Probabilità e Processi Stocastici (455AA)

Lezione 20

Dario Trevisan

4/12/2022

Section 1

Processi a stati continui

Richiami

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

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 $(s,t) \mapsto C(s,t) = \mathrm{Cov}(X_s, X_t)$,

Richiami

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 $(s,t) \mapsto C(s,t) = \mathrm{Cov}(X_s, X_t)$,
- La funzione di autocorrelazione $(s,t) \mapsto \mathrm{ACF}(s,t) = \rho_{X_s,X_t}$

Richiami

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

Richiami

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

Esempio 1: Rumore bianco gaussiano

Il più semplice processo a stati continui che consideriamo consiste di variabili aleatorie $(W_t)_{t\in\mathcal{T}}$, tutte con la medesima legge *gaussiana e indipendenti*, con media nulla e varianza σ^2: la densità della marginale è quindi

$$p(W_t = w) = \exp\left(-\frac{w^2}{2\sigma^2}\right)\frac{1}{\sqrt{2\pi\sigma^2}}.$$

- $(W_t)_{t\in\mathcal{T}}$ è detto **rumore bianco gaussiano** di **intensità** σ^2.

Il termine *rumore* è motivato da modelli di teoria dell'informazione.

- Un messaggio $(M_t)_{t \in \mathcal{T}}$ è trasmesso tramite un mezzo di comunicazione reale affetto da *distorsione*, perdite ecc.

Il termine *rumore* è motivato da modelli di teoria dell'informazione.

- Un messaggio $(M_t)_{t\in\mathcal{T}}$ è trasmesso tramite un mezzo di comunicazione reale affetto da *distorsione*, perdite ecc.
- Si suppone che il ricevitore osservi il processo

$$(M_t + W_t)_{t\in\mathcal{T}},$$

dove $(W_t)_{t\in\mathcal{T}}$ è un rumore bianco gaussiano di una intensità σ^2 (è un parametro che si dovrà stimare).

Il termine *rumore* è motivato da modelli di teoria dell'informazione.

- Un messaggio $(M_t)_{t\in\mathcal{T}}$ è trasmesso tramite un mezzo di comunicazione reale affetto da *distorsione*, perdite ecc.
- Si suppone che il ricevitore osservi il processo

$$(M_t + W_t)_{t\in\mathcal{T}},$$

  dove $(W_t)_{t\in\mathcal{T}}$ è un rumore bianco gaussiano di una intensità σ^2 (è un parametro che si dovrà stimare).
- Il rumore è sommato, per questo a volte è detto *additivo*.

Funzione di media e di autocovarianza

Sia $(W_t)_{t \in \mathcal{T}}$ un rumore bianco gaussiano di intensità σ^2.

- La funzione di media è identicamente nulla:

$$t \mapsto \mathbb{E}[W_t] = 0.$$

Funzione di media e di autocovarianza

Sia $(W_t)_{t \in \mathcal{T}}$ un rumore bianco gaussiano di intensità σ^2.

- La funzione di media è identicamente nulla:

$$t \mapsto \mathbb{E}[W_t] = 0.$$

- Ricordando che variabili indipendenti non sono correlate, e inoltre $\mathrm{Var}(W_t) = \sigma^2$ si ha

$$C(s,t) = \begin{cases} 0 & \text{se } s \neq t, \\ \sigma^2 & \text{se } s = t. \end{cases}$$

Funzione di media e di autocovarianza

Sia $(W_t)_{t\in\mathcal{T}}$ un rumore bianco gaussiano di intensità σ^2.

- La funzione di media è identicamente nulla:
$$t \mapsto \mathbb{E}[W_t] = 0.$$
- Ricordando che variabili indipendenti non sono correlate, e inoltre $\mathrm{Var}(W_t) = \sigma^2$ si ha
$$C(s,t) = \begin{cases} 0 & \text{se } s \neq t, \\ \sigma^2 & \text{se } s = t. \end{cases}$$
- Notazione $C(s,t) = \sigma^2 \delta_0(t-s)$ usando la delta di Dirac discreta:
$$\delta_0(x) = \begin{cases} 0 & \text{se } x \neq 0, \\ 1 & \text{se } x = 0. \end{cases}$$

Funzione di media e di autocovarianza

Sia $(W_t)_{t\in\mathcal{T}}$ un rumore bianco gaussiano di intensità σ^2.

