C(0, t - 1) = \dots = \alpha_1^t C(0, 0).$$

- ▶ Nel caso ARIMA(0, 0, q) si ha

$$C(0, t) = 0 \quad \text{se } t > q.$$



# Stima dei parametri

A partire dall'osservazione di una *serie storica*  $(x_t)_{t=0}^n$ , come stimare i parametri di un processo ARIMA che la modella?

- ▶ Abbiamo visto la stima di massima verosimiglianza negli esempi del rumore bianco gaussiano, della passeggiata aleatoria e dell'equazione lineare con smorzamento.

# Stima dei parametri

A partire dall'osservazione di una *serie storica*  $(x_t)_{t=0}^n$ , come stimare i parametri di un processo ARIMA che la modella?

- ▶ Abbiamo visto la stima di massima verosimiglianza negli esempi del rumore bianco gaussiano, della passeggiata aleatoria e dell'equazione lineare con smorzamento.
- ▶ Il metodo si riduce alla minimizzazione dei residui quadratici.

# Stima dei parametri

A partire dall'osservazione di una *serie storica*  $(x_t)_{t=0}^n$ , come stimare i parametri di un processo ARIMA che la modella?

- ▶ Abbiamo visto la stima di massima verosimiglianza negli esempi del rumore bianco gaussiano, della passeggiata aleatoria e dell'equazione lineare con smorzamento.
- ▶ Il metodo si riduce alla minimizzazione dei residui quadratici.
- ▶ Si può fare lo stesso per un ARIMA generale, ma la nozione di residuo va chiarita.



Partendo dall'equazione definente

$$p(L)(1-L)^d X_t = q(L)W_t$$

si ricava una equazione ricorsiva per il rumore bianco gaussiano  $W_t$ :

$$W_t = - \sum_{j=1}^q \beta_j W_{t-j} + p(L)(1-L)^d X_t$$

► Avendo osservato  $X_t = x_t$ , definiamo ricorsivamente i residui

$$w_t = - \sum_{j=1}^q \beta_j w_{t-j} + p(L)(1-L)^d x_t,$$

Partendo dall'equazione definente

$$p(L)(1 - L)^d X_t = q(L)W_t$$

si ricava una equazione ricorsiva per il rumore bianco gaussiano  $W_t$ :

$$W_t = - \sum_{j=1}^q \beta_j W_{t-j} + p(L)(1 - L)^d X_t$$

- ▶ Avendo osservato  $X_t = x_t$ , definiamo ricorsivamente i residui

$$w_t = - \sum_{j=1}^q \beta_j w_{t-j} + p(L)(1 - L)^d x_t,$$

- ▶ La stima di massima verosimiglianza si ottiene minimizzando la somma dei quadrati:

$$(\alpha_{\text{MLE}}, \beta_{\text{MLE}}) \in \arg \min_{\alpha, \beta} \sum_t w_t^2$$

# Stima dei coefficienti in R

- ▶ In R vi sono diverse funzioni che permettono la stima (fit) di un ARIMA sui dati osservati, ad esempio `arima()` oppure `Arima()` dalla libreria `forecast`.

# Stima dei coefficienti in R

- ▶ In R vi sono diverse funzioni che permettono la stima (fit) di un ARIMA sui dati osservati, ad esempio `arima()` oppure `Arima()` dalla libreria `forecast`.
- ▶ Oltre alle stime puntuali forniscono stime delle deviazioni standard associate ai parametri.

# Stima dei coefficienti in R

- ▶ In R vi sono diverse funzioni che permettono la stima (fit) di un ARIMA sui dati osservati, ad esempio `arima()` oppure `Arima()` dalla libreria `forecast`.
- ▶ Oltre alle stime puntuali forniscono stime delle deviazioni standard associate ai parametri.
- ▶ Inoltre con la funzione `forecast()` si ottengono previsioni (basate sulla stima ottenuta) dei valori futuri.

# Scelta del modello

Uno dei problemi principali oltre alla stima dei parametri è la scelta del modello (ossia i valori di  $(p, d, q)$ ).

- ▶ come per la regressione, un maggior numero di parametri permette di aderire meglio alle osservazioni, ma non garantisce previsioni accurate.

# Scelta del modello

Uno dei problemi principali oltre alla stima dei parametri è la scelta del modello (ossia i valori di  $(p, d, q)$ ).

