

APPUNTI DI

PROCESSI STOCASTICI E SERIE TEMPORALI

1 Processi stocastici stazionari.

Un processo stocastico è una famiglia di variabili aleatorie X_t , indiciate dal parametro $t \in T$. Spesso, t rappresenta un tempo. Una traiettoria del processo X_t di lunghezza n , è una realizzazione di X_t agli istanti t_1, t_2, \dots, t_n , ovvero $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$.

Un processo stocastico X_t si dice *stazionario in senso stretto*, se la funzione di distribuzione congiunta n -dimensionale del processo:

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) := P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n) \quad (1.1)$$

risulta invariante rispetto alle traslazioni temporali, ovvero se, per ogni $\tau \in T$ risulta:

$$F_{X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}}(x_1, x_2, \dots, x_n) = F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) \quad (1.2)$$

Per $n = 1$ la (1.2) diventa

$$P(X_{t_1} \leq x_1) = P(X_{t_1+\tau} \leq x_1), \quad (1.3)$$

ovvero, per ogni coppia di tempi t_1, t_2 , si ha $X_{t_1} \sim X_{t_2}$ cioè $E(X_t)$ e $Var(X_t)$ sono costanti; mentre per $n = 2$, si ottiene

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2) = P(X_{t_1+\tau} \leq x_1, X_{t_2+\tau} \leq x_2); \quad (1.4)$$

se, in quest'ultima uguaglianza si prende $\tau = -t_1$, si ottiene

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2) = P(X_0 \leq x_1, X_{t_2-t_1} \leq x_2); \quad (1.5)$$

ciò significa che la distribuzione congiunta delle v.a. X_{t_1} e X_{t_2} non dipende in modo assoluto da t_1 e t_2 , ma solo dalla loro differenza $t_2 - t_1$. Nel seguito, salvo avviso contrario, supporremo che esistano finite tutte le medie che andremo a calcolare.

Per quanto riguarda $E(X_{t_1}X_{t_2})$ si ha:

$$E(X_{t_1}X_{t_2}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1x_2 f_{X_{t_1}, X_{t_2}}(x_1, x_2) dx_1 dx_2, \quad (1.6)$$

dove $f_{X_{t_1}, X_{t_2}}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_{t_1}, X_{t_2}}(x_1, x_2)$ è la densità congiunta di X_{t_1} e X_{t_2} . Allora, per un processo stazionario X_t si ottiene:

$$E(X_{t_1}X_{t_2}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1x_2 f_{X_0, X_{t_2-t_1}}(x_1, x_2) dx_1 dx_2, \quad (1.7)$$

e quindi la covarianza o funzione di correlazione del processo X_t , definita da $R(t_1, t_2) = E(X_{t_1}X_{t_2}) - E(X_{t_1})E(X_{t_2})$, non dipende in modo assoluto da t_1 e t_2 , ma solo dalla loro differenza $t_2 - t_1$.

Tutte le caratteristiche di un processo stazionario X_t non dipendono dal tempo t .

Un processo stocastico X_t si dice *stazionario in senso lato* se sono verificate le due condizioni seguenti:

- 1) La media $E(X_t) = m$ e la varianza $Var(X_t) = \sigma^2$ non dipendono dal tempo t .
- 2) La funzione di correlazione $R(t_1, t_2)$ non dipende in modo assoluto da t_1 e t_2 , ma solo dalla loro differenza $t_2 - t_1$.

Da quanto detto, segue immediatamente che un processo stazionario in senso stretto lo è anche in senso lato, ma non è vero il viceversa.

Esempio 1.1 (processo di Poisson) N_t si dice un processo di *Poisson* con intensità $\lambda > 0$, se:

- (i) $N_0 = 0$;
- (ii) (incrementi indipendenti) per $t_1 < t_2 < t_3 < t_4$, $N_{t_4} - N_{t_3}$ è stocasticamente indipendente da $N_{t_2} - N_{t_1}$;
- (iii) per ogni $s, t > 0$:

$$P(N_{t+s} - N_s = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \dots$$

Da (iii) segue che un processo di Poisson ha incrementi stazionari, ovvero $N_{t+s} - N_s$ ha la stessa distribuzione di N_t (omogeneità temporale); inoltre, si ha $E(N_t) = Var(N_t) = \lambda t$. Quindi, il processo di Poisson non è stazionario.

Esempio 1.2 (white noise) Un processo stocastico X_t si dice *white noise*, o *rumore bianco* se:

- (i) $E(X_t) = 0 \forall t$.
- (ii) $\forall t$ si ha $Var(X_t) = E((X_t)^2) = \sigma^2$, ove σ^2 è una costante positiva.
- (iii)

$$R(t_1, t_2) = \sigma^2 \delta(t_2 - t_1) = \begin{cases} 0, & \text{se } t_1 \neq t_2 \\ \sigma^2, & \text{se } t_1 = t_2 \end{cases}, \quad (1.8)$$

dove $\delta(x)$ è la funzione di Dirac, ovvero $\delta(x) = 0$ per $x \neq 0$ e $\int_{-\infty}^{+\infty} \phi(x)\delta(x)dx = \phi(0)$, per ogni funzione ϕ continua in un intorno di $x = 0$. Un processo white noise è stazionario in senso lato.

Per poter parlare di stazionarietà in senso stretto, dovremmo specificare la distribuzione congiunta n -dimensionale del processo. Per esempio, se supponiamo che, per ogni $t_1 \neq t_2$ la v.a. X_{t_1} è stocasticamente indipendente dalla v.a. X_{t_2} e che per ogni t , X_t ha la stessa distribuzione di una v.a. Y con media 0 e varianza σ^2 , risulta che la distribuzione di X_t è invariante rispetto a traslazioni temporali (ovviamente, $E(X_t) = 0$, ed essendo X_{t_1} e X_{t_2} indipendenti, $cov(X_{t_1}, X_{t_2}) = 0$). Pertanto, X_t è stazionario in senso stretto. Spesso, si prende $Y \sim \mathcal{N}(0, \sigma^2)$, ovvero una v.a. normale di media zero e varianza σ^2 .

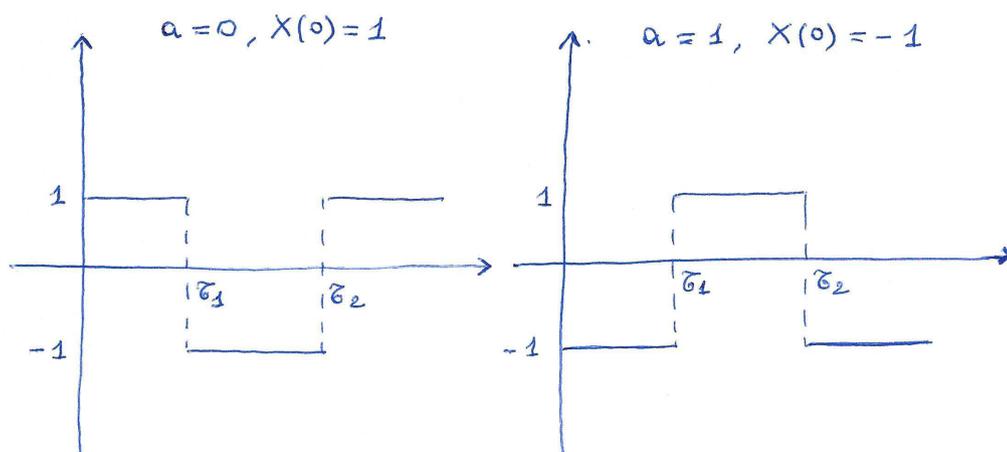


Figure 1: Due traiettorie del processo telegrafico casuale, rispettivamente nel caso che $X_0 = 1$ e $X_0 = -1$.

Esempio 1.3 (processo telegrafico casuale) E' definito da

$$X_t = X_0(-1)^{N_t} \quad (1.9)$$

dove N_t è un processo di Poisson omogeneo di parametro $\lambda > 0$, con $N_0 = 0$; inoltre X_0 è indipendente da N_t e $P(X_0 = 1) = P(X_0 = -1) = \frac{1}{2}$.

Si può anche scrivere $X_t = (-1)^{N_t+a} \in \{-1, 1\}$, dove a è una v.a. a valori in $\{0, 1\}$, tale che $P(a = 0) = P(a = 1) = \frac{1}{2}$. Naturalmente, si ha che $X_0 = (-1)^a = 1$, se $a = 0$, mentre $X_0 = -1$, se $a = 1$. Due tipiche traiettorie del processo sono illustrate nella Figura 1, rispettivamente nel caso che $X_0 = 1$ e $X_0 = -1$.

Si osservi che le traiettorie sono continue a tratti ed i tempi che intercorrono tra un istante di salto e il successivo sono variabili aleatorie indipendenti ed esponenziali di parametro λ . Calcoliamo ora $P(X_t = 1)$. Si ha:

$$\begin{aligned} P(X_t = 1) &= P(X_t = 1|X_0 = 1)P(X_0 = 1) + P(X_t = 1|X_0 = -1)P(X_0 = -1) \\ &= \frac{1}{2}[P(X_t = 1|X_0 = 1) + P(X_t = 1|X_0 = -1)]; \end{aligned}$$

ora, si vede facilmente che

$P(X_t = 1|X_0 = 1) = P(N_t \text{ è pari})$ e $P(X_t = 1|X_0 = -1) = P(N_t \text{ è dispari})$, dunque

$$P(X_t = 1) = \frac{1}{2}[P(N_t \text{ è pari}) + P(N_t \text{ è dispari})].$$

Dalla formula del processo di Poisson, si ha per $k = 0, 1, \dots$:

$$P(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

pertanto:

$$P(N_t \text{ è pari}) = e^{-\lambda t} \sum_{i=0}^{\infty} \frac{(\lambda t)^{2i}}{(2i)!} \text{ e } P(N_t \text{ è dispari}) = e^{-\lambda t} \sum_{i=0}^{\infty} \frac{(\lambda t)^{2i+1}}{(2i+1)!} \quad (1.10)$$

Dagli sviluppi in serie di Taylor di $e^{\lambda t}$ e $e^{-\lambda t}$, si ottiene:

$$e^{\lambda t} + e^{-\lambda t} = 2 \sum_{i=0}^{\infty} \frac{(\lambda t)^{2i}}{(2i)!}, \quad (1.11)$$

da cui

$$\sum_{i=0}^{\infty} \frac{(\lambda t)^{2i}}{(2i)!} = \frac{1}{2}(e^{\lambda t} + e^{-\lambda t}), \quad (1.12)$$

e quindi

$$P(N_t \text{ è pari}) = \frac{1}{2}e^{-\lambda t} (e^{\lambda t} + e^{-\lambda t}) = \frac{1}{2} (1 + e^{-2\lambda t}). \quad (1.13)$$

Analogamente, risulta

$$e^{\lambda t} - e^{-\lambda t} = 2 \sum_{i=0}^{\infty} \frac{(\lambda t)^{2i+1}}{(2i+1)!}, \quad (1.14)$$

da cui

$$\sum_{i=0}^{\infty} \frac{(\lambda t)^{2i+1}}{(2i+1)!} = \frac{1}{2}(e^{\lambda t} - e^{-\lambda t}), \quad (1.15)$$

e quindi

$$P(N_t \text{ è dispari}) = \frac{1}{2}e^{-\lambda t} (e^{\lambda t} - e^{-\lambda t}) = \frac{1}{2} (1 - e^{-2\lambda t}). \quad (1.16)$$

Tornando all' espressione di $P(X_t = 1)$, possiamo scrivere:

$$P(X_t = 1) = \frac{1}{2} \left(\frac{1}{2}(1 + e^{-2\lambda t}) + \frac{1}{2}(1 - e^{-2\lambda t}) \right) = \frac{1}{2}. \quad (1.17)$$

Abbiamo quindi concluso che

$$P(X_t = 1) = \frac{1}{2}, \quad (1.18)$$

e, in maniera analoga, possiamo mostrare che $P(X_t = -1) = \frac{1}{2}$, quindi:

$$X_t \in \begin{cases} +1, & \text{con probabilità } \frac{1}{2} \\ -1, & \text{con probabilità } \frac{1}{2} \end{cases}. \quad (1.19)$$

Da quanto trovato, segue subito che $E(X_t) = 0$. Calcoliamo ora $E(X_{t_1} X_{t_2})$, per $t_2 > t_1$; siccome $X_t \in \{-1, 1\}$, anche $Z(t_1, t_2) := X_{t_1} X_{t_2} \in \{-1, 1\}$. Quindi

$$E(X_{t_1} X_{t_2}) = 1 \cdot P(Z(t_1, t_2) = 1) + (-1) \cdot P(Z(t_1, t_2) = -1) = P(X_{t_2} = X_{t_1}) - P(X_{t_2} \neq X_{t_1}).$$

Calcoliamo prima $P(X_{t_2} = X_{t_1})$; dire che $X_{t_2} = X_{t_1}$ significa dire che il processo X_t assume lo stesso valore negli istanti t_1 e t_2 . Ciò accade quando nell' intervallo (t_1, t_2) avviene un numero pari di eventi di Poisson, per cui

$$P(X_{t_2} = X_{t_1}) = P(N_{t_2-t_1} \text{ è pari}).$$

Analogamente, dire che $X_{t_2} \neq X_{t_1}$ significa dire che il processo X_t assume valori diversi negli istanti t_1 e t_2 . Ciò accade quando nell' intervallo (t_1, t_2) avviene un numero dispari di eventi di Poisson, per cui

$$P(X_{t_2} \neq X_{t_1}) = P(N_{t_2-t_1} \text{ è dispari}). \quad (1.20)$$

Pertanto $E(X_{t_1}X_{t_2}) = P(N_{t_2-t_1} \text{ è pari}) - P(N_{t_2-t_1} \text{ è dispari})$. Sostituendo $t_2 - t_1$ in (1.13) e (1.16), si trova

$$P(N_{t_2-t_1} \text{ è pari}) = \frac{1}{2} (1 + e^{-2\lambda(t_2-t_1)}) \text{ e } P(N_{t_2-t_1} \text{ è dispari}) = \frac{1}{2} (1 - e^{-2\lambda(t_2-t_1)}).$$

Concludiamo che

$$E(X_{t_1}X_{t_2}) = \frac{1}{2} (1 + e^{-2\lambda(t_2-t_1)}) - \frac{1}{2} (1 - e^{-2\lambda(t_2-t_1)}) = e^{-2\lambda(t_2-t_1)},$$

e quindi

$$R(t_1, t_2) = E(X_{t_1}X_{t_2}) - E(X_{t_1})E(X_{t_2}) = E(X_{t_1}X_{t_2}) - 0 = e^{-2\lambda(t_2-t_1)}. \quad (1.21)$$

Per calcolare la media e la funzione di correlazione, si può anche procedere più semplicemente, come segue.

$$\begin{aligned} E[X_t] &= E[X_0(-1)^{N_t}] \\ &= E[X_0]E[(-1)^{N_t}] \text{ (per la indipendenza di } X_0 \text{ e } N_t) \\ &= 0 \text{ (visto che } E(X_0) = 0). \end{aligned} \quad (1.22)$$

$$\begin{aligned} E[X(t)X(t+s)] &= E[X_0^2(-1)^{N_t+N_{t+s}}] = E[(-1)^{2N_t}(-1)^{N_{t+s}-N_t}] \\ &= E[(-1)^{N_{t+s}-N_t}] = E[(-1)^{N_s}] \\ &= \sum_{i=0}^{\infty} (-1)^i e^{-\lambda s} \frac{(\lambda s)^i}{i!} = e^{-\lambda s} \sum_{i=0}^{\infty} \frac{(-\lambda s)^i}{i!} = e^{-\lambda s} e^{-\lambda s} = e^{-2\lambda s} \end{aligned} \quad (1.23)$$

che coincide con (1.21) per $t_1 = t$, $t_2 = t + s$.