- La funzione di media è identicamente nulla:

$$t \mapsto \mathbb{E}[W_t] = 0.$$

- Ricordando che variabili indipendenti non sono correlate, e inoltre $\mathrm{Var}(W_t) = \sigma^2$ si ha

$$C(s,t) = \begin{cases} 0 & \text{se } s \neq t, \\ \sigma^2 & \text{se } s = t. \end{cases}$$

- Notazione $C(s,t) = \sigma^2 \delta_0(t-s)$ usando la delta di Dirac discreta:

$$\delta_0(x) = \begin{cases} 0 & \text{se } x \neq 0, \\ 1 & \text{se } x = 0. \end{cases}$$

- Il processo è gaussiano e stazionario.

Stima dell'intensità σ^2 dalle osservazioni

Si può stimare σ^2 a partire da n osservazioni $W_{t_i} = w_i$.

- Il problema è lo stesso della stima della varianza di un campione gaussiano (di cui la media è nota e uguale a zero).

Stima dell'intensità σ^2 dalle osservazioni

Si può stimare σ^2 a partire da n osservazioni $W_{t_i} = w_i$.

- Il problema è lo stesso della stima della varianza di un campione gaussiano (di cui la media è nota e uguale a zero).
- In particolare la stima di massima verosimiglianza in questo caso è

$$\sigma^2_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} w_i^2.$$

Perché "bianco"?

- Supponiamo che l'insieme dei tempi sia finito $\mathcal{T} = \{0, 1, \ldots, n-1\}$.

Perché “bianco”?

- Supponiamo che l’insieme dei tempi sia finito $\mathcal{T} = \{0, 1, \ldots, n-1\}$.
- La trasformata di Fourier di $(W_t)_{t=0}^{n-1}$ è

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

Perché “bianco”?

- Supponiamo che l'insieme dei tempi sia finito $\mathcal{T} = \{0, 1, \ldots, n-1\}$.
- La trasformata di Fourier di $(W_t)_{t=0}^{n-1}$ è

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

- Per ciascuna frequenza $\xi \in \{0, \ldots, (n-1)\}$ la variabile aleatoria $\hat{W}(\xi)$ è una combinazione lineare (a coefficienti complessi) di variabili gaussiane indipendenti, quindi è gaussiana.

Perché “bianco”?

- Supponiamo che l’insieme dei tempi sia finito $\mathcal{T} = \{0, 1, \ldots, n-1\}$.
- La trasformata di Fourier di $(W_t)_{t=0}^{n-1}$ è

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

- Per ciascuna frequenza $\xi \in \{0, \ldots, (n-1)\}$ la variabile aleatoria $\hat{W}(\xi)$ è una combinazione lineare (a coefficienti complessi) di variabili gaussiane indipendenti, quindi è gaussiana.
- Il valor medio è per linearità

$$\mathbb{E}\left[\hat{W}(\xi)\right] = \sum_{t=0}^{n-1} \mathbb{E}\left[W_t\right] e^{-2\pi i \xi t/n} = 0,$$

- Il valor medio dell'energia $|\hat{W}(\xi)|^2$ sulla frequenza ξ è costante:

$$\mathbb{E}\left[|\hat{W}(\xi)|^2\right] = \sigma^2 n.$$

- Il valor medio dell'energia $|\hat{W}(\xi)|^2$ sulla frequenza ξ è costante:

$$\mathbb{E}\left[|\hat{W}(\xi)|^2\right] = \sigma^2 n.$$

- Dividendo per il tempo n si trova che la *potenza* media sulla frequenza ξ è costante e pari all'intensità σ^2.

Esempio 2: passeggiata aleatoria gaussiana

Posto $\mathcal{T} = \{0, 1, \ldots, n\}$ o eventualmente $\mathcal{T} = \mathbb{N}$, definiamo

$$S_0 = 0, \quad S_t = W_1 + W_2 + \ldots + W_t = \sum_{s=1}^{t} W_s,$$

dove $(W_s)_s$ è un rumore bianco gaussiano di instensità σ^2.

- Una definizione alternativa *ricorsiva* pone $S_0 = 0$ e per ogni $t \in \mathcal{T}$, $t \geq 1$,
$$S_t = S_{t-1} + W_t.$$

Esempio 2: passeggiata aleatoria gaussiana

Posto $\mathcal{T} = \{0, 1, \ldots, n\}$ o eventualmente $\mathcal{T} = \mathbb{N}$, definiamo

$$S_0 = 0, \quad S_t = W_1 + W_2 + \ldots + W_t = \sum_{s=1}^{t} W_s,$$

dove $(W_s)_s$ è un rumore bianco gaussiano di instensità σ^2.

- Una definizione alternativa *ricorsiva* pone $S_0 = 0$ e per ogni $t \in \mathcal{T}$, $t \geq 1$,
$$S_t = S_{t-1} + W_t.$$
- Il processo si intepreta come una "passeggiata", in cui ogni nuovo "passo" W_t sposta da S_{t-1} in $S_t = S_{t-1} + W_t$. È detto **passeggiata aleatoria gaussiana**.