- ▶ come per la regressione, un maggior numero di parametri permette di aderire meglio alle osservazioni, ma non garantisce previsioni accurate.
- ▶ vi sono indicatori particolari (AIC, BIC) che “guidano” opportunamente nella scelta.

# Scelta del modello

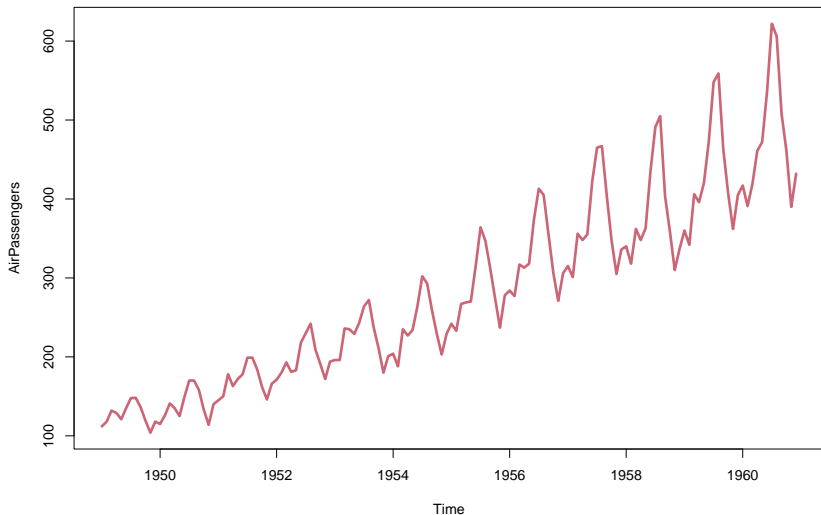
Uno dei problemi principali oltre alla stima dei parametri è la scelta del modello (ossia i valori di  $(p, d, q)$ ).

- ▶ come per la regressione, un maggior numero di parametri permette di aderire meglio alle osservazioni, ma non garantisce previsioni accurate.
- ▶ vi sono indicatori particolari (AIC, BIC) che “guidano” opportunamente nella scelta.
- ▶ In R il comando `auto.arima()` valuta automaticamente il migliore modello che aderisce alle osservazioni sulla base di tali indicatori.



# Un esempio

Consideriamo il dataset precaricato in R `AirPassengers`.



**Figure 1:** Grafico della serie storica `AirPassengers`

```

## Registered S3 method overwritten by 'quantmod':
##      method          from
##      as.zoo.data.frame zoo

## Series: AirPassengers
## ARIMA(2,1,1)(0,1,0)[12]
##
## Coefficients:
##              ar1      ar2      ma1
##          0.5960  0.2143 -0.9819
## s.e.    0.0888  0.0880  0.0292
##
## sigma^2 = 132.3:  log likelihood = -504.92
## AIC=1017.85   AICc=1018.17   BIC=1029.35
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE
## Training set 1.342299 10.84619 7.86754 0.4206976 2.80045

```

La funzione `auto.arima()` propone un modello ARIMA con stagionalità di periodo 12 (mesi) e ordine  $(2, 1, 1)(0, 1, 0)$  (la seconda tripla si riferisce alla stagionalità).

- ▶ In realtà funzione `auto.arima()` non determina automaticamente il periodo 12 e questo va indicato prima di applicarla ai dati, nel momento in cui si definisce un oggetto di tipo *serie storica* (in inglese *time series*) in R.

La funzione `auto.arima()` propone un modello ARIMA con stagionalità di periodo 12 (mesi) e ordine  $(2, 1, 1)(0, 1, 0)$  (la seconda tripla si riferisce alla stagionalità).

- ▶ In realtà funzione `auto.arima()` non determina automaticamente il periodo 12 e questo va indicato prima di applicarla ai dati, nel momento in cui si definisce un oggetto di tipo *serie storica* (in inglese *time series*) in R.
- ▶ Partendo da un vettore di dati osservati, il comando è `ts()`, che contiene l'opzione *frequency* (se non specificata è posta uguale ad 1).