La funzione di correlazione dipende solo da $t_2 - t_1$; quindi il processo telegrafico casuale è stazionario in senso lato.

Si può inoltre dimostrare che esso è stazionario anche in senso stretto.

Esempio 2.4 (processo Gaussiano)

E' un processo stocastico X_t per cui il vettore $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ ha distribuzione normale multivariata, per ogni t_1, t_2, \dots, t_n . Siccome la distribuzione di una variabile normale multivariata è completamente individuata dal vettore delle medie $(E(X_{t_1}), E(X_{t_2}), \dots, E(X_{t_n}))$ e dalle covarianze $cov(X_{t_i}, X_{t_j})$, $i, j = 1, 2, \dots, n$, un processo Gaussiano è completamente specificato, una volta assegnate le medie e le covarianze. Un processo Gaussiano debolmente stazionario, lo è anche in senso stretto.

Un esempio di processo Gaussiano è il seguente

Esempio 2.5 (moto Browniano, o processo di Wiener)

E' il processo $B(t)$ (denotato anche con B_t) con le seguenti proprietà:

- (i) $B(0) = 0$;
- (ii) (indipendenza degli incrementi) per $t_1 < t_2 < t_3 < t_4$, $B(t_2) - B(t_1)$ è indipendente da $B(t_4) - B(t_3)$;
- (iii) (incrementi normali) per $t > s$, $B(t) - B(s)$ ha distribuzione normale con media zero e varianza $t - s$; in particolare, $B(t) \sim \mathcal{N}(0, t)$.

Per ogni $t_1 < t_2 < \dots < t_n$, $(B_{t_1}, B_{t_2}, \dots, B_{t_n})$ è una v.a. n-dimensionale normale centrata, per cui B_t è un processo Gaussiano, e si ha: $\text{cov}(B(t), B(s)) = E(B(t)B(s)) = \min\{s, t\}$; infatti, se $s < t$:

$$E(B_t B_s) = E(B_s(B_s + B_t - B_s)) = E(B_s^2) + E(B_s(B_t - B_s)) =$$

(per l'indipendenza degli incrementi)

$$= E(B_s^2) + E(B_s)E(B_t - B_s) = E(B_s^2) = s$$

(poiché $E(B_s) = 0$). Inoltre $B(t)$ è una funzione continua di t (continuità della traiettoria). E' immediato constatare che, sebbene $E(B_t) = 0$, B_t non è un processo stazionario, poiché $\text{var}(B_t) = t$ (dipende dal tempo), e $\text{cov}(B_s, B_t) = \min\{s, t\}$ non è una funzione di $|t - s|$.

Moto Browniano come approssimazione continua di un Random Walk

Consideriamo una particella che parte dalla posizione 0 al tempo $t = 0$ e si muove, ad istanti successivi, $k\Delta t$, nella posizione $k\Delta x$, per $k = 0, 1, \dots$. Supponiamo che, se al tempo t si trova nella posizione x , al tempo $t + \Delta t$ si muove a destra nella posizione $x + \Delta x$ con probabilità $1/2$, e a sinistra nella posizione $x - \Delta x$ con probabilità $1/2$. Indichiamo con $u(t, x)$ la densità di probabilità che la particella si trovi nella posizione x all'istante t . Allora si ha:

$$u(t + \Delta t, x) = \frac{1}{2}u(t, x - \Delta x) + \frac{1}{2}u(t, x + \Delta x). \quad (1.24)$$

Dallo sviluppo di Taylor di $u(t, x)$ si trova:

$$u(t + \Delta t, x) = u(t, x) + \frac{\partial u}{\partial t}(t, x)\Delta t + o(\Delta t) \quad (\Delta t \rightarrow 0),$$

$$u(t, x - \Delta x) = u(t, x) + \frac{\partial u}{\partial x}(t, x) \cdot (-\Delta x) + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t, x) \cdot \Delta^2 x + o(\Delta^2 x) \quad (\Delta x \rightarrow 0),$$

$$u(t, x + \Delta x) = u(t, x) + \frac{\partial u}{\partial x}(t, x) \cdot \Delta x + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t, x) \cdot \Delta^2 x + o(\Delta^2 x) \quad (\Delta x \rightarrow 0).$$

Inserendo in (1.24), e trascurando i termini di ordine superiore ad 1 in Δt e di ordine superiore a 2 in Δx , otteniamo:

$$\begin{aligned} & u(t, x) + \frac{\partial u}{\partial t}(t, x)\Delta t \\ \approx & \frac{1}{2} \left[\left(u(t, x) - \Delta x \frac{\partial u}{\partial x}(t, x) + \frac{\Delta^2 x}{2} \frac{\partial^2 u}{\partial x^2}(t, x) \right) + \left(u(t, x) + \Delta x \frac{\partial u}{\partial x}(t, x) + \frac{\Delta^2 x}{2} \frac{\partial^2 u}{\partial x^2}(t, x) \right) \right], \end{aligned}$$

da cui si ottiene:

$$\frac{\partial u}{\partial t}(t, x)\Delta t \approx \frac{1}{2} \Delta^2 x \frac{\partial^2 u}{\partial x^2}(t, x),$$

e quindi:

$$\frac{\partial u}{\partial t}(t, x) \approx \frac{1}{2} \frac{\Delta^2 x}{\Delta t} \frac{\partial^2 u}{\partial x^2}(t, x).$$

Passando al limite, per $\Delta x \rightarrow 0$ e $\Delta t \rightarrow 0$, in maniera che $\frac{\Delta^2 x}{\Delta t} \rightarrow \sigma^2$, dove $\sigma^2 > 0$ è una costante (ovvero, se $\Delta x \sim \sigma\sqrt{\Delta t}$), si ottiene infine:

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \sigma^2 \frac{\partial^2 u}{\partial x^2}(t, x). \quad (1.25)$$

E' facile verificare che la soluzione di (1.25), che verifica la condizione di normalizzazione (si tratta di una densità di probabilità)

$$\int_{-\infty}^{+\infty} u(t, x) dx = 1,$$

è:

$$u(t, x) = \frac{1}{\sigma\sqrt{2\pi t}} e^{-x^2/2\sigma^2 t}, \quad (1.26)$$

che è la densità di una v.a. Normale di media 0 e varianza $\sigma^2 t$. Il moto Browniano standard B_t si ottiene per $\sigma = 1$, ed ha distribuzione $\mathcal{N}(0, t)$; una sua caratteristica è che $dB_t \sim \sqrt{dt}$. Il processo stocastico definito da $X(t) = \mu t + B_t$, dove μ è una costante, si chiama moto Browniano con drift μ .

Esempio 2.6 (processo di Ornstein-Uhlenbeck stazionario)

E' il processo definito da:

$$X_t = e^{-\beta t} B \left(\frac{\sigma^2}{2\beta} e^{2\beta t} \right),$$

dove $\beta, \sigma^2 > 0$ e $B(t)$ è un moto Browniano.

Si ha

$$E(X_t) = e^{-\beta t} E \left[B \left(\frac{\sigma^2}{2\beta} e^{2\beta t} \right) \right] = 0; \quad (1.27)$$

per $s < t$:

$$\begin{aligned} cov(X_s, X_t) &= E \left[e^{-\beta s} e^{-\beta t} B \left(\frac{\sigma^2}{2\beta} e^{2\beta s} \right) \cdot B \left(\frac{\sigma^2}{2\beta} e^{2\beta t} \right) \right] \\ &= e^{-\beta(t+s)} \min \left(\frac{\sigma^2}{2\beta} e^{2\beta s}, \frac{\sigma^2}{2\beta} e^{2\beta t} \right) = e^{-\beta(t-s)} \frac{\sigma^2}{2\beta}. \end{aligned} \quad (1.28)$$

Dunque $Var(X_t) = \frac{\sigma^2}{2\beta}$ e $R(s, t) = \frac{\sigma^2}{2\beta} e^{-\beta(t-s)}$.

Pertanto, X_t è stazionario in senso lato, ma lo è anche in senso stretto, visto che è un processo Gaussiano (essendolo B_t). Infatti, la distribuzione n-dimensionale di un processo Gaussiano è univocamente determinata da medie e covarianze.

Esempio 2.7 Sia U una v.a. uniformemente distribuita in $\{1, 2, 3, 4, 5\}$, ovvero $P(U = i) = 1/5$, $i = 1, 2, \dots, 5$. Costruiamo un processo stocastico X_t nel seguente modo:

- se $U = 1$, poniamo $X_t = g(t, 1) = -\sqrt{2} \cos t$;
- se $U = 2$, poniamo $X_t = g(t, 2) = -\sqrt{2} \sin t$;
- se $U = 3$, poniamo $X_t = g(t, 3) = \sqrt{2}(\cos t + \sin t)$;
- se $U = 4$, poniamo $X_t = g(t, 4) = \cos t - \sin t$;
- se $U = 5$, poniamo $X_t = g(t, 5) = \sin t - \cos t$.

Allora:

$$\begin{aligned} E(X_t) &= \frac{1}{5} \sum_{i=1}^5 g(t, i) \\ &= \frac{1}{5} [-\sqrt{2} \cos t - \sqrt{2} \sin t + \sqrt{2}(\cos t + \sin t) + \cos t - \sin t + \sin t - \cos t] = 0. \end{aligned} \quad (1.29)$$

Inoltre

$$R(t, t + \tau) = E(X_t X_{t+\tau}) = \frac{1}{5} \sum_{i=1}^5 g(t, i) g(t + \tau, i) \quad (1.30)$$

e, facendo i calcoli, si trova $R(t, t + \tau) = \frac{6}{5} \cos \tau$. Quindi, il processo X_t è stazionario in senso lato; si può però verificare che esso non lo è in senso stretto.

Esempio 2.8 Siano U e V v.a. indipendenti ed uniformemente distribuite in $(0, 1)$, e consideriamo il processo stocastico definito da $X_t = U \sin(tV)$.

Siccome U e V sono indipendenti, si ha $E(X_t) = E(U)E(\sin tV)$; inoltre, $E(U) = 1/2$ e

$$E(\sin tV) = \int_0^1 \sin tv dv = \frac{1}{t}(1 - \cos t).$$

Dunque, $E(X_t) = \frac{1}{2t}(1 - \cos t)$ è funzione di t , per cui il processo X_t non è stazionario, nemmeno in senso lato. Si può provare inoltre che la funzione di correlazione $R(s, t)$ dipende da s e t , e non solo da $t - s$.

2 Autocovarianza e autocorrelazione di una serie temporale. Correlogramma

Per un processo stocastico X_t è utile considerare un campione di ampiezza n di una sua realizzazione $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ agli istanti t_k , $k = 1, \dots, n$. Un tale campione di valori viene detto *serie temporale*; naturalmente, se $n = \infty$, la serie temporale diventa una sequenza infinita di valori, bilatera se si considerano anche t_k negativi.

Dato un processo stocastico stazionario X_t , di media costante μ , la sua *autocovarianza* al ritardo τ , è definita da

$$\gamma(\tau) = \text{cov}(X_t, X_{t-\tau}) = E[(X_t - \mu)(X_{t-\tau} - \mu)] = E[X_t X_{t-\tau}] - \mu^2 \quad (2.1)$$

ovvero la covarianza tra il valore del processo al tempo t e quello assunto in corrispondenza del tempo $t - \tau$, cioè ritardato di τ .

Analogamente, se i tempi sono discreti e $t \in \{1, 2, \dots\}$, l'*autocovarianza* al ritardo k , è definita da

$$\gamma_k = \text{cov}(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)] = E[X_t X_{t-k}] - \mu^2, \quad k = 0, 1, \dots \quad (2.2)$$

ovvero la covarianza tra il valore del processo al tempo t e quello assunto in corrispondenza del tempo $t - k$, cioè ritardato di k .

Naturalmente, risulta $\gamma_0 = \text{Var}(X_t)$. Inoltre, per la stazionarietà si ha anche $\gamma_k = \gamma_{-k}$; infatti,

$$\gamma_k = E(X_t X_{t-k}) - \mu^2 = E(X_{t-k} X_t) - \mu^2 = E(X_t X_{t+k}) - \mu^2 = \gamma_{-k}. \quad (2.3)$$

Si definisce poi la funzione di autocorrelazione globale, ovvero il coefficiente di correlazione lineare tra le v.a. X_t e X_{t-k} :

$$\rho(k) = \rho_k = \frac{\gamma_k}{\gamma_0}. \quad (2.4)$$

Si ha $\rho_0 = 1$, $\rho_k = \rho_{-k}$ (per la (2.3)) e, per le note proprietà del coefficiente di correlazione $|\rho_k| \leq 1$.

Inoltre ρ_k è invariante in modulo rispetto a qualsiasi trasformazione lineare di $\{X_t\}$. Infine, resta definita la matrice di correlazione:

$$P_n = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \dots & \rho_{n-2} \\ \dots & \dots & \dots & \dots \\ \rho_{n-1} & \rho_{n-2} & \dots & 1 \end{pmatrix} \quad (2.5)$$

i cui elementi rappresentano le autocorrelazioni tra X_t e X_{t-k} , $k = 0, 1, \dots, n-1$; essa è detta matrice di autocorrelazione di Toeplitz di ordine n . Gli elementi di ciascuna diagonale sono uguali tra loro, inoltre la matrice è simmetrica rispetto alle sue diagonali principali, il suo determinante assume valori in $(0, 1)$, ed è definita positiva, ovvero il suo determinante e tutti i minori principali sono maggiori di zero (la definita positività della matrice è anche equivalente al fatto che tutti i suoi autovalori sono positivi).

La funzione di autocorrelazione parziale

ρ_k è la funzione di autocorrelazione globale, poiché tiene conto delle correlazioni corrispondenti a tutti i ritardi. Per esempio, ρ_2 misura la correlazione tra X_t e X_{t-2} , ma incorpora anche quella tra X_t e X_{t-1} e tra X_{t-1} e X_{t-2} .