Funzione di media e di autocovarianza

La media di ciascuna S_t è nulla:

$$\mathbb{E}[S_t] = \mathbb{E}\left[\sum_{s=1}^{t} W_s\right] = \sum_{s=1}^{t} \mathbb{E}[W_s] = 0.$$

- La varianza vale

$$\begin{aligned}\operatorname{Var}(S_t) &= \operatorname{Var}\left(\sum_{i=1}^{t} W_i\right) = \sum_{i=1}^{t} \operatorname{Var}(W_i)\\ &= \sum_{i=1}^{t} \sigma^2 = t\sigma^2.\end{aligned}$$

e non è costante. **La passeggiata aleatoria non è stazionaria.**

Funzione di media e di autocovarianza

La media di ciascuna S_t è nulla:

$$\mathbb{E}\left[S_t\right] = \mathbb{E}\left[\sum_{s=1}^{t} W_s\right] = \sum_{s=1}^{t} \mathbb{E}\left[W_s\right] = 0.$$

- La varianza vale

$$\begin{aligned}\mathrm{Var}\left(S_t\right) &= \mathrm{Var}\left(\sum_{i=1}^{t} W_i\right) = \sum_{i=1}^{t} \mathrm{Var}\left(W_i\right) \\ &= \sum_{i=1}^{t} \sigma^2 = t\sigma^2.\end{aligned}$$

  e non è costante. **La passeggiata aleatoria non è stazionaria.**

- Possiamo anche calcolare la funzione di autocovarianza $C(s,t) = \sigma^2 \min\{s,t\}$.

Stima del parametro σ^2

Si può stimare σ^2 partendo da n osservazioni $S_t = s_t$, per $t = 1, 2, \ldots, n$.

- Tramite una *differenza finita* (o derivata discreta) passiamo dalla passeggiata aleatoria al rumore bianco gaussiano:

  $$w_t = s_t - s_{t-1}$$

  si trovano n osservazioni.

Stima del parametro σ^2

Si può stimare σ^2 partendo da n osservazioni $S_t = s_t$, per $t = 1, 2, \ldots, n$.

- Tramite una *differenza finita* (o derivata discreta) passiamo dalla passeggiata aleatoria al rumore bianco gaussiano:

$$w_t = s_t - s_{t-1}$$

  si trovano n osservazioni.

- La stima di massima verosimiglianza è

$$\sigma^2_{\mathrm{MLE}} = \frac{1}{n} \sum_{t=1}^{n} (s_t - s_{t-1})^2.$$

Esempio 3: equazione lineare con smorzamento

Consideriamo una variante della passeggiata aleatoria: prima di ogni nuovo passo *trasformiamo* lo stato tramite una *dilatazione* di un parametro α. In formule:

- posto $(W_i)_i$ un rumore bianco gaussiano di instensità σ^2,

Esempio 3: equazione lineare con smorzamento

Consideriamo una variante della passeggiata aleatoria: prima di ogni nuovo passo *trasformiamo* lo stato tramite una *dilatazione* di un parametro α. In formule:

- posto $(W_i)_i$ un rumore bianco gaussiano di instensità σ^2,
- l'equazione ricorsiva è, per ogni $t \in \mathcal{T}$, $t \geq 1$,

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

Esempio 3: equazione lineare con smorzamento

Consideriamo una variante della passeggiata aleatoria: prima di ogni nuovo passo *trasformiamo* lo stato tramite una *dilatazione* di un parametro α. In formule:

- posto $(W_i)_i$ un rumore bianco gaussiano di instensità σ^2,
- l'equazione ricorsiva è, per ogni $t \in \mathcal{T}$, $t \geq 1$,

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

- Se $\alpha = 1$ è la passeggiata aleatoria.

Esempio 3: equazione lineare con smorzamento

Consideriamo una variante della passeggiata aleatoria: prima di ogni nuovo passo *trasformiamo* lo stato tramite una *dilatazione* di un parametro α. In formule:

- posto $(W_i)_i$ un rumore bianco gaussiano di instensità σ^2,
- l'equazione ricorsiva è, per ogni $t \in \mathcal{T}$, $t \geq 1$,

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

- Se $\alpha = 1$ è la passeggiata aleatoria.
- Se $|\alpha| < 1$, l'effetto è di riavvicinare X_{t-1} verso l'origine, uno *smorzamento*. Se non ci fosse il rumore, sarebbe esponenziale:

$$X_t = \alpha X_{t-1} = \alpha^2 X_{t-2} = \ldots = \alpha^t X_0.$$

Funzione di media e varianza

Supponiamo che X_0 abbia densità gaussiana di parametri $\mathcal{N}(0, \sigma_0^2)$ e sia indipendente dal rumore bianco gaussiano.