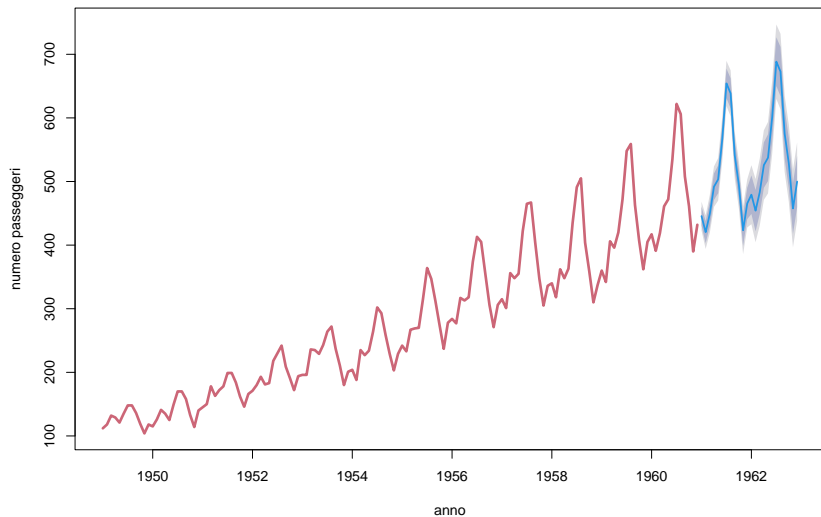
La funzione `auto.arima()` propone un modello ARIMA con stagionalità di periodo 12 (mesi) e ordine  $(2, 1, 1)(0, 1, 0)$  (la seconda tripla si riferisce alla stagionalità).

- ▶ In realtà funzione `auto.arima()` non determina automaticamente il periodo 12 e questo va indicato prima di applicarla ai dati, nel momento in cui si definisce un oggetto di tipo *serie storica* (in inglese *time series*) in R.
- ▶ Partendo da un vettore di dati osservati, il comando è `ts()`, che contiene l'opzione *frequency* (se non specificata è posta uguale ad 1).
- ▶ In generale si può ricorrere alla funzione di autocorrelazione empirica o ad analisi spettrale per determinare eventuali stagionalità e il loro periodo.

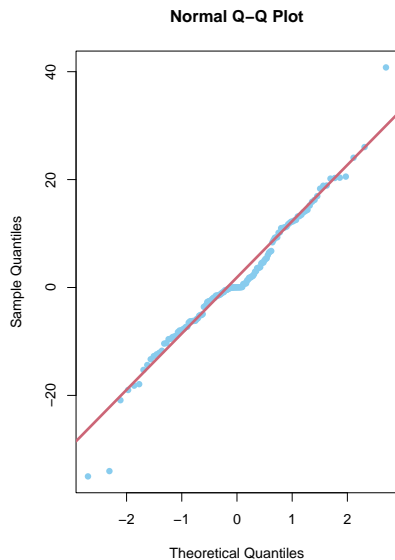
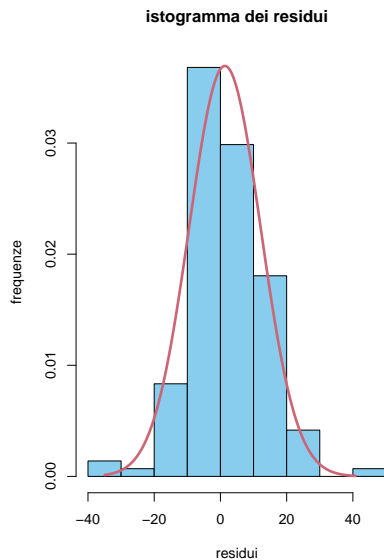
La funzione `auto.arima()` propone un modello ARIMA con stagionalità di periodo 12 (mesi) e ordine (2, 1, 1)(0, 1, 0) (la seconda tripla si riferisce alla stagionalità).

- ▶ In realtà funzione `auto.arima()` non determina automaticamente il periodo 12 e questo va indicato prima di applicarla ai dati, nel momento in cui si definisce un oggetto di tipo *serie storica* (in inglese *time series*) in R.
- ▶ Partendo da un vettore di dati osservati, il comando è `ts()`, che contiene l'opzione *frequency* (se non specificata è posta uguale ad 1).
- ▶ In generale si può ricorrere alla funzione di autocorrelazione empirica o ad analisi spettrale per determinare eventuali stagionalità e il loro periodo.
- ▶ Un comando automatico è invece `findfrequency()`.

Forecasts from ARIMA(2,1,1)(0,1,0)[12]



**Figure 2:** dati osservati e previsione con un modello SARIMA



**Figure 3:** istogramma e QQ-plot dei residui



## **Stima della funzione di autocovarianza**

# Stima della funzione di autocovarianza

Affrontiamo il problema generale di stimare la funzione di autocovarianza di un processo  $X$  a partire dall'osservazione dei valori  $(X_t)_{t=0}^{n-1} = (x_t)_{t=0}^{n-1}$  (una *serie storica*).