La funzione di autocorrelazione parziale di un processo stazionario X_t al ritardo k è definita come la correlazione lineare fra X_t e X_{t-k} , al netto di tutte le correlazioni lineari intermedie, ossia quelle tra X_t e X_{t-1} , tra X_t e X_{t-2} , e così' via fino alla correlazione lineare tra X_t e X_{t-k+1} . La funzione di autocorrelazione parziale ϕ_{kk} può essere espressa come:

$$\phi_{kk} = E[(X_t - E(X_t|X_{t-1}, X_{t-2}, \dots, X_{t-k+1}))(X_{t-k} - E(X_{t-k}|X_{t-1}, X_{t-2}, \dots, X_{t-k+1}))], \quad (2.6)$$

per $k = 1, 2, \dots$. Si può dimostrare che ([Box e Jenkins])

$$\phi_{kk} = \frac{|Q_k|}{|P_k|}, \quad k = 1, 2, \dots \quad (2.7)$$

dove P_k è la matrice di Toeplitz di ordine k , che è funzione delle autocorrelazioni globali fino al ritardo $k-1$ e Q_k è la stessa matrice, dove però l'ultima colonna è sostituita dal vettore $(\rho_1, \rho_2, \dots, \rho_k)^T$:

$$\phi_{kk} = \frac{\det \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \dots & \rho_2 \\ \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_k \end{pmatrix}}{\det \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{pmatrix}} \quad (2.8)$$

Posto per convenzione $\phi_{00} = 1$, le successive tre autocorrelazioni parziali sono:

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\det \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{pmatrix}}{\det \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\det \begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{pmatrix}}{\det \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}} = \frac{\rho_3(1 - \rho_1^2) + \rho_1(\rho_1^2 + \rho_2^2 - 2\rho_2)}{(1 - \rho_2)(1 + \rho_2 - 2\rho_1^2)}.$$

2.1 Stima di quantità di interesse

Dato un processo stazionario y_t , si può stimarne la media η mediante:

$$\hat{\eta} = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t. \quad (2.9)$$

In maniera analoga, si stimano la funzione di autocovarianza e autocorrelazione mediante:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=h+1}^n (y_t - \bar{y})(y_{t-h} - \bar{y}), \quad \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \quad (2.10)$$

Osservazione 1. Si osservi che dividiamo per n e non per $n - h$ che è il numero degli addendi; in questo modo anche la stima (e non solo quello che si vuole stimare) gode delle proprietà richieste. Ad esempio, dividendo per $n - h$ potremmo ottenere stime dei coefficienti di autocorrelazione, in modulo, più grandi di 1. Dividendo per n introduciamo, però, una distorsione verso lo zero nella stima (sottostimiamo in maniera sistematica la correlazione esistente).

Osservazione 2. $\hat{\gamma}(h)$ non è definita se $h > n - 1$; questo è ovvio: con n osservazioni non abbiamo nessuna coppia di osservazioni distanti n o $n + 1$ o etc.

Osservazione 3. Ha senso calcolare $\hat{\gamma}(h)$ solo se $n - h$, ovvero il numero di addendi su cui è basata la stima, è sufficientemente grande. Questo non è un grande problema nelle applicazioni, visto che tipicamente si è interessati alla funzione di autocovarianza (o di autocorrelazione) solamente per ritardi non grandi. Però va sempre tenuto presente.

Osservazione 4. Per il calcolo di η e $\hat{\rho}(h)$ non è necessaria la stazionarietà. A volte si usa $\hat{\rho}(h)$ per avere una idea “media nel tempo” della dipendenza lineare esistente.

DEFINIZIONE Il grafico di ρ_k in funzione di k . viene chiamato *correlogramma*.

Per ogni h , $\rho(h)$ rappresenta il coefficiente di correlazione tra due variabili del processo X_t separate da un intervallo di lunghezza h . Al crescere di h , ci aspettiamo che la correlazione tra X_t e X_{t+h} diminuisca. Infatti, se h è grande, in generale il processo all’istante $t + h$ avrà “dimenticato” il valore assunto all’istante t . Di conseguenza, ci aspettiamo che sia $\gamma(h)$ che $\rho(h)$ tendano a zero per $|h| \rightarrow \infty$. Il modo in cui $\rho(h)$ decresce a zero può essere interpretato come una misura della “memoria” del processo. Si ricordi comunque che stiamo parlando di casi “tipici”; ci sono infatti processi per i quali $\rho(h)$ non decresce a zero per $|h| \rightarrow \infty$, ma si comporta in maniera sinusoidale.

Esempio 2.1 Se abbiamo a disposizione un campione di numerosità $n = 5$, diciamo

$\{y_1, y_2, \dots, y_5\}$, nell' ipotesi che $\bar{y} = \eta = 0$, si ha:

$$\hat{\gamma}(1) = \frac{1}{5} \sum_{t=2}^5 y_t y_{t-1} = \frac{1}{5} (y_2 y_1 + y_3 y_2 + y_4 y_3 + y_5 y_4)$$

$$\hat{\gamma}(2) = \frac{1}{5} \sum_{t=3}^5 y_t y_{t-2} = \frac{1}{5} (y_3 y_1 + y_4 y_2 + y_5 y_3)$$

$$\hat{\gamma}(3) = \frac{1}{5} \sum_{t=4}^5 y_t y_{t-3} = \frac{1}{5} (y_4 y_1 + y_5 y_2)$$

$$\hat{\gamma}(4) = \frac{1}{5} \sum_{t=5}^5 y_t y_{t-4} = \frac{1}{5} y_5 y_1$$

$\hat{\gamma}(h)$ è definito solo per $h \leq n - 1 = 4$, quindi $\hat{\gamma}(5)$ non è definito.

Ricordiamo che $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$ e $\hat{\gamma}(0) = \frac{1}{n} \sum_{t=1}^m (y_t - \bar{y})^2$ è una stima di $\sigma^2 = Var(y_t)$.

Si può dimostrare che se non esiste autocorrelazione nel processo (ovvero se $\rho(h) = 0$ quando $h \neq 0$, cioè se, come si usa dire, il processo osservato è un rumore bianco (white noise)) allora la distribuzione asintotica di $\sqrt{n}\hat{\rho}(h)$ è una normale di media nulla e varianza uno.

Quindi nel caso di una serie senza autocorrelazione, $\hat{\rho}(h)$ “cadrà” nell'intervallo $[-\frac{1}{\sqrt{n}}\phi_{1-\alpha/2}, \frac{1}{\sqrt{n}}\phi_{1-\alpha/2}]$ (ϕ_β è il quantile di ordine β di una normale standard, ovvero $\Phi(\phi_\beta) = \beta$, dove Φ denota la funzione di distribuzione di una normale standard) approssimativamente con una probabilità maggiore o uguale a $1 - \alpha$ (ovviamente n deve essere sufficientemente grande, $n > 50$ sembra essere sufficiente). Infatti, se $\hat{\rho}(h)\sqrt{n} \sim \mathcal{N}(0, 1)$, allora

$$P(-\delta < \hat{\rho}(h)\sqrt{n} < \delta) = \Phi(\delta) - \Phi(-\delta) = 2\Phi(\delta) - 1;$$

per $\delta = \phi_{1-\alpha/2}$, si ottiene

$$P(-\phi_{1-\alpha/2} < \hat{\rho}(h)\sqrt{n} < \phi_{1-\alpha/2}) = 1 - \alpha,$$

ovvero

$$P(-\frac{1}{\sqrt{n}}\phi_{1-\alpha/2} < \hat{\rho}(h) < \frac{1}{\sqrt{n}}\phi_{1-\alpha/2}) = 1 - \alpha.$$

Per questo motivo, nei grafici della funzione di autocorrelazione empirica (ovvero quella stimata dai dati), sono spesso indicate delle bande del tipo $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$ (si osservi che $\phi_{0.975} = 1.96$).

Valori di $\hat{\rho}(h)$, per quanto diversi da zero, ma all'interno di queste bande suggeriscono che l'autocorrelazione “stimata” potrebbe essere in realtà dovuta al caso (ovvero non essere una proprietà del processo).

Si osservi, comunque, che anche in assenza di autocorrelazione, ci aspettiamo, utilizzando le bande precedenti, un $\hat{\rho}(h)$ ogni 20 “fuori” dalle bande. Ovvero, se calcoliamo i primi 30 coefficienti di autocorrelazione, trovarne uno, due o anche tre fuori dalle bande può essere attribuito all'effetto del caso. Ovviamente però ce li aspettiamo non di molto esterni alle bande stesse. Pochi $\hat{\rho}(h)$ fuori di poco dalle bande possono essere attribuiti all'errore di stima.

Una stima della funzione di autocorrelazione parziale può essere ottenuta sostituendo, nella definizione (2.8), ai coefficienti di autocorrelazione globale ρ_k , le corrispondenti stime $\hat{\rho}_k$. Esistono però altri metodi, basati su regressioni ricorsive, che vengono utilizzati dai principali software, ad esempio R .

2.2 Il test di Ljung-Box (e quello di Box-Pierce)

Una statistica test che può essere utilizzata per verificare l'ipotesi che il processo sia un white noise è

$$T_{L\&B} = n(n+2) \sum_{h=1}^H \frac{\hat{\rho}^2(h)}{n-h}, \quad (2.11)$$

dove H è un intero prescelto. Sotto l'ipotesi nulla (assenza di autocorrelazione, ovvero rumore bianco) $T_{L\&B}$ si distribuisce asintoticamente come una variabile casuale χ^2 con H gradi di libertà (ovvero $\Gamma(H/2, 1/2)$). Valori troppo grandi rispetto a quelli che ci aspettiamo da questa distribuzione sono evidenza che l'autocorrelazione "non è solo apparente".

Un test, asintoticamente analogo a quello di Ljung e Box, si basa sulla statistica test proposta e studiata da Box e Pierce

$$T_{B\&P} = n \sum_{h=1}^H \hat{\rho}^2(h). \quad (2.12)$$

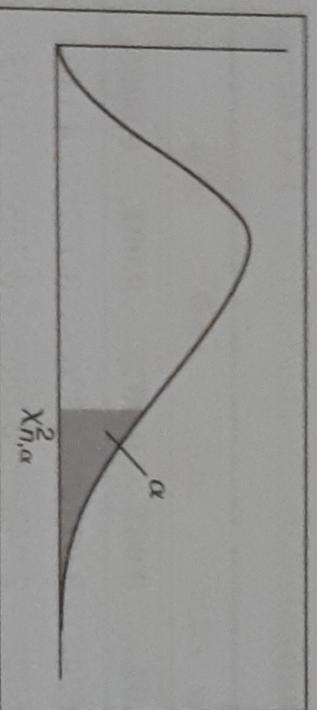
La differenza tra le due statistiche consiste semplicemente nella differente ponderazione adottata: nella prima il quadrato di $\hat{\rho}(h)$ "entra" con peso $n(n+2)/(n-h)$, mentre nella seconda con peso n . Asintoticamente sono equivalenti. Si può però mostrare che la prima statistica converge più rapidamente alla sua distribuzione asintotica. E' quindi consigliabile utilizzare $T_{L\&B}$.

Esempio 2.2 Supponiamo che $T_{L\&B}$, calcolato sulla base dei primi 20 coefficienti di autocorrelazione campionari valga circa 31.4. Una approssimazione del livello di significativà del test è quindi $Prob(\chi_{20}^2 \geq 31.41) \approx 0.05$, ovvero $Prob(\chi_{20}^2 \leq 31.41) \approx 0.95$; quindi con alta probabilità il processo può ritenersi White Noise. Le differenze da zero delle autocorrelazioni campionarie potrebbero quindi essere semplicemente dovute all'errore di stima.

Esempio 2.3 Supponiamo ora che $T_{L\&B}$, calcolato sulla base dei primi 20 coefficienti di autocorrelazione campionari valga circa 7. Una approssimazione del livello di significativà del test è quindi $Prob(\chi_{20}^2 \geq 7.43) \approx 0.995$, ovvero $Prob(\chi_{20}^2 \leq 7.43) \approx 0.05$. Allora, con probabilità piccola, il processo può essere ritenuto un white noise, dunque il processo evidenzia correlazione. Questo valore ci dice che applicando la medesima procedura a serie storiche incorrelate ci aspettiamo un valore della statistica test grande come quello osservato circa due volte ogni dieci miliardi di occasioni. Quindi, ci suggerisce che l'autocorrelazione segnalata dal grafico è reale (e non semplicemente dovuta all'errore di stima).

2.3 Spettro di un processo stazionario.

Per tempi continui, si definisce *spettro* di un processo stazionario X_t con media $E(X_t) = \mu$, la trasformata di Fourier della funzione di autocovarianza $\gamma(\tau) = E[(X(t+\tau) - \mu)(X(t) - \mu)]$,

Tavola D.3 Percentili $\chi_{n, \alpha}^2$ delle distribuzioni chi-quadrato

n	α														
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005				
1	0.00+	0.00+	0.00+	0.00+	0.02	0.45	2.71	3.84	5.02	6.63	7.88				
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60				
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84				
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86				
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75				
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55				
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28				
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96				
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59				
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19				
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76				
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30				
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82				
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32				
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80				
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27				
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72				
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16				
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58				
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00				
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40				
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80				
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18				
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56				
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93				
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29				
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65				
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99				
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34				
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67				

divisa per 2π , ovvero

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega\tau} \gamma(\tau) d\tau. \quad (2.13)$$

Poiché $\gamma(-\tau) = \gamma(\tau)$, si ha anche

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega\tau} \gamma(\tau) d\tau. \quad (2.14)$$

Antitrasformando, si ottiene:

$$\gamma(\tau) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} e^{-i\omega\tau} g(\omega) d\omega. \quad (2.15)$$

Il valore

$$\gamma(0) = \text{Var}(X_t) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} g(\omega) d\omega \quad (2.16)$$

si chiama potenza media dello spettro. E' facile vedere che $g(-\omega) = g(\omega)$, ovvero g è una funzione pari.

Per tempi discreti, l'integrale di sopra va sostituito con una sommatoria, ovvero con

$$\sum_{k=-\infty}^{+\infty} e^{i\omega k} \gamma_k = \gamma_0 + \sum_{k=-\infty}^{-1} e^{i\omega k} \gamma_k + \sum_{k=1}^{+\infty} e^{i\omega k} \gamma_k \quad (2.17)$$

La prima serie può essere riscritta come

$$\sum_{h=1}^{+\infty} e^{-i\omega h} \gamma_{-h} = \sum_{h=1}^{+\infty} e^{-i\omega h} \gamma_h.$$

visto che $\gamma_{-h} = \gamma_h$. Dunque, riprendendo il calcolo:

$$\sum_{k=-\infty}^{+\infty} e^{i\omega k} \gamma_k = \gamma_0 + \sum_{k=1}^{+\infty} (e^{i\omega k} + e^{-i\omega k}) \gamma_k = \gamma_0 + 2 \sum_{k=1}^{\infty} \cos(k\omega) \gamma_k.$$

Quindi la definizione di spettro diventa:

$$g(\omega) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(\omega k) \right] \quad (0 \leq \omega \leq \pi). \quad (2.18)$$

Per $0 \leq \omega \leq \pi$, la *densità spettrale* è la trasformata di Fourier discreta della funzione di autocorrelazione ρ_k :

$$f(\omega) = \frac{1}{2\pi} \left[\rho_0 + 2 \sum_{k=1}^{\infty} \rho_k \cos(\omega k) \right] = \frac{1}{2\pi} \left[1 + 2 \sum_{k=1}^{\infty} \frac{\gamma_k}{\gamma_0} \cos(\omega k) \right] = \frac{g(\omega)}{\gamma_0}. \quad (2.19)$$

Si può mostrare che $f(\omega)$ è non negativa $\forall \omega \in [0, \pi]$. Infatti, risulta

$$g(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} E [|F_T(\omega)|^2] \geq 0,$$

dove $F_T(\omega)$ è la trasformata di Fourier del processo troncato $X_T(t) = \mathbf{1}_{[-T, T]}(t)X(t)$. La *funzione di ripartizione spettrale* $F(\omega)$ di un processo stazionario con densità spettrale $f(\omega)$ è definita da

$$F(\omega) = \int_{-\pi}^{\omega} f(\eta) d\eta. \quad (2.20)$$

$F(\omega)$ è una funzione monotona non decrescente, limitata, con $F(-\pi) = 0$. Inoltre:

$$F(\pi) = \int_{-\pi}^{\pi} \frac{1}{2\pi} \left[1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(\omega k) \right] = 1 + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho_k \int_{-\pi}^{\pi} \cos(\omega k) d\omega = 1,$$

poiché gli integrali sono tutti nulli. Quindi $F(\omega)$ possiede tutte le proprietà di una funzione di ripartizione di una v.a. Analogamente, la densità spettrale $f(\omega)$ possiede le proprietà di una densità di probabilità, $f(\omega) = F'(\omega) \geq 0$ e $\int_{-\pi}^{\pi} f(\omega) d\omega = 1$.