- La funzione di media del processo è costante e nulla:

$$\mathbb{E}[X_t] = \mathbb{E}[\alpha X_{t-1} + W_t] = \alpha \mathbb{E}[X_{t-1}] + \mathbb{E}[W_{t-1}] = \alpha \mathbb{E}[X_{t-1}],$$

e quindi, ripetendo t volte,

$$\mathbb{E}[X_t] = \alpha \mathbb{E}[X_{t-1}] = \alpha^2 \mathbb{E}[X_{t-2}] = \ldots = \alpha^t \mathbb{E}[X_0] = 0.$$

Funzione di media e varianza

Supponiamo che X_0 abbia densità gaussiana di parametri $\mathcal{N}(0, \sigma_0^2)$ e sia indipendente dal rumore bianco gaussiano.

- La funzione di media del processo è costante e nulla:

$$\mathbb{E}\left[X_t\right] = \mathbb{E}\left[\alpha X_{t-1} + W_t\right] = \alpha \mathbb{E}\left[X_{t-1}\right] + \mathbb{E}\left[W_{t-1}\right] = \alpha \mathbb{E}\left[X_{t-1}\right],$$

  e quindi, ripetendo t volte,

$$\mathbb{E}\left[X_t\right] = \alpha \mathbb{E}\left[X_{t-1}\right] = \alpha^2 \mathbb{E}\left[X_{t-2}\right] = \ldots = \alpha^t \mathbb{E}\left[X_0\right] = 0.$$

- Per la varianza usiamo che X_{t-1} è indipendente da W_t,

$$\mathrm{Var}\left(X_t\right) = \mathrm{Var}\left(\alpha X_{t-1} + W_t\right) = \alpha^2 \,\mathrm{Var}\left(X_{t-1}\right) + \sigma^2.$$

Condizione di stazionarietà

Sotto quali condizioni la varianza è costante $\mathrm{Var}(X_t) = \sigma^2 = \sigma_0^2$? deve valere

$$\sigma_0^2 = \alpha^2\sigma_0^2 + \sigma^2,$$

da cui

$$\sigma_0^2 = \frac{\sigma^2}{1-\alpha^2}.$$

- Il termine $1-\alpha^2$ deve essere positivo, e quindi troviamo la condizione

$$|\alpha| < 1.$$

Funzione di autocovarianza

Funzione di autocorrelazione

La funzione di autocorrelazione è $\rho(t) = \alpha^t$.

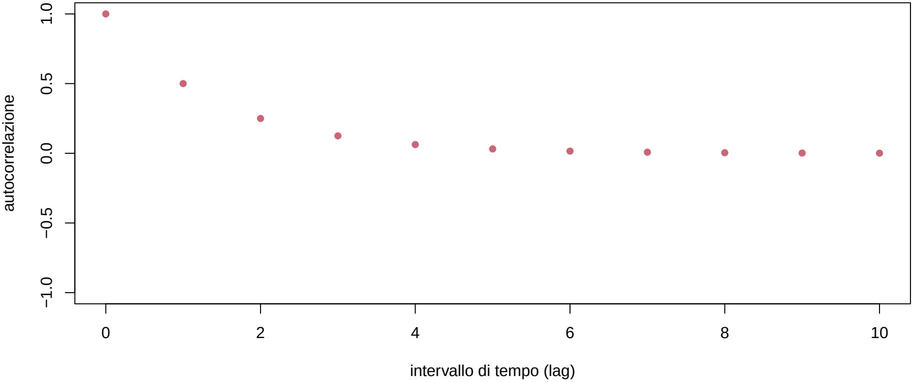


Figure 1: funzione di autocorrelazione dell'equazione lineare con smorzamento per $\alpha = 1/2$

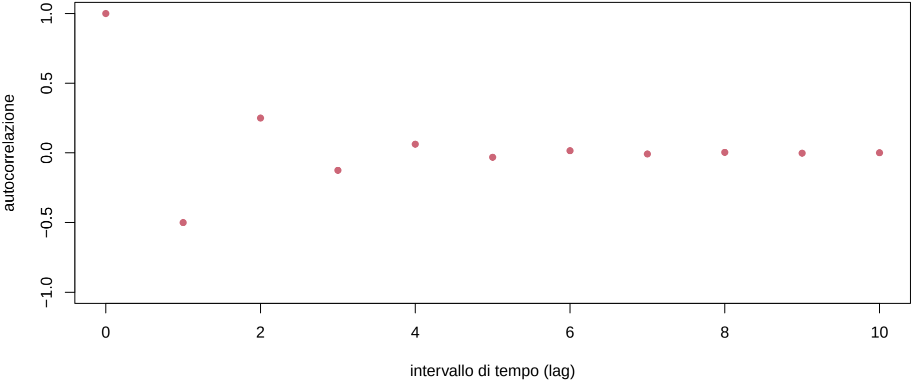


Figure 2: funzione di autocorrelazione dell'equazione lineare con smorzamento per $\alpha = -1/2$

Stima dei parametri

Possiamo stimare i parametri α e σ^2 da $n+1$ osservazioni $X_t = x_t$ per $t = 0, 1, \ldots, n$.