- ▶ È una generalizzazione del problema di stimare valor medio e varianza di una famiglia di variabili aleatorie indipendenti, tutte con la stessa legge.

# Stima della funzione di autocovarianza

Affrontiamo il problema generale di stimare la funzione di autocovarianza di un processo  $X$  a partire dall'osservazione dei valori  $(X_t)_{t=0}^{n-1} = (x_t)_{t=0}^{n-1}$  (una *serie storica*).

- ▶ È una generalizzazione del problema di stimare valor medio e varianza di una famiglia di variabili aleatorie indipendenti, tutte con la stessa legge.
- ▶ Nel caso gaussiano avevamo ottenuto la media e la covarianza campionarie.

Invece dell'indipendenza, supporremo la *stazionarietà del processo*, ossia la matrice di covarianza delle variabili  $(X_t)_{t=0}^{n-1}$  è costante sulle diagonali:

$$(C(s, t))_{s, t=0}^{n-1} = (C(|t - s|))_{s, t=0}^{n-1}.$$

- Supponiamo pure che il processo  $X$  sia *gaussiano e centrato* (ossia la funzione di media è nota e costantemente nulla)

Invece dell'indipendenza, supporremo la *stazionarietà del processo*, ossia la matrice di covarianza delle variabili  $(X_t)_{t=0}^{n-1}$  è costante sulle diagonali:

$$(C(s, t))_{s, t=0}^{n-1} = (C(|t - s|))_{s, t=0}^{n-1}.$$

- ▶ Supponiamo pure che il processo  $X$  sia *gaussiano e centrato* (ossia la funzione di media è nota e costantemente nulla)
- ▶ La verosimiglianza per la matrice di covarianza è

$$\begin{aligned} L(C; x) &= p((X_t)_{t=0}^{n-1} = (x_t)_{t=0}^{n-1} | C) \\ &\propto \exp\left(-\frac{1}{2}x^T C^{-1}x\right) \frac{1}{\sqrt{\det C}}, \end{aligned}$$

dove abbiamo posto, per alleggerire la notazione,  $x = (x_t)_{t=0}^{n-1}$ .

# Stima di massima verosimiglianza

Possiamo determinare la stima di massima verosimiglianza per  $C$  con i soliti passaggi: si tratta di minimizzare la funzione

$$C \mapsto x^T C^{-1} x + \log \det C.$$

- È difficile calcolare analiticamente  $C_{\text{MLE}}$ , ma si può ricorrere a metodi numerici.

# Una ipotesi semplificativa

Per ottenere delle espressioni elementari per  $C_{MLE}$  introduciamo una ulteriore ipotesi nella struttura della matrice di covarianza.

- ▶ Oltre al fatto che  $C(s, t)$  sia simmetrica, semidefinita positiva e costante sulle diagonali (per la stazionarietà), chiediamo che sia *circolante*:

$$C(k) = C(n - k) \quad \text{per ogni } k = 1, 2, \dots, n - 1.$$

# Una ipotesi semplificativa

Per ottenere delle espressioni elementari per  $C_{MLE}$  introduciamo una ulteriore ipotesi nella struttura della matrice di covarianza.

- ▶ Oltre al fatto che  $C(s, t)$  sia simmetrica, semidefinita positiva e costante sulle diagonali (per la stazionarietà), chiediamo che sia *circolante*:

$$C(k) = C(n - k) \quad \text{per ogni } k = 1, 2, \dots, n - 1.$$

- ▶ È solo una ipotesi di comodo, una volta ottenute delle stime esplicite cerchiamo di rimuoverla.



# Passaggio alla base delle frequenze

Tramite trasformata di Fourier a tempi finiti,  $C$  diventa diagonale.  
Richiamiamo la notazione:

► Sia  $F \in \mathbb{C}^{n \times n}$ ,

$$F_{\xi t} = e^{-2\pi i \xi t / n}, \quad \text{per } \xi, t = 0, 1, \dots, (n-1).$$

# Passaggio alla base delle frequenze

Tramite trasformata di Fourier a tempi finiti,  $C$  diventa diagonale.  
Richiamiamo la notazione:

► Sia  $F \in \mathbb{C}^{n \times n}$ ,

$$F_{\xi t} = e^{-2\pi i \xi t / n}, \quad \text{per } \xi, t = 0, 1, \dots, (n-1).$$

► La trasformata di Fourier a tempi finiti di  $(x_t)_{t=0}^{n-1}$  è

$$\hat{x}(\xi) = \sum_{t=0}^{n-1} x_t e^{-2\pi i \xi t / n} = \sum_{t=0}^{n-1} F_{\xi t} x_t,$$

ossia  $\hat{x} = Fx$ .