La *funzione di spettro cumulato* è definita da:

$$G(\omega) = \int_{-\pi}^{\omega} g(\eta) d\eta = \int_{-\pi}^{\omega} \gamma_0 f(\eta) d\eta = \gamma_0 F(\omega). \quad (2.21)$$

Per un processo white noise di varianza σ^2 abbiamo uno spettro costante; infatti, siccome $\gamma_k = \sigma^2 \delta(k - 0)$, ovvero $\gamma_0 = \sigma^2$ e $\gamma_k = 0$ se $k \neq 0$, si ha

$$g(\omega) = \frac{\gamma_0}{2\pi} = \frac{\sigma^2}{2\pi}. \quad (2.22)$$

2.4 Filtri lineari

Dato un nprocesso stocastico X_t , non necessariamente white noise, si definisce *filtro lineare* il processo Y_t dato da:

$$Y_t = \sum_{j=-\infty}^{+\infty} \psi_j X_{t-j} \quad (2.23)$$

dove i coefficienti ψ_j sono reali e fissati, ovvero non dipendono da t . Se $\psi_j = 0$ per $j < 0$, si parla di *filtro causale*.

Se in (2.23) poniamo $-j = i$ e indichiamo ψ_{-i} ancora con ψ_i , si ottiene anche

$$Y_t = \sum_{j=-\infty}^{+\infty} \psi_j X_{t+j}. \quad (2.24)$$

Calcoliamo ora lo spettro di Y_t ; si ha:

$$g_Y(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} e^{i\omega k} \gamma_k^Y, \quad (2.25)$$

dove $i = \sqrt{-1}$ e γ_k^Y denota la covarianza di Y_t e Y_{t-k} , ovvero:

$$\gamma_k^Y = \text{cov}(Y_t, Y_{t-k}) = \text{cov} \left(\sum_{j=-\infty}^{+\infty} \psi_j X_{t-j}, \sum_{h=-\infty}^{+\infty} \psi_h X_{t-k-h} \right)$$

$$= \sum_j \sum_h \psi_j \psi_h \text{cov}(X_{t-j}, X_{t-k-h}) = \sum_h \sum_j \psi_h \psi_j \gamma_{k+h-j}^X \quad (2.26)$$

ove $\gamma_k^X = \text{cov}(X_t, X_{t-k})$. Inserendo il valore di γ_k^Y in (2.25), si ottiene:

$$\begin{aligned} g_Y(\omega) &= \frac{1}{2\pi} \sum_k e^{i\omega k} \sum_h \sum_j \psi_h \psi_j \gamma_{k+h-j}^X = \frac{1}{2\pi} \sum_k \sum_h \sum_j e^{i\omega(k-h+j)} \gamma_{k-h+j}^X e^{i\omega h} \psi_h e^{-i\omega j} \psi_j \\ &= g^X(\omega) \cdot \left[\sum_h e^{i\omega h} \psi_h \cdot \sum_j e^{-i\omega j} \psi_j \right]. \end{aligned} \quad (2.27)$$

Ma, l'ultima $\left[\right]$ vale $\left| \sum_h e^{i\omega h} \psi_h \right|^2$; infatti:

$$\begin{aligned} &\sum_h \sum_j e^{i\omega h} e^{-i\omega j} \psi_h \psi_j \\ &= \sum_h \sum_j (\cos \omega h + i \sin \omega h)(\cos \omega j - i \sin \omega j) \psi_h \psi_j \\ &= \sum_h \sum_j (\cos \omega h \cdot \cos \omega j - i \sin \omega j \cdot \cos \omega h + i \sin \omega h \cdot \cos \omega j + \sin \omega h \cdot \sin \omega j) \psi_h \psi_j \\ &= \sum_h \sum_j \psi_h \psi_j \cos \omega h \cdot \cos \omega j + \sum_h \sum_j \psi_h \psi_j \sin \omega h \cdot \sin \omega j \\ &= \left(\sum_h \psi_h \cos \omega h \right)^2 + \left(\sum_h \psi_h \sin \omega h \right)^2; \end{aligned} \quad (2.28)$$

d'altra parte:

$$\begin{aligned} \left| \sum_h e^{i\omega h} \psi_h \right|^2 &= \left| \sum_h \psi_h (\cos \omega h + i \sin \omega h) \right|^2 \\ &= \left(\sum_h \psi_h (\cos \omega h + i \sin \omega h) \right) \left(\sum_k \psi_k (\cos \omega h - i \sin \omega h) \right) \\ &= \left(\sum_h \psi_h \cos \omega h \right)^2 + \left(\sum_h \psi_h \sin \omega h \right)^2, \end{aligned}$$

che coincide con (2.28). Dunque, concludiamo che:

$$g_Y(\omega) = \left| \sum_h e^{i\omega h} \psi_h \right|^2 g^X(\omega). \quad (2.29)$$

Il termine $\sum_h e^{i\omega h} \psi_h$ è la *funzione di risposta del filtro*, il suo modulo è la *funzione guadagno*:

$$\mathcal{G}(\omega) = \left| \sum_h e^{i\omega h} \psi_h \right|, \quad (2.30)$$

mentre il quadrato del modulo è la *funzione di trasferimento*. La funzione guadagno \mathcal{G} determina il fattore di riduzione o aumento dell'ampiezza di una componente ciclica in seguito all'applicazione del filtro. Se $\mathcal{G}(\omega)$ è grande per ω piccola, ma piccola per ω grande, allora il

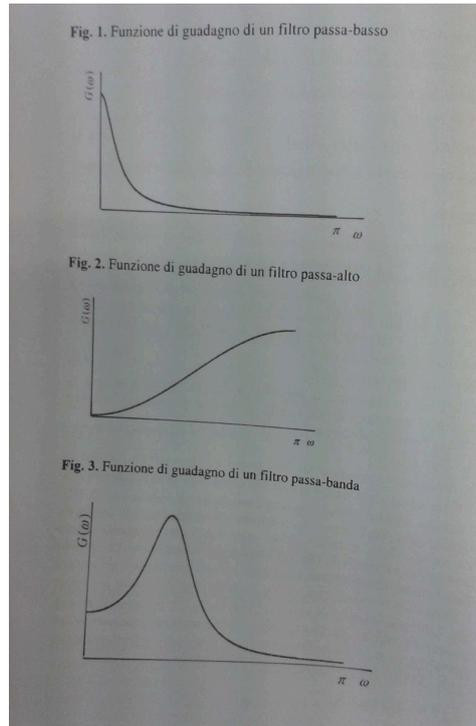


Figure 2: Funzioni di guadagno per filtri passa-basso, passa-alto e passa-banda

filtro è detto *passa-basso*. Ciò significa che se l'output è una misura di oscillazione di differenti frequenze, allora solamente quelle componenti con frequenze basse passano attraverso il filtro, mentre le frequenze alte vengono significativamente ridotte (v. Fig. 2, pannello 1).

Se $\mathcal{G}(\omega)$ è piccola per ω piccola e grande per ω grande, il filtro si dice *passa-alto* (v. Fig. 2, pannello 2). Infine, un filtro può essere costruito per far passare solo le componenti di frequenza che appartengono ad un determinato intervallo, e viene detto *filtro passa-banda* (v. Fig. 2, pannello 3).

2.5 L'operatore ritardo

L'operatore *ritardo* (backshift operator) B è definito dalla trasformazione del processo: $BX_t = X_{t-1}$.

Si ha pure $BX_{t-1} = X_{t-2}$, e dunque $B(BX_t) = B^2(X_t) = X_{t-2}$.

In generale $B^j X_t = X_{t-j}$, $j = 1, 2, \dots$ e si pone, per convenzione, $B^0 = I$ (identità).

2.6 Medie mobili

Una media mobile è una media aritmetica ponderata costituita da un numero fisso di termini e applicata in maniera sequenziale aggiungendo (ed eliminando) una osservazione alla volta. La stima ottenuta con una media mobile \hat{X}_t , relativa ad una osservazione X_t è calcolata a partire dalla osservazione X_{t-m} fino a X_{t+m} nel modo seguente:

$$\hat{X}_t = \sum_{j=-m}^m w_j X_{t+j} = w_{-m} X_{t-m} + \dots + w_0 X_t + \dots + w_m X_{t+m} \quad (2.31)$$

$E(X^{\wedge}(t)) = \sum_j w_j E(X_{t+j}) = \eta \sum_j w_j$,
ove η = media del processo stazionario $X(t)$; allora
 $E(X^{\wedge}(t))/E(X(t)) = \sum_j w_j$

Tabella 1. Esempio di smoothing di una serie storica attraverso una media mobile.

Serie storica: valori originali		Pesi della media mobile				Valori smooth
378	-0.073					
371	0.294	-0.073				
395	0.558	0.294	-0.073			378.76
413	0.294	0.558	0.294	-0.073		426.25
487	-0.073	0.294	0.558	0.294	-0.073	475.05
499		-0.073	0.294	0.558	0.294	499.55
498			-0.073	0.294	0.558	503.09
525				-0.073	0.294	
552					-0.073	

Figure 3: Esempio di smoothing di una serie storica mediante una media mobile

dove w_j sono i pesi, w_0 è il peso principale e per medie mobili simmetriche risulta $w_{-j} = w_j$; si osservi che i pesi possono essere anche negativi, ma $\sum_{j=-m}^m w_j$ deve essere 1. La lunghezza della media mobile è data dal numero di osservazioni utilizzate nel calcolo, ovvero $2m + 1$. L'operazione di applicare una media mobile a una serie storica, detta anche *smoothing*, è un caso particolare del processo più generale di filtraggio, un concetto mutuato dall'ingegneria elettrica. Le medie mobili sono filtri lineari invarianti nel tempo, i cui pesi sommano a uno. Dato che **la somma dei pesi di un filtro determina il rapporto fra la media dell'output e quella dell'input**, pesi che sommano a uno implicano che la media dell'input non viene cambiata dall'applicazione di una media mobile. Naturalmente, una media mobile non deve essere confusa con un *processo media mobile*, i cui pesi non sommano a uno, e di cui parleremo più avanti. Le medie mobili sono dette funzioni di smoothing poiché riducono l'ampiezza delle fluttuazioni cicliche di alta frequenza, senza modificare significativamente le componenti di bassa frequenza (nello spettro); esse hanno la caratteristica dei filtri *passa-basso*. Una media mobile \mathcal{F} è invariante rispetto a cambiamenti di scala se ad una amplificazione dell'input X_t per un fattore scalare α segue una stessa amplificazione dell'output, ovvero

$$\mathcal{F}(\alpha X_t) = \alpha \mathcal{F}(X_t) \quad (2.32)$$

Una applicazione di una media mobile a 5 termini è mostrata nella Tabella di Figura 2.

Per esempio, il primo valore smooth 387.76 si ottiene moltiplicando i pesi della media mobile $-0.073, 0.294, 0.558, 0.294, -0.773$ per i valori della serie originale che vanno da 378 a 487, cioè:

$$-0.073 \cdot 378 + 0.294 \cdot 371 + 0.558 \cdot 395 + 0.294 \cdot 413 - 0.773 \cdot 487 = 378.76$$

Il secondo valore smooth 426.25 si ottiene moltiplicando i pesi della media mobile per i valori che vanno da 371 a 499, e così via. Si noti che la lunghezza della media mobile è sempre fissa (cinque termini) e che essa è applicata alla serie originale in maniera sequenziale, da cui l'aggettivo mobile.

La stima dei pesi di una media mobile può essere ottenuta o adattando polinomi locali ai dati col metodo dei minimi quadrati, o attraverso formule dette “di sommatoria”.

Una importante classe di medie mobili è quella dei *kernel smoother* che per ogni osservazione X_t restituiscono la stima

$$\widehat{X}_t = \sum_{j=1}^n w_j X_{t+j}$$

dove i pesi w_j sono determinati dalla funzione kernel K , tale che $K(-x) = K(x)$, $\int_{-\infty}^{+\infty} K(x)dx = 1$, nel seguente modo;

$$w_j = \frac{K(|X_t - X_{t_j}|/b)}{\sum_{i=1}^n K(|X_t - X_{t_i}|/b)}, \quad b > 0.$$

Il parametro di smoothing b determina il numero di osservazioni X_{t_j} vicine a X_t che hanno peso w_j diverso da zero. Il kernel gaussiano, definito da $K(x) = e^{-x^2/2}/\sqrt{2\pi}$ è ampiamente utilizzato. Altri kernel noti in letteratura sono il kernel triangolare, a scatola, e il kernel uniforme.

2.7 Processi stocastici lineari

Si chiama processo lineare con pesi ψ_j , $j = 1, \dots$, dove $\psi_0 = 1$ e $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$, un processo definito da:

$$X_t = \eta + u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \dots = \eta + u_t + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \quad (2.33)$$

o anche, se si pone $\psi_0 = 1$:

$$X_t = \eta + \sum_{j=0}^{\infty} \psi_j u_{t-j},$$

dove η è la media del processo, u_t è un white noise. Si può dimostrare (si veda ad esempio [Box, Jenkins] o [Bee Dagum]) che un processo lineare della forma (2.33) è stazionario se la serie (funzione generatrice dei pesi ψ_j):

$$\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j \quad (2.34)$$

converge per $|B| \leq 1$.

Si può dimostrare (Teorema di Wold) che un processo stocastico debolmente stazionario, può essere sempre rappresentato nella forma (2.33). Questa rappresentazione va intesa in media quadratica, cioè

$$E \left[\left(X_t - \eta - \sum_{j=0}^n \psi_j u_{t-j} \right)^2 \right] \rightarrow 0, \quad \text{per } n \rightarrow \infty.$$

La condizione $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$ garantisce la stazionarietà del processo. Infatti

$$E(X_t) = \eta + \lim_{n \rightarrow \infty} E \left[\sum_{j=0}^n \psi_j u_{t-j} \right] = \eta + \lim_{n \rightarrow \infty} \sum_{j=0}^n \psi_j E[u_{t-j}] = \eta + 0 = \eta,$$

$$\text{Var}(X_t) = \lim_{n \rightarrow \infty} \sum_{j=0}^n \psi_j^2 E[(u_{t-j})^2] = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

Inoltre

$$\begin{aligned} \gamma(h) &= E[(X_t - E(X_t))(X_{t-h} - E(X_{t-h}))] \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \sum_{j=0}^n \sum_{i=0}^m \psi_j \psi_i E[u_{t-j} u_{t-h-i}] = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}, \end{aligned}$$

poiché $E[u_r u_s] = \sigma^2 \delta_{rs}$.