- Ci riconduciamo a n osservazioni di rumore bianco gaussiano

$$w_t = x_t - \alpha x_{t-1}.$$

Stima dei parametri

Possiamo stimare i parametri α e σ^2 da $n+1$ osservazioni $X_t = x_t$ per $t = 0, 1, \ldots, n$.

- Ci riconduciamo a n osservazioni di rumore bianco gaussiano

$$w_t = x_t - \alpha x_{t-1}.$$

- La stima di massima verosimiglianza si ottiene minimizzando

$$(\alpha, \sigma^2) \mapsto \sum_{t=1}^{n} \frac{(x_t - \alpha x_{t-1})^2}{\sigma^2} - n \log(\sigma^2)$$

In particolare, α_{MLE} minimizza la somma dei quadrati dei "residui"

$$\alpha \mapsto \sum_{t=1}^{n} (x_t - \alpha x_{t-1})^2,$$

Stima dei parametri

Possiamo stimare i parametri α e σ^2 da $n+1$ osservazioni $X_t = x_t$ per $t = 0, 1, \ldots, n$.

- Ci riconduciamo a n osservazioni di rumore bianco gaussiano

$$w_t = x_t - \alpha x_{t-1}.$$

- La stima di massima verosimiglianza si ottiene minimizzando

$$(\alpha, \sigma^2) \mapsto \sum_{t=1}^{n} \frac{(x_t - \alpha x_{t-1})^2}{\sigma^2} - n \log(\sigma^2)$$

  In particolare, α_{MLE} minimizza la somma dei quadrati dei "residui"

$$\alpha \mapsto \sum_{t=1}^{n} (x_t - \alpha x_{t-1})^2,$$

- Si trova $\alpha_{\mathsf{MLE}} = \frac{\sum_{t=1}^{n} x_t x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2}$, $\sigma^2_{\mathsf{MLE}} = \frac{1}{n} \sum_{t=1}^{n} (x_t - \alpha_{\mathsf{MLE}} x_{t-1})^2$.

Section 2

Modelli ARIMA

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:

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)

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)

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.

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 che $\mathcal{T} = \{0, 1, \ldots, n\}$ 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 σ^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

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

  si possono abbreviare come

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

Modelli AR

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}.

Modelli AR

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.

Modelli AR

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 AR(p) (autoregressivo di ordine p) se esistono parametri $\alpha_1, \alpha_2, \ldots, \alpha_p \in \mathbb{R}$ tali che, per ogni $t \in \mathcal{T}$ (tale che $t - p \in \mathcal{T}$) si abbia
$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + W_t.$$

Notazione compatta

Possiamo scrivere la ricorsione

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

in modo compatto:

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

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

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

Media mobile

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

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

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

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

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

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

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

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

- La funzione di autocovariaza è, usando la bilinearità,

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

  che dipende da $t - s$ solamente.

Modelli MA

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

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

- Un processo a media mobile MA(q) è semplicemente del tipo $W * g$, dove g è dato dai coefficienti 1, β_1, β_2, $\ldots$, β_q (e nullo altrove). In particolare, X è **gaussiano e stazionario**.

Modelli MA

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

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

- Un processo a media mobile MA(q) è semplicemente del tipo $W * g$, dove g è dato dai coefficienti 1, β_1, β_2, ..., β_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

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

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 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:

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

Esempi

- il rumore bianco gaussiano è ARIMA$(0,0,0)$,

Esempi

- il rumore bianco gaussiano è ARIMA$(0, 0, 0)$,
- la passeggiata aleatoria è ARIMA$(0, 1, 0)$

Esempi

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

Section 3

Proprietà dei processi ARIMA

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:

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

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:

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

- 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:

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

- 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 $k \to \infty$ 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

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

Criterio di stazionarietà

Il problema sta nella convergenza della serie, che dipende dalla crescita dei coefficienti b_k al tendere di $k \to \infty$ 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

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

- 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

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

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

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

- Nel caso ARIMA$(0,0,q)$ si ha

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

Dimostrazione

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 modellizza?