Se il vettore  $x$  è l'osservazione del processo  $X$ , la matrice di covarianza del vettore aleatorio  $\hat{X} = FX$  si calcola tramite la formula per le trasformazioni affini:

$$\Sigma_{FX} = F\Sigma_X\bar{F}^T = FC\bar{F}^T,$$

- Questo cambio di coordinate: dalla base dei “tempi” a quella delle “frequenze” diagonalizza la matrice delle covarianze.

Se il vettore  $x$  è l'osservazione del processo  $X$ , la matrice di covarianza del vettore aleatorio  $\hat{X} = FX$  si calcola tramite la formula per le trasformazioni affini:

$$\Sigma_{FX} = F\Sigma_X\bar{F}^T = FC\bar{F}^T,$$

- ▶ Questo cambio di coordinate: dalla base dei “tempi” a quella delle “frequenze” diagonalizza la matrice delle covarianze.
- ▶ Introduciamo la notazione

$$\hat{C}(\xi) = \sum_{k=0}^{n-1} e^{-2\pi i \xi k/n} C(k).$$

## Stima di massima verosimiglianza

I numeri  $n\hat{C}(\xi)$  sono quindi multipli degli autovalori della matrice di covarianza  $C$  e quindi sono tutti positivi.

- In queste nuove coordinate, le componenti del vettore  $FX$  sono non correlate e quindi, essendo gaussiane, indipendenti.

# Stima di massima verosimiglianza

I numeri  $n\hat{C}(\xi)$  sono quindi multipli degli autovalori della matrice di covarianza  $C$  e quindi sono tutti positivi.

- ▶ In queste nuove coordinate, le componenti del vettore  $FX$  sono non correlate e quindi, essendo gaussiane, indipendenti.
- ▶ La verosimiglianza assume quindi una espressione molto più trattabile:

$$L(C; \hat{x}) = p(FX = \hat{x} | C) \propto \exp \left( -\frac{1}{2n} \sum_{\xi=0}^{n-1} \frac{|\hat{x}(\xi)|^2}{\hat{C}(\xi)} \right) \frac{1}{\sqrt{\prod_{\xi=0}^{n-1} \hat{C}(\xi)}}$$

# Stima di massima verosimiglianza

I numeri  $n\hat{C}(\xi)$  sono quindi multipli degli autovalori della matrice di covarianza  $C$  e quindi sono tutti positivi.

- ▶ In queste nuove coordinate, le componenti del vettore  $FX$  sono non correlate e quindi, essendo gaussiane, indipendenti.
- ▶ La verosimiglianza assume quindi una espressione molto più trattabile:

$$L(C; \hat{x}) = p(FX = \hat{x} | C) \propto \exp \left( -\frac{1}{2n} \sum_{\xi=0}^{n-1} \frac{|\hat{x}(\xi)|^2}{\hat{C}(\xi)} \right) \frac{1}{\sqrt{\prod_{\xi=0}^{n-1} \hat{C}(\xi)}}$$

- ▶ La stima di massima verosimiglianza si ottiene minimizzando la funzione

$$C \mapsto \sum_{\xi=1}^{n-1} \left[ \frac{1}{n} \frac{|\hat{x}(\xi)|^2}{\hat{C}(\xi)} + \log \hat{C}(\xi) \right]$$





Discutiamo l'espressione

$$\frac{1}{n^2} \sum_{s,t=0}^{n-1} x_t x_s \sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n}$$

► Se  $t = s + k$ , allora il termine

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} 1 = n,$$

Discutiamo l'espressione

$$\frac{1}{n^2} \sum_{s,t=0}^{n-1} x_t x_s \sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n}$$

► Se  $t = s + k$ , allora il termine

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} 1 = n,$$

► Se  $t = s + k - n$  ugualmente si avrebbe

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} e^{2\pi i\xi} = \sum_{\xi=0}^{n-1} 1 = n.$$