Di conseguenza la funzione di autocorrelazione risulta:

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+h}}{\sum_{i=0}^{\infty} \psi_i^2}$$

Chiaramente le funzioni di autocovarianza e autocorrelazione dipendono solo dal ritardo h . Comunque, dal momento che coinvolgono somme infinite, per la stazionarietà occorre mostrare che $\gamma(h)$ è finita per ogni h :

$$|\gamma(h)| = |E[(X_t - \eta)(X_{t+h} - \eta)]| \leq [\text{Var}(X_t) \text{Var}(X_{t+h})]^{1/2} = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$$

(abbiamo usato il fatto che, se U e V sono v.a. con varianza finita, allora $|E(UV)| \leq \sqrt{E(U^2)} \sqrt{E(V^2)}$).

Utilizzando l'operatore ritardo B , e ponendo $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, la (2.33) si può scrivere nella forma:

$$X_t = \eta + \psi(B)u_t$$

Un'altra utile rappresentazione per il processo X_t è nota come rappresentazione autoregressiva (AR):

$$X_t = \eta + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + u_t,$$

dove u_t è un rumore bianco. L'equazione precedente si può riscrivere nella forma

$$\pi(B)X_t = u_t + \eta \tag{2.35}$$

dove

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

con

$$1 + \sum_{j=1}^{\infty} |\pi_j| < +\infty.$$

Secondo la definizione di Box & Jenkins un processo si dice invertibile se si può scrivere nella forma (2.35).

Andiamo ora a considerare modelli finiti.

3 Processi media mobile di ordine q

Un processo a media mobile di ordine q (moving average, MA(q)) è un modello del tipo (2.33) ridotto al finito. Si tratta di un modello dinamico lineare che genera un processo stazionario; la sua caratteristica principale è quella di rappresentare/approssimare la struttura di autocorrelazione di un qualsiasi processo stazionario. La serie osservata y_t è generata a partire da un white noise u_t (ovvero un processo completamente incorrelato a tutti i ritardi, di media nulla e varianza costante σ^2), nel seguente modo:

$$y_t = \eta + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}, \quad (3.1)$$

dove $(\eta, \theta_1, \dots, \theta_q)$ è un vettore di parametri costanti.

E' immediato verificare che:

$$E(y_t) = \eta \quad (3.2)$$

e che, per $0 \leq h \leq q$:

$$\text{cov}(y_t, y_{t-h}) = \begin{cases} \sigma^2(\theta_h + \theta_1\theta_{h+1} + \cdots + \theta_{q-h}\theta_q), & \text{se } h \leq q \\ 0, & \text{se } h > q \end{cases} \quad (3.3)$$

Poiché la media è costante e $\text{cov}(y_t, y_{t-h})$ dipende solo da h , il processo è stazionario, almeno del secondo ordine.

Per un processo MA(1), si ha $y_t = \eta + u_t + \theta u_{t-1} = \eta + (1 + \theta B)u_t$ e quindi

$$\gamma_0 = \text{cov}(y_t, y_t) = \text{Var}(y_t) = (\theta_0^2 + \theta_1^2)\sigma^2 = (1 + \theta^2)\sigma^2 \quad (\theta_0 \equiv 1, \theta_1 \equiv \theta)$$

$$\gamma_1 = \text{cov}(y_t, y_{t-1}) = \theta_1\sigma^2 = \theta\sigma^2$$

$$\gamma_k = 0, \text{ se } k > q = 1$$

Pertanto, se $\rho_k = \gamma_k/\gamma_0$, si ha

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{\theta}{1+\theta^2}, & k = 1 \\ 0, & k > 1 \end{cases} \quad (3.4)$$

Lo spettro risulta

$$g(\omega) = \frac{1}{2\pi}[\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(\omega k)] = \frac{1}{2\pi}[(1+\theta^2)\sigma^2 + 2\theta\sigma^2 \cos(\omega)] = \frac{\sigma^2}{2\pi}(1+\theta^2 + 2\theta \cos \omega). \quad (3.5)$$

La densità spettrale, per $\omega \in [0, \pi]$, è:

$$f(\omega) = \frac{g(\omega)}{\gamma_0} = \frac{1}{2\pi} \left(1 + 2 \frac{\theta \cos \omega}{1 + \theta^2} \right) = \frac{g(\omega)}{\gamma_0}. \quad (3.6)$$

La densità spettrale $f(\omega)$ è crescente per $\theta < 0$, decrescente per $\theta > 0$.

Per un processo MA(2), si ha $y_t = \eta + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} = \eta + (1 + \theta_1 B + \theta_2 B^2)u_t$, $\gamma(h) = \sigma^2(\theta_h + \theta_1\theta_{h+1} + \cdots + \theta_{2-h}\theta_2)$, e quindi

$$\gamma_0 = \text{cov}(y_t, y_t) = \text{Var}(y_t) = (1 + \theta_1^2 + \theta_2^2)\sigma^2$$

$$\begin{aligned}\gamma_1 &= \text{cov}(y_t, y_{t-1}) = \sigma^2(\theta_1 + \theta_1\theta_2) = \sigma^2\theta_1(1 + \theta_2) \\ \gamma_2 &= \sigma^2\theta_2 \\ \gamma_k &= 0, \text{ se } k > q = 2\end{aligned}$$

Pertanto:

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2}, & k = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases} \quad (3.7)$$

La densità spettrale, per $\omega \in [0, \pi]$, è:

$$f(\omega) = \frac{1}{2\pi} \left[1 + 2 \sum_{k=1}^2 \rho_k \cos(\omega k) \right] = \frac{1}{2\pi} \left[1 + 2 \frac{\theta_1(1+\theta_2)}{(1+\theta_1^2+\theta_2^2)} \cos \omega + 2 \frac{\theta_2}{1+\theta_1^2+\theta_2^2} \cos(2\omega) \right]. \quad (3.8)$$

3.1 Invertibilità di un modello MA(q)

Un modello MA(q) è detto *invertibile* se, per ogni t , u_t è calcolabile a partire da y_t, y_{t-1}, \dots . Ciò non è sempre possibile. Ad esempio, consideriamo un modello MA(1) con media nulla, ovvero $y_t = u_t + \theta u_{t-1}$. Mediante sostituzioni successive, si trova, per m fissato:

$$\begin{aligned}u_t &= y_t - \theta u_{t-1} \\ &= y_t - \theta y_{t-1} + \theta^2 u_{t-2} \\ &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 u_{t-3} \\ &= y_t - \theta y_{t-1} + \dots + (-\theta)^m y_{t-m} + (-\theta)^{m+1} u_{t-m-1} \\ &= y_t + \sum_{i=1}^m (-\theta)^i y_{t-i} + (-\theta)^{m+1} u_{t-m-1}.\end{aligned}$$

Se $|\theta| < 1$, il termine $(-\theta)^{m+1} u_{t-m-1} \rightarrow 0$ per $m \rightarrow +\infty$, quindi, se $|\theta| < 1$ e m è “grande”, si ottiene

$$u_t \simeq y_t + \sum_{i=1}^m (-\theta)^i y_{t-i} \simeq y_t + \sum_{i=1}^{\infty} (-\theta)^i y_{t-i} = \sum_{i=0}^{\infty} (-\theta)^i y_{t-i}.$$

Dunque, un modello MA(1) è invertibile se e solo se $|\theta| < 1$. In generale, si ha:

Proposition 3.1 *Condizione sufficiente per l'invertibilità di un modello MA(q) è che le q soluzioni (che possono essere anche complesse) dell'equazione caratteristica:*

$$1 + \theta_1 x + \dots + \theta_q x^q = 0 \quad (3.9)$$

siano in modulo maggiori di 1 (per $q = 1$, (3.9) diventa $|-1/\theta| > 1$, ovvero $|\theta| < 1$, come abbiamo già trovato).

Prima di dimostrare la Proposizione 3.1, occorre fare una considerazione preliminare.

Osservazione

Per prima cosa, affinché un processo $MA(\infty)$, $y_t = \eta + \sum_{j=0}^{\infty} \theta_j u_{t-j}$, sia stazionario occorre che la funzione generatrice dei pesi (2.34) converga, e questo accade se $|B| \leq 1$. Per q finito, consideriamo un processo $MA(q)$, mantenendo la condizione $|B| \leq 1$, ovvero $y_t = \eta + \theta(B)u_t$, dove

$$\theta(B)u_t = \sum_{j=0}^q B^j u_t = u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q};$$

affinché y_t sia invertibile occorre che si possa scrivere: $u_t = (\theta(B))^{-1}(y_t - \eta)$, dove $\theta^{-1}(B) = (1 + \theta_1 B + \cdots + \theta_q B^q)^{-1}$. Ora, l'operatore $\theta(B)$ è un polinomio in B di ordine q che può essere fattorizzato nel seguente modo:

$$\theta(B) = (1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_q B) = \prod_{i=1}^q (1 - \lambda_i B). \quad (3.10)$$

Quindi

$$(\theta(B))^{-1} = \prod_{i=1}^q (1 - \lambda_i B)^{-1}.$$

Affinché esista l'inverso di ogni fattore $(1 - \lambda_i B)$, deve essere $|\lambda_i B| < 1$ e, visto che deve essere $|B| \leq 1$, ciò è equivalente a $|\lambda_i| < 1$. Infatti, sotto tale condizione la serie $\sum_{k=0}^{\infty} \lambda_i^k B^k$ converge e, come è facile verificare, si ha:

$$\left(\sum_{k=0}^{\infty} \lambda_i^k B^k \right) (1 - \lambda_i B) = I,$$

ovvero:

$$(1 - \lambda_i B)^{-1} = \sum_{k=0}^{\infty} \lambda_i^k B^k.$$

Dimostrazione della Proposizione 3.1 La fattorizzazione (3.10) consiste nel trovare i valori di $\lambda_1, \lambda_2, \dots, \lambda_q$ in modo che risulti

$$1 + \theta_1 B + \cdots + \theta_q B^q = (1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_q B)$$

Moltiplicando entrambi i membri per B^q e ponendo $\lambda = B^{-1}$, si ottiene:

$$\lambda^{-q} + \theta_1 \lambda^{-(q+1)} + \cdots + \theta_q \lambda^{-2q} = (1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_q B) \lambda^{-q}$$

ovvero

$$1 + \theta_1 \lambda^{-1} + \cdots + \theta_q \lambda^{-q} = (1 - \lambda_1/\lambda)(1 - \lambda_2/\lambda) \cdots (1 - \lambda_q/\lambda) \quad (3.11)$$

Il secondo membro di (3.11) si annulla per $\lambda = \lambda_i$, dove i può assumere ciascuno dei valori da 1 a q ; i valori λ_i annullano quindi anche il membro sinistro, ovvero, per $\lambda = \lambda_i$ si ha:

$$1 + \theta_1 \lambda^{-1} + \cdots + \theta_q \lambda^{-q} = 0.$$

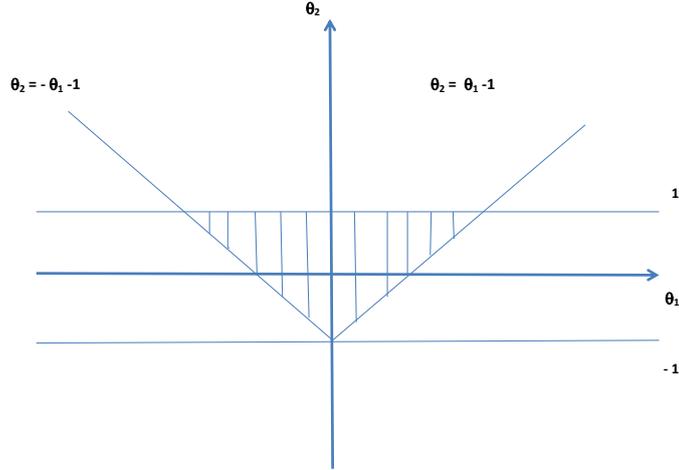


Figure 4: Valori di θ_1 e θ_2 che soddisfano la (3.12) nel piano $\theta_1\theta_2$.

Infine, dire che i λ_i hanno modulo minore di 1, significa dire che le radici di $1 + \theta_1 B + \dots + \theta_q B^q$ sono esterne al cerchio unitario, visto che si è posto $\lambda = B^{-1}$.

□

Per esempio, per un processo MA(1), la condizione di invertibilità è $|\theta| < 1$, come abbiamo già detto; per un processo MA(2), l'equazione caratteristica $1 + \theta_1 x + \theta_2 x^2 = 0$ ha soluzioni x di modulo maggiore di 1 se i valori dei parametri θ_1 e θ_2 soddisfano le seguenti relazioni (si veda ad es. [Dagum]; si osservi che in [Dagum] un processo MA(q) è definito come $y_t = \eta + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$, quindi, rispetto alla nostra definizione, occorre sostituire θ_i con $-\theta_i$):

$$\begin{cases} \theta_1 + \theta_2 > -1 \\ \theta_1 - \theta_2 < 1 \\ |\theta_2| < 1 \end{cases} \quad (3.12)$$

I valori di θ_1 e θ_2 che soddisfano la (3.12) nel piano $\theta_1\theta_2$ sono evidenziati nella parte col tratteggio della figura 2. Nel caso che un modello MA(q) sia invertibile è possibile mostrare

che esiste una successione λ_n di pesi per cui

$$u_t = y_t - \eta - \sum_{i=1}^{\infty} \lambda_i y_{t-i}, \quad (\eta = E(y_t)) \quad (3.13)$$

dove l'uguaglianza va intesa nel senso della convergenza in media quadratica, ovvero;

$$\lim_{m \rightarrow \infty} E \left[u_t - \left(y_t - \sum_{i=1}^m \lambda_i y_{t-i} \right) \right]^2 = 0$$

I pesi convergono a zero e in particolare $\sum_{i=1}^{\infty} \lambda_i^2 < \infty$. Quindi, si può anche scrivere

$$y_t = \eta + \sum_{i=1}^{\infty} \lambda_i y_{t-i} + u_t \quad (3.14)$$

ovvero, un modello MA(q) invertibile può essere visto come una specie di modello di regressione infinito in cui le variabili esplicative sono il “passato” di y_t . Gli infiniti coefficienti di regressione λ_i non sono però completamente liberi, poiché sono funzioni dei q coefficienti $\theta_1, \dots, \theta_q$.