- 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 modellizza?

- 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 modellizza?

- 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:

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

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 deviazoni 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 deviazoni 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.

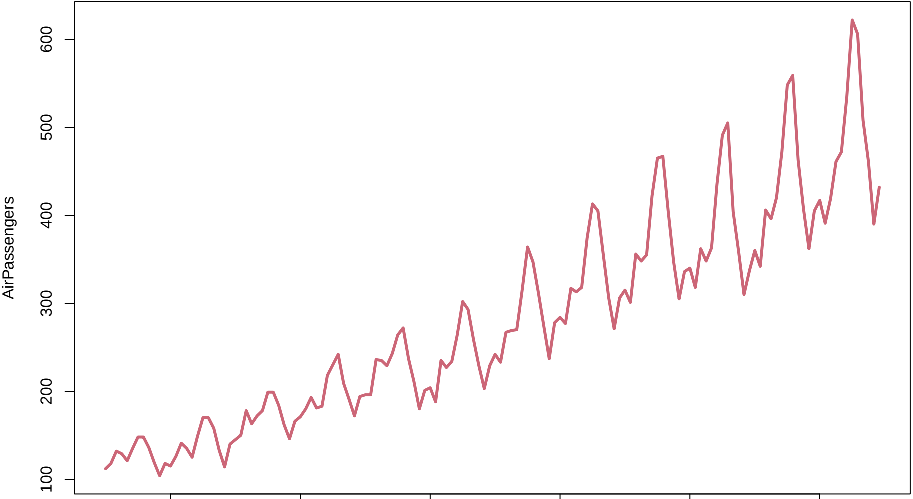
Scelta del modello

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`.



```
## 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 2.800458 (
```

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).

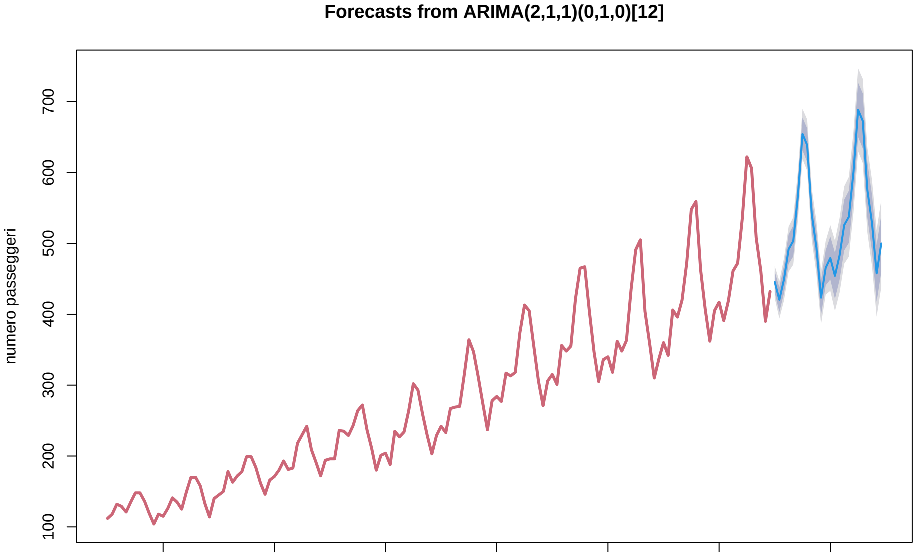
La funzione `auto.arima()` propone un modello ARIMA con stagionalità di periodo 12 (mesi) e ordine $(2,1,1)(0,1,0)$ (la seconda tripla si riferisce alla stagionalità).

- 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.

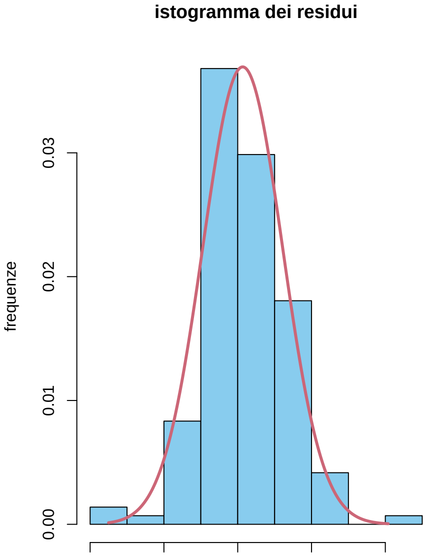
La funzione `auto.arima()` propone un modello ARIMA con stagionalità di periodo 12 (mesi) e ordine $(2,1,1)(0,1,0)$ (la seconda tripla si riferisce alla stagionalità).