Discutiamo l'espressione

$$\frac{1}{n^2} \sum_{s,t=0}^{n-1} x_t x_s \sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n}$$

- ▶ Se  $t = s + k$ , allora il termine

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} 1 = n,$$

- ▶ Se  $t = s + k - n$  ugualmente si avrebbe

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = \sum_{\xi=0}^{n-1} e^{2\pi i\xi} = \sum_{\xi=0}^{n-1} 1 = n.$$

- ▶ In tutti gli altri casi possibili, si trova invece

$$\sum_{\xi=0}^{n-1} e^{-2\pi i(t-s-k)\xi/n} = 0,$$

Concludiamo che

$$C_{\text{MLE}}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} x_s x_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} x_s x_{s+k-n}.$$

► È la somma due contributi:

Concludiamo che

$$C_{\text{MLE}}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} x_s x_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} x_s x_{s+k-n}.$$

► È la somma due contributi:

1. la covarianza campionaria tra  $X_s$  e il processo “traslato” avanti nel tempo di  $k$  istanti,  $X_{s+k}$

Concludiamo che

$$C_{\text{MLE}}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} x_s x_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} x_s x_{s+k-n}.$$

► È la somma due contributi:

1. la covarianza campionaria tra  $X_s$  e il processo “traslato” avanti nel tempo di  $k$  istanti,  $X_{s+k}$
2. la covarianza campionaria tra  $X_s$  e il traslato indietro di  $n - k$  istanti,  $X_{s+k-n}$ .

Concludiamo che

$$C_{\text{MLE}}(k) = \frac{1}{n} \sum_{s=0}^{n-1-k} x_s x_{s+k} + \frac{1}{n} \sum_{s=n-k+1}^{n-1} x_s x_{s+k-n}.$$

- ▶ È la somma due contributi:
  1. la covarianza campionaria tra  $X_s$  e il processo “traslato” avanti nel tempo di  $k$  istanti,  $X_{s+k}$
  2. la covarianza campionaria tra  $X_s$  e il traslato indietro di  $n - k$  istanti,  $X_{s+k-n}$ .
- ▶ Se  $k \ll n$ , il primo è rilevante.

# Autocovarianza campionaria

I calcoli svolti suggeriscono di proporre come stima generale per  $C$  solo il primo contributo: la covarianza campionaria tra  $X_s$  e il traslato  $X_{s+k}$ .

- Definiamo la funzione di **autocovarianza campionaria** (o empirica) come

$$c(k) = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_s - \bar{x}_0)(x_{s+k} - \bar{x}_k),$$

dove le medie campionarie  $\bar{x}_0$ ,  $\bar{x}_k$  sono rispettivamente sui primi  $n-k$  e sugli ultimi  $n-k$  valori.



# Autocovarianza campionaria

I calcoli svolti suggeriscono di proporre come stima generale per  $C$  solo il primo contributo: la covarianza campionaria tra  $X_s$  e il traslato  $X_{s+k}$ .

- Definiamo la funzione di **autocovarianza campionaria** (o empirica) come

$$c(k) = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_s - \bar{x}_0)(x_{s+k} - \bar{x}_k),$$

dove le medie campionarie  $\bar{x}_0$ ,  $\bar{x}_k$  sono rispettivamente sui primi  $n-k$  e sugli ultimi  $n-k$  valori.

- $c(k)$  è la covarianza tra due variabili aleatorie così definite: si sceglie  $S \in \{0, 1, \dots, n-1-k\}$  casuale uniforme e si considerano i valori  $x_S$  (prima variabile) e  $x_{S+k}$  (seconda variabile).

# Funzione di autocorrelazione campionaria

Possiamo definire anche le varianze campionarie

$$\sigma_0^2 = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_s - \bar{x}_0)^2$$