4 Modelli autoregressivi di ordine p

Un processo stocastico y_t è detto generato da un modello autoregressivo di ordine p (“acronimo” AR(p)) se

$$y_t = \eta + \phi_1(y_{t-1} - \eta) + \dots + \phi_p(y_{t-p} - \eta) + u_t, \quad (4.1)$$

dove $(\eta, \phi_1, \dots, \phi_p)$ è un vettore di parametri costanti. Come si può vedere si tratta di un normale modello di regressione lineare in cui la variabile risposta è il valore presente del processo mentre le variabili esplicative sono i valori passati del processo stesso. E’ possibile dimostrare che il processo generato da un processo AR(p) è stazionario se e solo se le radici dell’equazione

$$1 - \phi_1 x - \dots - \phi_p x^p = 0 \quad (4.2)$$

sono in modulo maggiori di uno. In questo caso, η coincide con la media di y_t ed è possibile far vedere che il processo è rappresentabile come una “media mobile infinita” di white noise, ovvero che esistono dei pesi ψ_1, ψ_2, \dots tali che

$$y_t = \eta + u_t + \sum_{i=1}^{\infty} \psi_i u_{t-i} \quad (4.3)$$

E’ possibile far vedere che la funzione di autocorrelazione di un processo AR(1) soddisfa l’equazione alle differenze

$$\rho(h) = \phi_1 \rho(h-1) \quad (4.4)$$

con condizione iniziale $\rho(0) = 1$, e quindi che

$$\rho(h) = \phi_1^h. \quad (4.5)$$

4.1 La funzione di autocorrelazione parziale di un AR(p)

In un modello AR(1) l’autocorrelazione ai ritardi superiori al primo può essere diversa da zero ed essere anche non banale. In realtà, si sa che, in un certo senso, questa correlazione è spuria. Ad esempio la correlazione che troviamo al ritardo 2 è legata al fatto che y_{t-1} è generato a partire da y_{t-2} e y_t a partire da y_{t-1} . Quindi, la correlazione tra y_t e y_{t-2} è diversa da zero ma tutta “mediata” da y_{t-1} . Per un processo AR(2), invece, una parte della correlazione al ritardo 2 è “mediata” da y_{t-1} , mentre una parte esprime l’impatto “diretto” di y_{t-2} su y_t . Per isolare la correlazione “diretta” dalla correlazione “mediata” è possibile utilizzare i coefficienti di *autocorrelazione parziale*

$$\pi(h) = \text{corr}(y_t, y_{t-h} | y_{t-1}, \dots, y_{t-h+1}), \quad h = 1, 2, \dots \quad (4.6)$$

che misura la correlazione esistente tra y_t e y_{t-h} quando da ambedue le variabili venga eliminato la parte “spiegabile linearmente” da $y_{t-1}, \dots, y_{t-h+1}$. Si osservi che $\pi(h)$ è stata indicata con ϕ_{hh} nella sezione 2 (formula (2.6)).

Per un processo AR(p) si ha $\pi(h) = 0$, se $h > p$. Per esempio, per un AR(1) si ha $\rho_h = (\phi_1)^h$ e (vedi pag. 8, 9):

$$\begin{aligned}\pi_1 &= \phi_{11} = \rho_1 = \phi_1; \\ \pi_2 &= \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi_1^2 - \phi_1^2}{1 - \rho_1^2} = 0; \\ \pi_3 &= \phi_{33} = \frac{\rho_3(1 - \rho_1^2) + \rho_1(\rho_1^2 + \rho_2^2 - 2\rho_2)}{(1 - \rho_2)(1 + \rho_2 - 2\rho_1^2)} = \frac{\phi_1^3(1 - \phi_1^2) + \phi_1(\phi_1^2 + \phi_1^4 - 2\phi_1^2)}{(1 - \rho_2)(1 + \rho_2 - 2\rho_1^2)} = 0.\end{aligned}$$

In esempi pratici si può osservare come la funzione di autocorrelazione parziale di un processo MA(q) approssimi zero per h grande, ma senza diventare esattamente zero. Ed inoltre, come la funzione di autocorrelazione parziale di un modello MA(p) possa, a parte il segno, mostrare gli andamenti qualitativi della funzione di autocorrelazione di un AR(p) e viceversa.

4.2 Modelli autoregressivi a media mobile ARMA(p, q)

Un processo stocastico y_t si dice generato da un modello autoregressivo a media mobile di ordine (p, q) (abbreviato in ARMA(p, q)) quando è generato dall'equazione alle differenze

$$y_t = \eta + \phi_1(y_{t-1} - \eta) + \dots + \phi_p(y_{t-p} - \eta) + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q} \quad (4.7)$$

I modelli ARMA sono quindi una combinazione dei modelli MA e dei modelli AR.

Un processo ARMA è invertibile se tutte le radici dell'equazione analoga alla (3.9), ovvero $\theta(B) = 0$, sono in modulo maggiore di 1. In questo caso gode anche di una rappresentazione AR(∞) del tipo della (3.14).

Un processo ARMA è stazionario se tutte le radici dell'equazione analoga alla (4.2), ovvero $\phi(B) = 0$, sono in modulo maggiore di 1. In questo caso gode anche di una rappresentazione MA(∞) del tipo della (4.3). La funzione di autocorrelazione può avere andamenti molto diversi. La regola generale è che “i primi p coefficienti sono sostanzialmente arbitrari; dopo, la funzione di autocorrelazione converge verso lo zero come fa, a partire da zero, la funzione di autocorrelazione di un AR(p)”. La funzione di autocorrelazione parziale si comporta nella stessa maniera sostituendo però q a p e MA(q) a AR(p). La convergenza verso lo zero avviene in accordo all'equazione $\rho(h) = \phi_1 \rho(h-1)$ che caratterizza un modello AR(1), ma iniziando in questo caso da $h = q = 2$.

Utilizzando l'operatore ritardo B , possiamo riscrivere l'equazione (4.7) come

$$\phi(B)(y_t - \eta) = \theta(B)u_t \quad (4.8)$$

dove

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (4.9)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q. \quad (4.10)$$

Infatti, si ha:

$$y_t - \eta - \phi_1(y_{t-1} - \eta) - \dots - \phi_p(y_{t-p} - \eta) = u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q}$$

ovvero

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(y_t - \eta) = (1 + \theta_1 B + \theta_2 B + \dots + \theta_q B^q)u_t.$$

Equivalentemente, si può scrivere

$$y_t = \eta + (\phi(B))^{-1}\theta(B)u_t, \quad (4.11)$$

se $(\phi(B))^{-1}$ esiste.

4.3 Modelli ARIMA (p,d,q)

Sono modelli ARMA integrati con *trend*, e servono ad approssimare processi non stazionari. Cominciamo col considerare un processo deterministico $y_t = \zeta + \eta t$ (ζ, η costanti), la cui realizzazione coincide con una retta. Ovviamente, $(I - B)y_t = y_t - y_{t-1} = \eta$, ovvero le differenze successive di una serie costituita solo da un trend lineare sono costanti.

Che cosa ci aspettiamo di ottenere, se nella equazione precedente, sostituiamo a η un processo stocastico, ad esempio un ARMA(p, q) di media η ? Ovvero, quali caratteristiche avranno le traiettorie di un processo generato da una equazione alle differenze del tipo

$$y_t = y_{t-1} + z_t$$

dove z_t indica un processo stazionario di media η generato da un modello ARMA? La risposta è facile. Mediamente, y_t dovrebbe aumentare di un valore η per ogni unità di tempo. Quindi ci aspettiamo che y_t esibisca un trend lineare.

Considerazioni empiriche suggeriscono che un modello del tipo:

$$(y_t - y_{t-1})^d = ((1 - B)y_t)^d = z_t \quad (4.12)$$

dove $d > 0$ e z_t è un modello ARMA, con

$$\phi(B)(z_t - \eta) = \theta(B)u_t, \quad (4.13)$$

è adatto ad approssimare il meccanismo generatore di una serie temporale *non stazionaria*. In tal caso, si parla di modello *autoregressivo a media mobile integrato* di ordine (p, d, q) (abbreviazione ARIMA (p, d, q)).

Se $(I - B)y_t = z_t$, ovvero $y_t - y_{t-1} = z_t$, visto che $y_t = y_0 + (y_1 - y_0) + (y_2 - y_1) + \dots + (y_t - y_{t-1})$, si ottiene $y_t = y_0 + \sum_{i=1}^t z_i$; cosicché l'aggettivo "integrato" discende dal fatto che la "somma" è l'analogo a tempo discreto di un integrale.

Si osservi che possiamo non menzionare esplicitamente z_t nella definizione di un modello ARIMA(p, d, q). Infatti, applicando l'operatore $\phi(B)$ ad ambedue i termini dell'equazione (4.12), otteniamo

$$\phi(B)(1 - B)^d y_t = \phi(B)z_t \quad (4.14)$$

D'altra parte, $\phi(B)(z_t - \eta) = \phi(B)z_t - \phi(B)\eta$, per (4.13) è uguale a $\theta(B)u_t$, e quindi $\phi(B)z_t = \theta(B)u_t + \phi(B)\eta$. Posto

$$\delta = \phi(B)\eta = (1 - \phi_1 B - \dots - \phi_p B^p)\eta = \eta - \phi_1 B(\eta) - \dots - \phi_p B^p(\eta),$$

siccome $B(\eta) = \eta$, si ottiene

$$\delta = \eta - \phi_1\eta - \cdots - \phi_p\eta = (1 - \phi_1 - \cdots - \phi_p)\eta;$$

sostituendo in (4.14) otteniamo infine

$$\phi(B)((1 - B)y_t)^d = \delta + \theta(B)u_t. \quad (4.15)$$

Osservare che, talvolta in letteratura viene scritto

$$\phi(B)(1 - B)^d y_t = \delta + \theta(B)u_t,$$

intendendo con $(1 - B)^d y_t$ la quantità $((1 - B)y_t)^d$.

RIASSUMENDO

Se

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \\ \eta &= E(y_t) \text{ e } u_t \text{ denota un rumore bianco :}\end{aligned}$$

1. MA (q) è definito da:

$$\theta(B)u_t = y_t - \eta$$

2. AR(p) è definito da:

$$\phi(B)(y_t - \eta) = u_t$$

3. ARMA (p,q) è definito da:

$$\phi(B)(y_t - \eta) = \theta(B)u_t$$

4. ARIMA (p,d,q) è definito da:

$$\phi(B)(z_t - \eta) = \theta(B)u_t, \quad z_t = ((1 - B)y_t)^d$$

oppure

$$\phi(B)((1 - B)y_t)^d = \delta + \theta(B)u_t$$

con $\delta = (1 - \phi_1 - \dots - \phi_p)\eta$.

Osserviamo che *ARIMA*(0,0,0) è un white noise con l'aggiunta di δ , ovvero $y_t = \delta + u_t$, ed ha media $\eta = \delta$ diversa da zero.

Inoltre:

1'. un processo MA (q) può essere visto come

$$(\theta(B))^{-1}(y_t - \eta) = u_t, \text{ ovvero un } AR(\infty)$$

2'. un processo AR (p) può essere visto come

$$y_t - \eta = (\phi(B))^{-1}u_t, \text{ ovvero un } MA(\infty)$$

(cfr. equazione (4.3).

5. La funzione di autocorrelazione globale di un $MA(q)$ si annulla per ritardi superiori a q e, data l'equivalenza tra un processo $MA(q)$ e un processo $AR(\infty)$, la funzione di autocorrelazione parziale non si annulla mai, ma decade verso zero, è dominata da funzioni esponenziali smorzate e/o da funzioni cosinusoidali smorzate, a seconda della natura dell'equazione caratteristica. Il viceversa vale per un processo $AR(p)$.

6. Un processo $AR(p)$ è sempre invertibile, ma affinché sia stazionario, le radici del polinomio $\phi(B)$ devono stare all'esterno del cerchio unitario nel piano complesso. Al contrario, un processo $MA(q)$ non richiede alcuna condizione per la stazionarietà, mentre per garantire l'invertibilità occorre che le radici del polinomio $\theta(b)$ siano all'esterno del cerchio unitario.

5 BIBLIOGRAFIA

Per redarre questi appunti si è attinto dalle seguenti fonti:

G.E.P. Box, G.M. Jenkins: *Time series Analysis: forecasting and control*. Holden –Day, California, 1976.

E. Bee Dagum: *Analisi delle serie storiche* Springer-Verlag Italia, Milano, 2002.

A. Lijoi: *Corso di Statistica per le applicazioni Economiche*, Università di Pavia (per l' esercitazione con **R** della sezione 5).

G. Massarotto: *Lucidi delle Lezioni di “Analisi delle serie temporali”*, Facoltà di Scienze Statistiche, Università di Padova.

6 Esercitazione con R

Esercitazione con R

A. Lijoi, Università di Pavia

Corso di Statistica per le Applicazioni Economiche

In questa esercitazione lavoreremo con serie storiche generate da modelli ARMA. Introduciamo, quindi, il trend ed esamineremo metodi per la stima e l'eliminazione del trend.

Simulazione di processi ARMA

Sia $\{X_t\}$ un processo ARMA(p, q), ovvero

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

In R è possibile simulare le realizzazioni di un processo ARMA sfruttando il comando `arma.sim()`. Questo, come suggerisce il nome, consente di simulare più in generale processi ARIMA.

Iniziamo con il considerare modelli molto semplici. Se si vuole simulare un processo white-noise $\{\varepsilon_t\}$ si invia l'istruzione

```
eps <- arma.sim(list(order=c(0,0,0)), n=200)
```

In questo caso, gli argomenti di `arma.sim()` sono

- **order** che consente di specificare l'ordine del processo ARIMA(p, d, q). Nel caso specifico in esame si è interessati alla simulazione di un processo white-noise e si pone, di conseguenza, $p = d = q = 0$;
- **n** che individua il numero di realizzazioni (o di valori simulati) che si richiede a R di generare.

L'istruzione

```
> plot(eps, ylab=" ", xlab="t", main="Simulazione di un processo white-noise")
```

genera il grafico della serie storica `eps` rappresentato nella Figura 1. Come si può rilevare la serie oscilla attorno a $E(\varepsilon_t) = 0$ e le oscillazioni sono di ampiezza pressoché costante. Quest'ultima circostanza può essere spiegata con il fatto che $\text{Var}(\varepsilon_t) = \sigma^2$ ed è, quindi, costante nel tempo. Le funzioni di autocorrelazione e di autocorrelazione parziale (campionarie) sono rappresentate in Figura 2.

I comandi usati per generare i grafici delle stime delle funzioni di autocorrelazione e di autocorrelazione parziale sono i seguenti:

```
> par(mfrow=c(2,1))
> acf(eps, lag.max=20, main="ACF campionaria (h=0,1,...,20)")
> pacf(eps, lag.max=20, main="PACF campionaria (h=0,1,...,20)")
```

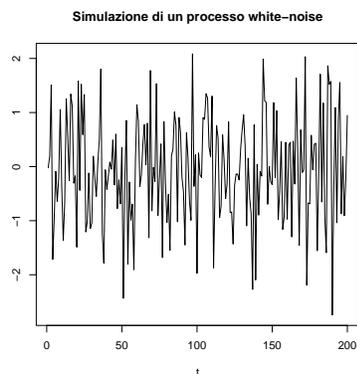


Figure 1: Processo white-noise simulato con la funzione `arma.sim`

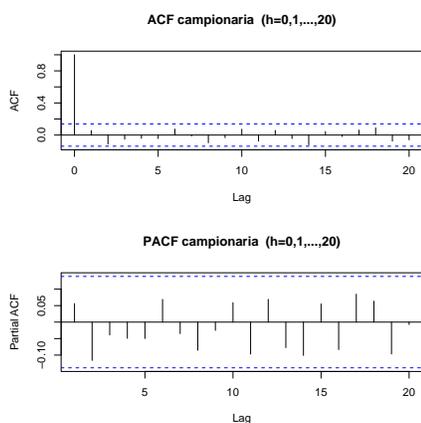


Figure 2: Funzione di autocorrelazione e autocorrelazione parziale stimate in corrispondenza delle 200 realizzazioni simulate del processo white-noise $\{\varepsilon_t\}$

Il comando `par(mfrow=c(p,q))` serve per indicare a R che si intende creare una tabella di grafici in un'unica finestra. La tabella contiene `p` grafici per colonna e `q` grafici per riga. In questo caso si è scelto di far apparire il grafico della ACF sopra il grafico della PACF: quindi `p=2` e `q=1`.