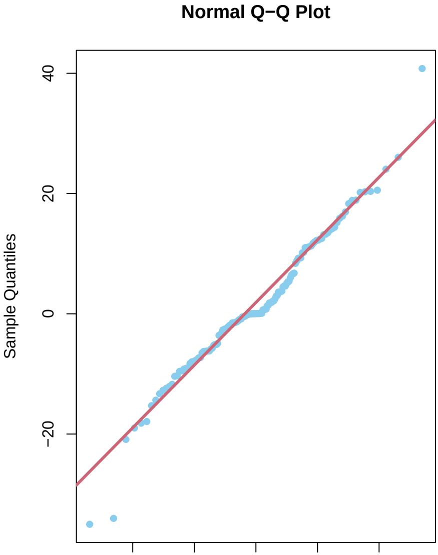
- In realtà funzione `auto.arima()` non determina automaticamente il periodo 12 e questo va indicato prima di applicarla ai dati, nel momento in cui si definisce un oggetto di tipo *serie storica* (in inglese *time series*) in R.
- Partendo da un vettore di dati osservati, il comando è `ts()`, che contiene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

Previsione


Forecasts from ARIMA(2,1,1)(0,1,0)[12]
numero passeggeri
100
200
300
400
500
600
700

Residui


istogramma dei residui
frequenze
0.00
0.01
0.02
0.03


Normal Q−Q Plot
Sample Quantiles
−20
0
20
40

Section 4

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 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 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.

Verosimiglianza

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)

Verosimiglianza

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_{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*:

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

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*:

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

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

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

Passaggio alla base delle frequenze

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

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

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

- 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

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

- È la somma due contributi:

Concludiamo che

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

- 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

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

- 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

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

- 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 $k << n$, 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 $S \in \{0, 1, \ldots, n-1-k\}$ 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

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

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

ACF in R

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

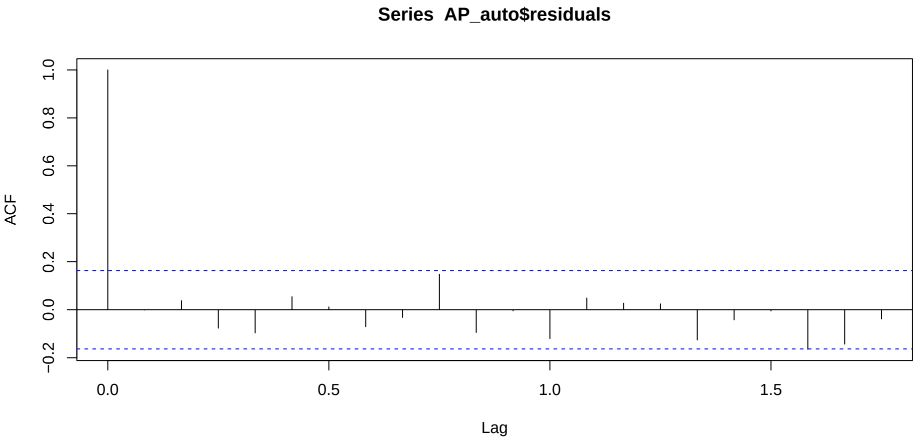


Figure 6: autocorrelazione empirica dei residui

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

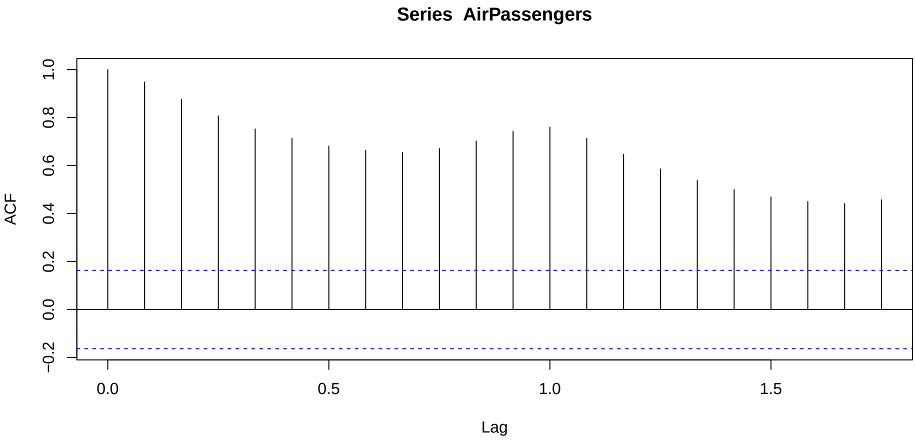


Figure 7: la funzione di autocorrelazione è uniformemente grande per via del trend