e

$$\sigma_k^2 = \frac{1}{n-k} \sum_{s=0}^{n-1-k} (x_{s+k} - \bar{x}_k)^2$$

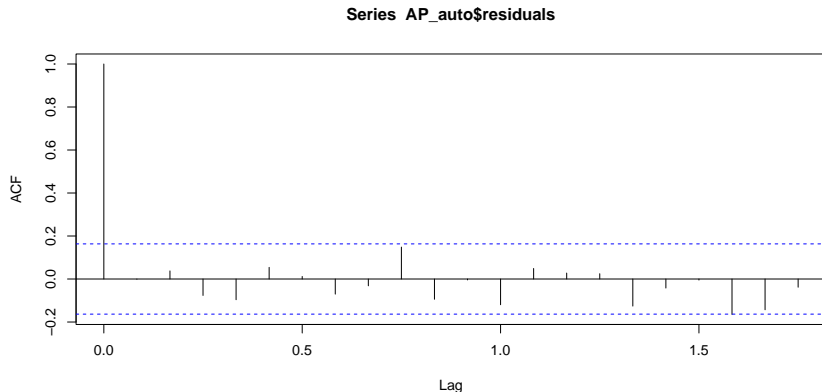
e quindi la funzione di **autocorrelazione campionaria**, data da

$$\text{acf}(k) = \frac{c(k)}{\sigma_0 \sigma_k},$$

che assume sempre valori tra  $[-1, 1]$

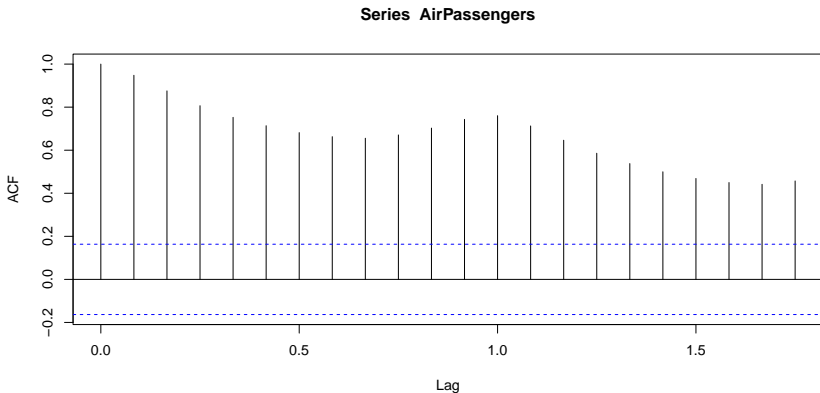
# ACF in R

In R è l'autocorrelazione campionaria si calcola tramite la funzione `acf()`.



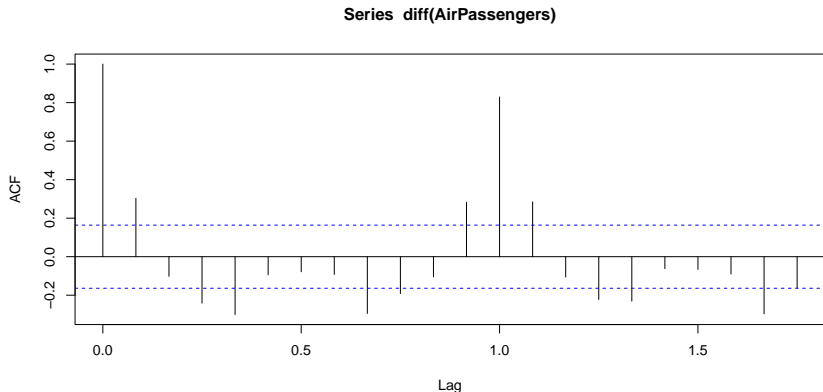
**Figure 4:** autocorrelazione empirica dei residui

Attenzione: la presenza di un “trend” può fare apparire la ACF uniformemente grande.



**Figure 5:** la funzione di autocorrelazione è uniformemente grande per via del trend

Per rimuovere questo effetto basta passare ad una derivata discreta tramite la funzione `diff()`.



**Figure 6:** la derivata discreta rimuove il trend e la funzione di autocorrelazione evidenzia la stagionalità a 12 mesi

# Densità spettrale di potenza

Torniamo alla base delle frequenze: cosa accade ai calcoli se non vale l'ipotesi semplificativa  $C(k) = C(n - k)$ ?

- ▶ Se  $n$  è molto grande la stima

$$\hat{C}(\xi) \approx \frac{|\hat{x}(\xi)|^2}{n} = \frac{1}{n} \left| \sum_{t=0}^{n-1} x_t e^{-2\pi i t \xi / n} \right|^2$$

è comunque una buona approssimazione.

# Densità spettrale di potenza

Torniamo alla base delle frequenze: cosa accade ai calcoli se non vale l'ipotesi semplificativa  $C(k) = C(n - k)$ ?