La funzione `acf()` contiene i seguenti argomenti

- Occorre, anzitutto, specificare la serie $\{x_1, x_2, \dots, x_n\}$ di cui si vuole determinare la funzione di autocorrelazione campionaria

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h} (x_{t+h} - \bar{x}_n)(x_t - \bar{x}_n)}{\sum_{t=1}^n (x_t - \bar{x}_n)^2}$$

In questo caso, la serie cui si è interessati coincide con i 200 valori simulati del processo white-noise che abbiamo denominato `eps`.

- `lag.max` indica il numero di ritardi h in corrispondenza dei quali si vuole far calcolare la funzione $\hat{\rho}(h)$. La specificazione usata nell'esempio, `lag.max=20`, equivale a chiedere a R di calcolare, e di rappresentare graficamente, $\hat{\rho}(1)$, $\hat{\rho}(2)$, ... e $\hat{\rho}(20)$. Si noti che $\hat{\rho}(0) = 1$ per definizione.

Gli altri argomenti che sono stati specificati in `acf()` possono essere utili per attribuire etichette agli assi o imporre un titolo principale al grafico (come nel caso della Figura 2). Considerazioni analoghe valgono per l'uso della funzione `pacf`.

Si osservi che nei grafici sopra sono anche presenti due segmenti orizzontali tratteggiati. Questi corrispondono alla soglia della regione di rifiuto per il test

$$H_0 : \rho(h) = 0 \quad \text{vs} \quad H_1 : \rho(h) \neq 0$$

ad un livello $\alpha = 0.05$. In base a risultati asintotici noti per la funzione di autocorrelazione campionaria, se è vera H_0 e n è sufficientemente elevato,

$$\hat{\rho}(h) \approx N\left(0; \frac{1}{\sqrt{n}}\right).$$

una regione di rifiuto, di livello $\alpha = 0.05$, per verificare H_0 contro H_1 è

$$\mathcal{R} = \left\{ (x_1, \dots, x_n) : |\hat{\rho}(h)| > \frac{1.96}{\sqrt{n}} \right\}.$$

Pertanto le due linee tratteggiate in blu nel grafico della Figura 2 corrispondono alle rette $y = 1.96/\sqrt{200} \approx 0.1386$ e $y = -1.96/\sqrt{200} \approx -0.1386$. Nel caso specifico in esame, tutti i picchi della funzione di autocorrelazione campionaria non sfiorano le soglie descritte e sono tali da indurre ad accettare $H_0 : \rho(h) = 0$.

Procediamo, ora, a simulare $n = 200$ realizzazioni del processo AR(1)

$$X_t = 0.2 X_{t-1} + \varepsilon_t$$

e dei processi MA(1)

$$Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$

$$Z_t = \varepsilon_t - 0.5\varepsilon_{t-1}$$

La sequenza di istruzioni da inviare a R è

```
> x <- arima.sim(list(order=c(1,0,0), ar=0.2), n=200)
> y <- arima.sim(list(order=c(0,0,1), ma=0.5), n=200)
> z <- arima.sim(list(order=c(0,0,1), ma=-0.5), n=200)
```

I grafici delle serie simulate sono generate mediante

```
> par(mfrow=c(3,1))
> plot(x, type="l", main="Processo AR(1)")
```

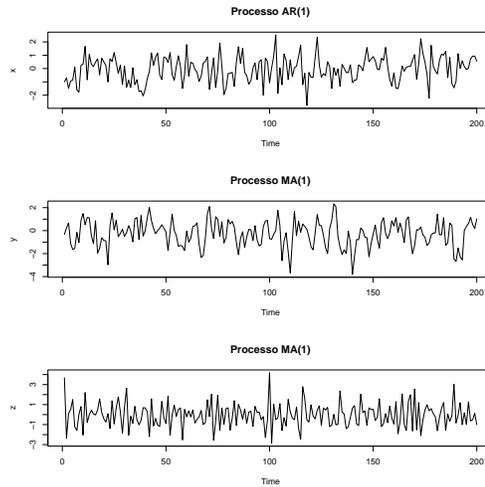


Figure 3: Realizzazioni di un processo AR(1) e di due processi MA(1).

```
> plot(y, type="l", main="Processo MA(1)")
> plot(z, type="l", main="Processo MA(1)")
```

e sono rappresentate nei grafici in Figura 3. Si noti il diverso comportamento delle serie $\{Y_t\}$ e $\{Z_t\}$, che differiscono solo per il segno del coefficiente associato al termine del processo white-noise ε_{t-1} . In particolare, le oscillazioni della serie $\{Z_t\}$ sono molto iù frequenti e una spiegazione del diverso comportamento delle due serie può essere spiegato sulla base dell funzione di autocorrelazione. Se ρ_Y è la funzione di autocorrelazione di $\{Y_t\}$ e ρ_Z è la funzione di autocorrelazione di $\{Z_t\}$ risulta

$$\rho_Y(h) = \begin{cases} 1 & h = 0 \\ \mathbf{0.4} & h = 1 \\ 0 & h = 2, 3, \dots \end{cases} \quad \rho_Z(h) = \begin{cases} 1 & h = 0 \\ \mathbf{-0.4} & h = 1 \\ 0 & h = 2, 3, \dots \end{cases}$$

La correlazione negativa tra osservazioni ritardate di un periodo nella serie $\{Z_t\}$ spiega la differenza dell'andamento osservato rispetto a quello della serie $\{Y_t\}$.

Nonostante le funzioni di autocorrelazione delle serie che abbiamo simulato siano già note, possiamo chiedere a R di stimarle. Limitiamoci alla serie $\{Y_t\}$ e stimiamo la funzione di autocorrelazione ρ_Y (che sappiamo avere la forma descritta sopra) con campioni di ampiezza crescente. Proviamo con $n = 200$, $n = 1000$ e, infine, con $n = 10000$ e verifichiamo se la funzione di autocorrelazione campionaria $\hat{\rho}_Y$ si avvicina a quella vera ρ_Y . Con i comandi

```
> y1 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=200)
> y2 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=1000)
> y3 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=10000)
```

generiamo tre serie storiche dal medesimo processo MA(1). Nella Figura 4 si osserva che, all'aumentare del numero di osservazioni, il grafico della funzione di autocorrelazione cam-

pionaria $\hat{\rho}_Y$ diventa sempre più simile a quello della funzione di autocorrelazione vera ρ_Y .

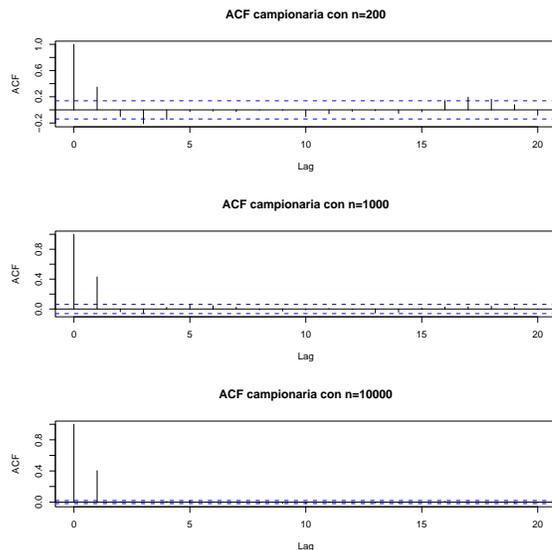


Figure 4: Grafici di $\hat{\rho}_Y$ con: (1) $n = 200$, (2) $n = 1000$ e (3) $n = 10000$ osservazioni.

Con R è anche possibile valutare esattamente la funzione di autocorrelazione di un qualsiasi processo ARMA. Il comando da usare è `ARMAacf()`. Il comando prevede di inserire alcuni argomenti

- `ar=c(...)` i valori numerici dei coefficienti delle componenti autoregressive, se presenti. Qualora si intenda calcolare la funzione di autocorrelazione di un processo MA (senza componenti AR) è sufficiente porre `ar=0`
- `ma=c(...)` i valori numerici dei coefficienti delle componenti a media mobile, se presenti. Come in precedenza, se non sono presenti componenti MA, occorre inviare `ma=0`
- `lag.max` ovvero il numero di ritardi h in corrispondenza dei quali si intende far calcolare $\rho(h)$.

Per calcolare la funzione di autocorrelazione del processo MA(3)

$$X_t = \varepsilon_t + 0.2 \varepsilon_{t-1} - 0.15 \varepsilon_{t-2} + 0.2 \varepsilon_{t-3}$$

occorre inviare l'istruzione

```
> rho <- ARMAacf(ar=0, ma=c(0.2,-0.15,0.2), lag.max=10)
```

o, più sinteticamente,

```
> rho <- ARMAacf(0, c(0.2,-0.15,0.2), lag.max=10)
```

Questi sono casi in cui l'espressione analitica esatta della funzione di autocorrelazione ρ è nota e può essere facilmente valutata. Infatti, se $\{X_t\}$ è un processo MA(q) con

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}$$

la funzione di autocorrelazione coincide con

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \beta_j \beta_{j+h}}{\sum_{j=0}^q \beta_j^2} & h = 0, 1, \dots, q \\ 0 & h > q \end{cases}$$

Anche per un processo AR(1)

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

con $|\alpha| < 1$ la funzione di autocorrelazione può essere valutata facilmente, essendo $\rho(h) = \alpha^h$, per $h = 0, 1, 2, \dots$. Più complicata è la determinazione della funzione di autocorrelazione di un processo ARMA(p, q) con $p \geq 1$ e $q \geq 1$. In questo caso, l'uso di R è di aiuto. Ad esempio, la funzione di autocorrelazione del processo

$$X_t = 0.8 X_{t-1} + \varepsilon_t - 0.45 \varepsilon_{t-1}$$

in corrispondenza ai primi 10 ritardi è calcolata da R con il comando

```
> r <- ARMAacf(ar=0.8, ma=-0.45, lag.max=10)
```

da cui

```
> r
```

	0	1	2	3	4	5
	1.00000000	0.46424870	0.37139896	0.29711917	0.23769534	0.19015627
	6	7	8	9	10	
	0.15212502	0.12170001	0.09736001	0.07788801	0.06231041	

Il grafico della funzione `r` può essere generato con l'istruzione

```
> h <- seq(0, 10, 1)
```

```
> matplot(h,r,ylab=" ",xlab="Lag",type="h",main="ACF di un processo ARMA(1,1)")
```

Con il medesimo comando `ARMAacf()` è possibile far calcolare a R anche la funzione di autocorrelazione parziale di un processo ARMA(p, q). Occorre specificare, tra le opzioni di `ARMAacf()`, `pacf=TRUE`. Senza questa specificazione, infatti, R valuta (in modo automatico) la funzione di autocorrelazione. Nell'esempio in questione si può digitare

```
> r <- ARMAacf(ar=0.8, ma=-0.45, lag.max=10, pacf=TRUE)
```

Chiudiamo questo paragrafo con una interessante funzione di R che consente di determinare un numero finito dei coefficienti nella rappresentazione $MA(\infty)$ di un processo $ARMA(p, q)$. Come visto durante il corso, se

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

è un processo $ARMA(p, q)$ stazionario, allora

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

dove la successione di coefficienti ψ_1, ψ_2, \dots è determinata dalla seguente identità tra polinomi

$$\sum_{j=0}^{\infty} \psi_j B^j = \psi(B) = \frac{\beta(B)}{\alpha(B)} = \frac{\sum_{j=0}^q \beta_j B^j}{\sum_{j=0}^p \alpha_j B^j}.$$

Sebbene sia nota l'esistenza dei coefficienti ψ_j , è difficile determinarne concretamente il valore numerico. Con R è possibile valutare un numero finito di tali coefficienti ψ_1, \dots, ψ_N mediante la funzione `ARMAtoMA()`. Per illustrarne l'impiego, si ipotizzi $p = 1$ e $q = 2$, $\beta_1 = 0.3$ e $(\alpha_1, \alpha_2) = (0.5, -0.1)$. In questo caso

```
> x <- ARMAtoMA(ar=c(0.5,-0.1), ma=0.3, lag.max=5)
> x
[1] 0.8000 0.3000 0.0700 0.0050 -0.0045
```

e R ha determinato il valore numerico dei primi 5 coefficienti ψ_1, \dots, ψ_5 nella rappresentazione $MA(\infty)$ del processo $ARMA(2, 1)$ sopra descritto.

Trend e stagionalità

La presenza di trend e/o stagionalità contribuisce a rendere non stazionarie gran parte delle seire storiche effettivamente osservate. In questi casi occorre depurare la serie da tali componenti in modo da renderle stazionarie e identificare il modello ARMA che meglio si adatti alla nuova serie (stazionaria) di dati. Per la stima del trend si possono usare filtri lineari del tipo “medie mobili” semplici e ponderate. La stima del trend in t mediante una media mobile semplice di ordine m coincide con

$$\hat{T}_t = \frac{1}{2m+1} \sum_{k=-m}^m x_{t+k}.$$

Per dati aventi periodicità stagionale, la medesima stima è data da

$$\hat{T}_t = \frac{1}{4m} (x_{t-m} + 2x_{t-m+1} + \dots + 2x_{t+m-1} + x_{t+m})$$

dove $m = 2$ per dati trimestrali e $m = 6$ per dati mensili. Per definire e valutare questi filtri lineari in R occorre usare la funzione `filter()`. Prima di usare la funzione `filter()` è

necessario memorizzare in R il vettore dei coefficienti con i quali sono ponderate le osservazioni nel filtro lineare. Ad esempio, se si vuole calcolare una media mobile semplice ponderata verrà memorizzato il seguente vettore di coefficienti

$$\boldsymbol{\alpha} = \left(\frac{1}{2m+1}, \dots, \frac{1}{2m+1} \right)$$

Se $m = 2$, allora $\boldsymbol{\alpha} = (1/5, 1/5, 1/5, 1/5, 1/5)$ e il vettore è memorizzato da R con il comando

```
> gamma <- rep(1/5, 5)
```

Supponiamo, ora, di applicare una media mobile semplice di ordine $m = 2$ per stimare il trend della serie storica di dati (trimestrali) sul PIL americano

```
> pil.us <- read.table("C:/Dati/GDP96USD.txt", col.names=c("Anno", "Trimestre", "PIL"))
```

```
> PIL <- pil.us$PIL
```

```
> xt <- ts(PIL, start=c(1947,1), freq=4)
```

La serie, come evidenziato nel grafico ottenuto nella precedente esercitazione, presenta un trend crescente nel tempo. Per stimare T_t utilizziamo una media mobile semplice di ordine $m = 2$ che è generata nel modo seguente

```
> ma <- filter(xt, gamma, method="convolution", sides=2, circular=FALSE)
```

Gli argomenti di `filter` sono

- La serie di cui si vuole stimare il trend mediante un filtro lineare, in questo caso la serie del PIL statunitense che è stata memorizzata da R con il nome `xt`
- Il vettore di pesi $(\alpha_1, \dots, \alpha_{2m+1})$ da utilizzare nel filtro

$$\hat{T}_t = \alpha_1 x_{t-m} + \dots + \alpha_{m+1} x_t + \dots + \alpha_{2m+1} x_{t+m}.$$

Nel caso in esame si vuole determinare la media mobile semplice di ordine $m = 2$. Di conseguenza $\alpha_1 = \dots = \alpha_5 = 1/5$.