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

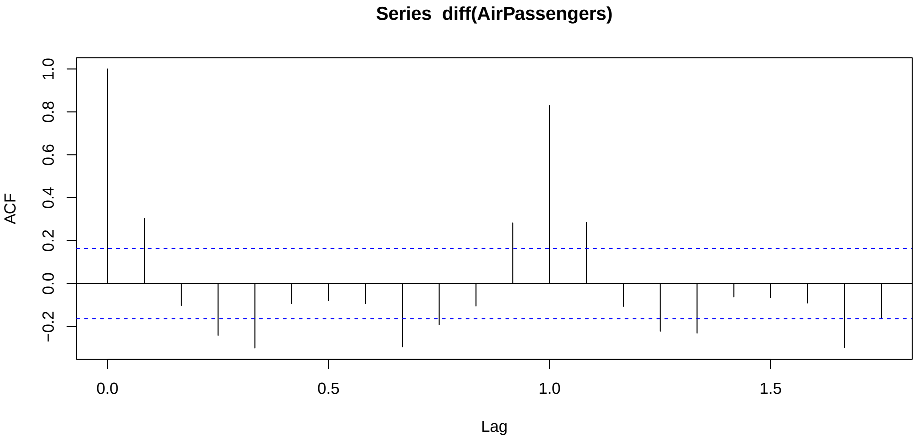


Figure 8: la derivata discreta rimuove il trend e la funzione di autocorrelazione evidenzia la stagionalità a 12 mesi

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 ξ diviso il tempo n, quindi si intepreta 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 ξ 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 ξ la quantità

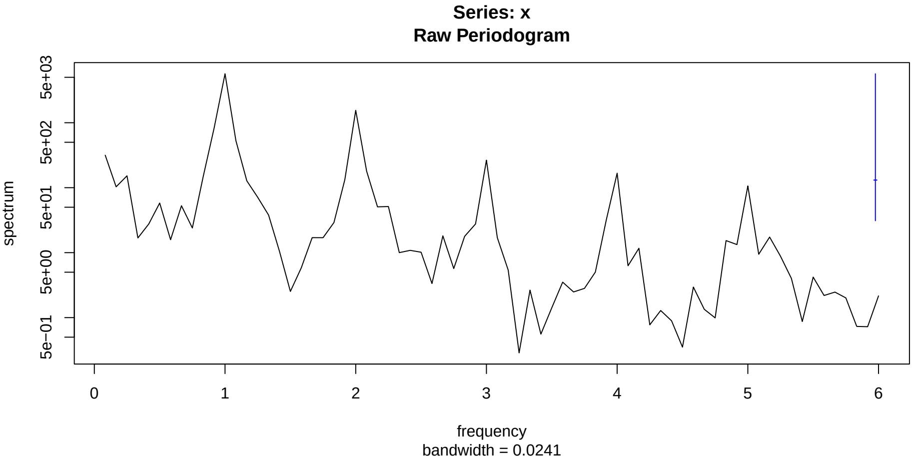
$$\hat{C}(\xi) = \sum_{k \in \mathbb{Z}} C(|k|) e^{-2\pi i k \xi}.$$

- Vale il limite

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

Spettrogramma in R

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



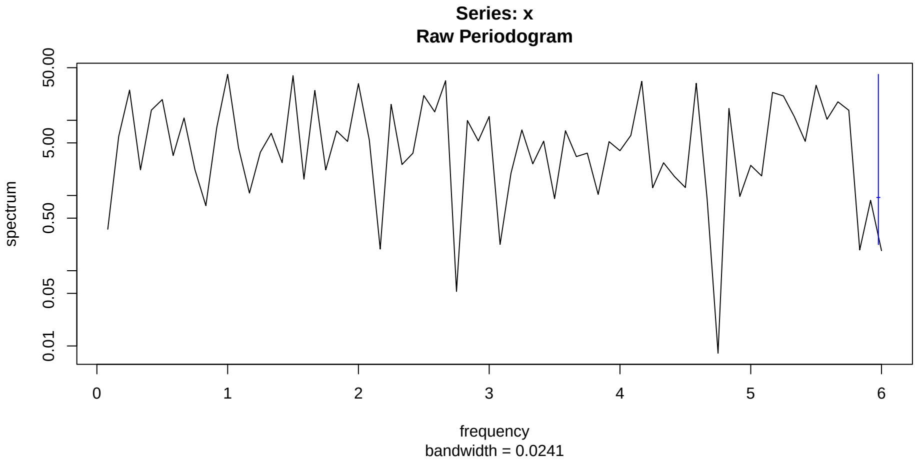


Figure 10: Stima della densità spettrale di potenza dei residui della serie AirPassengers dopo un fit con un modello ARIMA