- ▶ Se  $n$  è molto grande la stima

$$\hat{C}(\xi) \approx \frac{|\hat{x}(\xi)|^2}{n} = \frac{1}{n} \left| \sum_{t=0}^{n-1} x_t e^{-2\pi i t \xi / n} \right|^2$$

è comunque una buona approssimazione.

- ▶ Il membro a destra è l'energia associata alla frequenza  $\xi$  diviso il tempo  $n$ , quindi si interpreta come *potenza*.

## Il Teorema di Wiener-Khinchin

Sia  $(X_t)_{t=0}^{\infty}$  un processo a valori reali, stazionario in senso lato, con media nulla  $\mathbb{E}[X_t] = 0$ , e tale che

$$\sum_{k=0}^{\infty} |C(0, k)| < \infty.$$

- Per ogni  $\xi \in [0, 1]$ , si definisce *densità spettrale di potenza* associata alla frequenza  $\xi$  la quantità

$$\hat{C}(\xi) = \sum_{k \in \mathbb{Z}} C(|k|) e^{-2\pi i k \xi}.$$



## Il Teorema di Wiener-Khinchin

Sia  $(X_t)_{t=0}^{\infty}$  un processo a valori reali, stazionario in senso lato, con media nulla  $\mathbb{E}[X_t] = 0$ , e tale che

$$\sum_{k=0}^{\infty} |C(0, k)| < \infty.$$

- Per ogni  $\xi \in [0, 1]$ , si definisce *densità spettrale di potenza* associata alla frequenza  $\xi$  la quantità

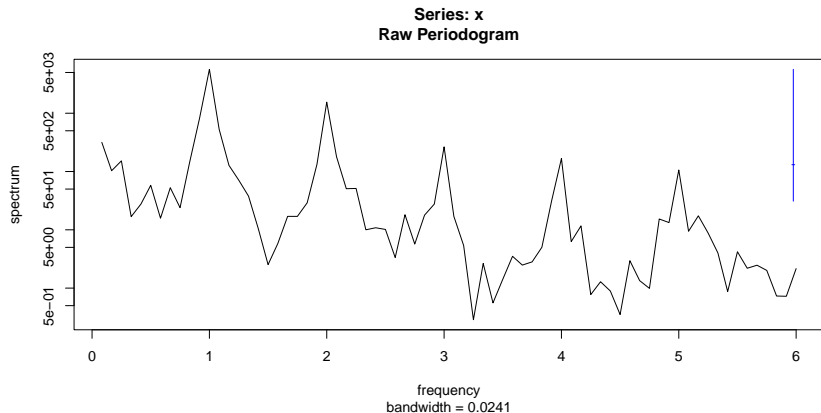
$$\hat{C}(\xi) = \sum_{k \in \mathbb{Z}} C(|k|) e^{-2\pi i k \xi}.$$

- Vale il limite

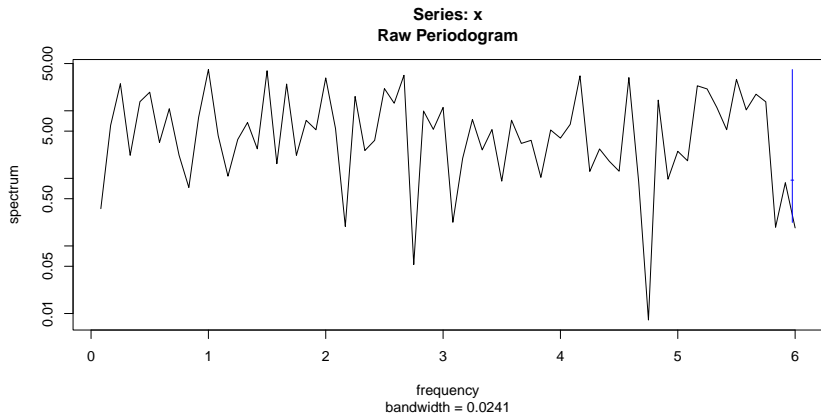
$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \left| \sum_{t=0}^{n-1} X_t e^{-2\pi i t \xi} \right|^2 \right] = \hat{C}(\xi).$$

# Spettrogramma in R

In R si può utilizzare direttamente il comando `spectrum()`, che fornisce una stima della densità spettrale di potenza da una serie storica osservata.



**Figure 7:** Stima della densità spettrale di potenza della serie AirPassengers



**Figure 8:** Stima della densità spettrale di potenza dei residui della serie AirPassengers dopo un fit con un modello ARIMA