- Il metodo. Qui vi sono due opzioni: `"convolution"` e `"recursive"`. Per generare una media mobile occorre usare `"convolution"`
- Il parametro `sides`. Questo determina le osservazioni utilizzate nel filtro. Se `sides=1`, allora nel filtro vengono utilizzate $2m$ osservazioni precedenti l'istante t e l'osservazione stessa in t . Se si pone `sides=2` si usano le m osservazioni precedenti t , le m osservazioni successive e l'osservazioni in t . Per la media mobile occorre fissare `sides=2`

- Il termine `circular` è specificato allo scopo di inserire NA in corrispondenza alle $2m$ osservazioni che vengono perse all'inizio e alla fine della serie.

I valori della media mobile calcolati in `ma` da R possono essere visualizzati

```
> ma
      Qtr1   Qtr2   Qtr3   Qtr4
1947    NA    NA 1503.7 1519.76
1948 1535.56 1551.22 1557.82 1559.26
...     ...     ...     ...     ...
2001 9310.12 9327.82 9364.76    NA
2002    NA
```

Come si può notare, in corrispondenza alle prime due e alle due ultime osservazioni della serie è presente l'indicazione NA, ovvero valori mancanti. I grafici della media mobile `ma` così calcolata e della serie originaria `xt` sono creati inviando le seguenti istruzioni

```
> plot(ma, ylab=" ", xlab="t", main="Media mobile di ordine 2 per la serie del PIL")
> lines(xt, col="red", lty=2)
```

Dal momento che i dati della serie del PIL sono trimestrali, è più sensato utilizzare una media mobile centrata di ordine $m = 2$, ovvero

$$\hat{T}_t = \frac{1}{8} (x_{t-2} + 2x_{t-1} + 2x_t + 2x_{t+1} + x_{t+2}).$$

Tale media mobile è generata mediante la sequenza di istruzioni

```
> alpha <- c(1/8, 1/4, 1/4, 1/4, 1/8)
> mac <- filter(xt, alpha, method="convolution", sides=2, circular=FALSE)
```

Per valutare la serie detrendizzata, supponiamo che la serie originaria sia descritta mediante un modello additivo. In questo caso, la serie detrendizzata è ottenuta mediante l'operazione

$$\hat{X}_t = X_t - \hat{T}_t.$$

Per creare la serie \hat{X}_t e fornirne una rappresentazione grafica, in R si può procedere nel modo seguente

```
> detrendx <- xt-mac
> plot(detrendx,ylab=" ",xlab="t",main="Serie del PIL US detrendizzata")
```

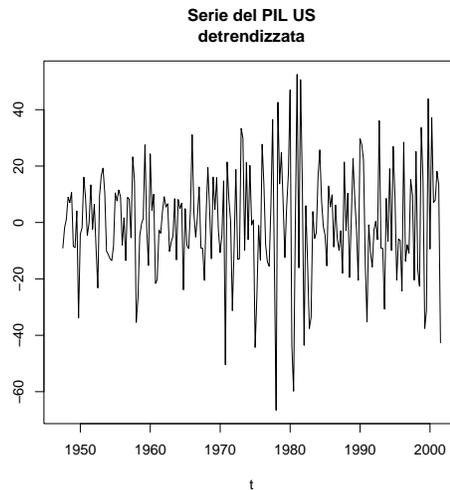


Figure 5: Serie del PIL detrendizzata, ovvero $\hat{X}_t = X_t - \hat{T}_t$, ove \hat{T}_t è stimato mediante una media mobile centrata di ordine $m = 2$.

Il grafico è rappresentato in Figura 5. Per detrendizzare la serie in presenza di un modello moltiplicativo occorre effettuare la seguente divisione

$$\hat{X}_t = \frac{X_t}{\hat{T}_t}$$

e in R viene specificato il comando

```
> detrendxx <- xt/mac
```

In R esiste anche una procedura automatizzata per detrendizzare e destagionalizzare una serie e si basa sulla funzione `decompose()`. Questa funzione, applicata ad una qualsiasi serie storica, restituisce la stima delle componenti di trend e stagionalità e la componente di errore residua che nel modello additivo è data da

$$\hat{E}_t = X_t - \hat{T}_t - \hat{S}_t$$

mentre nel modello moltiplicativo coincide con

$$\hat{E}_t = \frac{X_t}{\hat{T}_t \hat{S}_t}$$

La decomposizione della serie storica del PIL statunitense mediante un modello additivo è realizzata in R mediante l'istruzione

```
> d.pil <- decompose(xt, type="additive")
```

Il programma R ha memorizzato un oggetto, denominato `d.pil`, che contiene le seguenti informazioni

- `d.pil$seasonal` : la stima della componente stagionale. In base a quanto già detto in una delle precedenti lezioni, le componenti stagionali stimate (per dati trimestrali) sono quattro e queste si mantengono costanti per tutto il periodo di osservazione. Per dati mensili, le componenti stagionali stimate sono 12.
- `d.pil$trend` : la stima della componente di trend. La stima, in questo caso specifico, è una media mobile di ordine $m = 2$ cosicché vengono persi i primi due dati e gli ultimi due dati della serie.
- `d.pil$random` : la stima della componente di errore. Dal momento che è stato specificato l'opzione "additive", la componente di errore è ottenuta dall'operazione

$$xt - d.pil$trend - d.pil$seasonal$$

L'istruzione

```
> plot(d.pil)
```

crea in R una finestra con 4 grafici: quello della serie originaria `xt`, quello della componente di trend, quello della componente stagionale e, infine, il grafico della componente di errore.

Differenziazione di una serie storica

Come è noto, una serie storica può essere detrendizzata (o destagionalizzata) applicando gli operatori differenza

$$\nabla^d = (1 - B)^d \quad \nabla_p^D = (1 - B^p)^D$$

essendo B l'operatore di retro-traslazione (o ritardo) e p la periodicità stagionale. Il comando che consente di generare in R una serie differenziata è `diff()`.

Si supponga di lavorare ancora con la serie `xt` del PIL e che il trend nella serie sia lineare. Se si vuole detrendizzare, si può applicare l'operatore differenza di ordine $d = 1$ e l'istruzione da inviare a R è

```
> dxt <- diff(xt, lag=1, differences=1)
```

Si noti che in `diff()` vanno specificati: (1) la serie cui applicare l'operatore differenza ∇^d ; (2) il valore numerico di `lag` che rappresenta la potenza da applicare all'operatore ritardo B ; (3) il valore numerico del parametro `differences` che coincide con l'ordine d (o D se si destagionalizza). Indicata con X_t la serie originaria, la precedente istruzione ha definito in R il seguente oggetto

$$dxt \equiv Y_t = \nabla X_t = (1 - B)X_t = X_t - X_{t-1}$$

Se, oltre a detrendizzare, si intende altresì destagionalizzare la serie si può procedere come segue

```
> ddxt <- diff(dxt, lag=4, differences=1).
```

La nuova serie storica così generata risulta uguale a

$$ddxt = Y_t - Y_{t-4} = \nabla_4 Y_t = \nabla_4 \nabla X_t.$$

Se i dati della serie fossero stati mensili, la precedente istruzione avrebbe dovuto essere modificata come segue

```
> ddxt.mensile <- diff(dxt, lag=12, differences=1).
```

Nel caso della serie del PIL statunitense, il grafico della serie

$$Z_t = Y_t - Y_{t-4} = \nabla_4 \nabla X_t$$

ottenuta differenziando sia la componente stagionale e quella non stagionale è presentato in Figura 6. Differenze di ordine più elevato (in questo non necessarie) si ottengono modificando

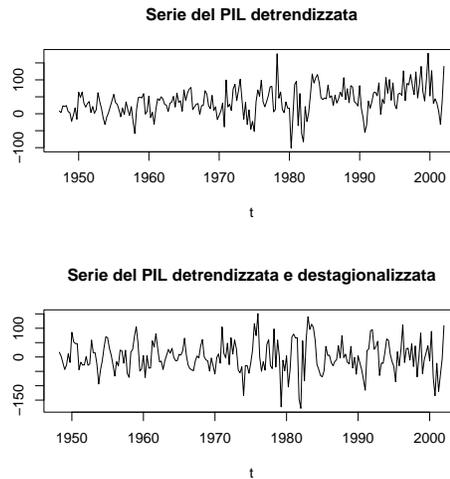


Figure 6: In alto il grafico della serie differenziata $Y_t = X_t - X_{t-1}$. In basso il grafico della serie differenziata anche nella sua componente stagionale $Z_t = Y_t - Y_{t-4} = \nabla_4 \nabla X_t$.

gli argomenti di `diff()`. Ad esempio, se si vuole differenziare due volte la serie X_t per ottenere

$$Z_t = \nabla^2 X_t = (1 - B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

si digita

```
> zt <- diff(xt, lag=1, differences=2)
```

Lo studio (preliminare) della stagionalità mediante strumenti grafici è possibile in R utilizzando varie funzioni. Una prima funzione è il cosiddetto `monthplot()`. Questa fornisce il

grafico della serie nei singoli mesi trimestri. Se i dati sono mensili, il `monthplot()` genera 12 curve distinte che evidenziano l'andamento (negli anni) per i singoli mesi. Applichiamo questa funzione al dataset `AirPassengers`. Questo dataset è già presente in R e si riferisce al numero di passeggeri statunitensi (espressi in migliaia) imbarcati su voli di linea diretti verso l'estero. I dati sono mensili e vanno da Gennaio 1949 fino a Dicembre 1960. Per richiamare il dataset in R è necessario inviare

```
> data(AirPassengers)
```

Dal grafico si notano un trend crescente e delle fluttuazioni cicliche stagionali. Per visualizzare in termini grafici il comportamento della componente stagionale, si digiti il comando

```
> monthplot(AirPassengers)
```

generando, così, il grafico della Figura 7. Le linee orizzontali corrispondono alle medie dei

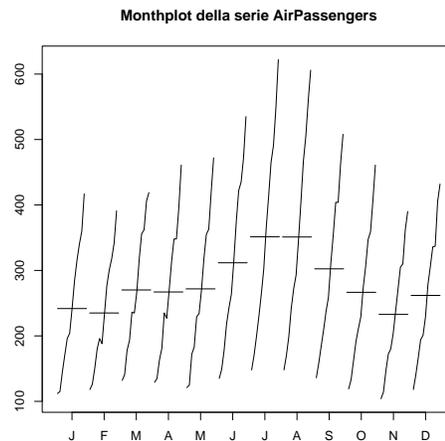


Figure 7: Il monthplot della serie `AirPassengers` evidenzia la presenza di un trend nella componente stagionale.

valori delle singole sottoserie corrispondenti a ciascuno dei 12 mesi dell'anno. I valori medi sono più elevati nei mesi di Luglio e Agosto e si nota anche un deciso incremento (trend) del numero di passeggeri in ciascuno dei mesi dell'anno. Questa osservazione ci suggerisce che occorre applicare alla serie `AirPassengers` sia l'operatore ∇ sia l'operatore ∇_{12} .

Un altro diagramma utile per individuare la presenza di trend e/o stagionalità è il cosiddetto *diagramma di autodispersione* che viene creato da R utilizzando la funzione `lagplot`. Il grafico di autodispersione è individuato dalle coppie di punti $\{(X_t, X_{t+k}) : t = 1, 2, \dots, n-k\}$. Nel caso della serie `AirPassengers` si potrà digitare

```
> lag.plot(AirPassengers, set.lags=1:12, type="p", do.lines=FALSE)
```

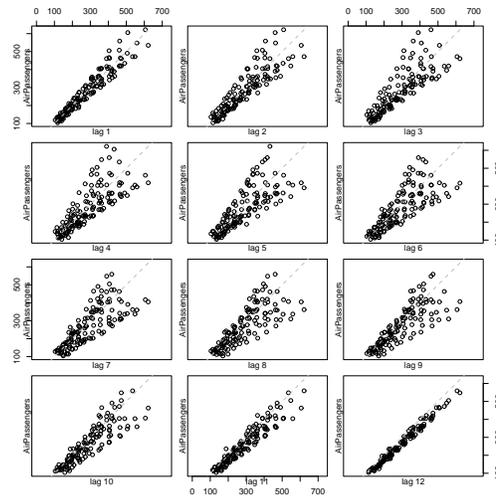


Figure 8: Grafici di autodispersione della serie `AirPassengers`.

Il grafico corrispondente è rappresentato nella Figura 8. Il parametro `set.lags` identifica i valori del ritardo k per i quali si vuole far tracciare il diagramma di autodispersione. Se, ad esempio, si desidera visualizzare solo i grafici corrispondenti ai ritardi $k = 1$ e $k = 12$ è sufficiente specificare la seguente istruzione

```
> lag.plot(AirPassengers, set.lags=c(1,12), type="p", do.lines=FALSE)
```

Un ultimo strumento grafico che consente di estrapolare alcune informazioni utili circa la stagionalità di una serie storica è il cosiddetto *seasonal plot*. Il grafico non è generato in automatico da R ma può essere creato facilmente ricorrendo ad una sequenza di comandi di R. Il *seasonal plot* è il grafico della serie storica limitata ad un singolo anno. Ciò consente di visualizzare il comportamento dei dati nel corso dell'anno e verificare se si manifestano regolarità che si ripetono nei vari anni di osservazione. Lavorando sempre con il dataset `AirPassengers`, estrapoliamo le serie degli anni 1949, 1951, 1954, 1956, 1958 e 1960 con il comando `window`

```
> y1 <- window(AirPassengers, start=c(1949,1), end=c(1949,12))
> y2 <- window(AirPassengers, start=c(1951,1), end=c(1951,12))
> y3 <- window(AirPassengers, start=c(1954,1), end=c(1954,12))
> y4 <- window(AirPassengers, start=c(1956,1), end=c(1956,12))
> y5 <- window(AirPassengers, start=c(1958,1), end=c(1958,12))
> y6 <- window(AirPassengers, start=c(1960,1), end=c(1960,12))
```

Le sei serie memorizzate da R sono costituite da 12 osservazioni ciascuna (corrispondenti ai

12 mesi dell'anno). Per generare il grafico occorre creare un vettore $x = (1, \dots, 12)$ che sarà utilizzato sull'asse delle ascisse. Dal momento che

```
> range(AirPassengers)
[1] 104 622
```

dovremo generare una finestra grafica che ricomprenda tutti i valori da 104 a 622 specificando in modo opportuno il parametro `ylim`. La sequenza di istruzioni da fornire a R è la seguente

```
> x <- seq(1,12,1)
> matplot(x, y1, type="l", ylim=c(100,650), ylab=" ", xlab="Mese")
> lines(x, y2, type="l", lty=2, col="red")
> lines(x, y3, type="l", lty=3, col="blue", lwd=3)
> lines(x, y4, type="l", lty=4, col="green", lwd=3)
> lines(x, y5, type="l", lty=5, col="pink", lwd=2)
> lines(x, y6, type="l", lty=6, col="brown", lwd=2)
```

Si noti che per tracciare il seasonal plot, vedi Figura 9, occorre trattare y_1, \dots, y_6 come vettori e non come serie storiche. Di conseguenza per tracciare il grafico, abbiamo usato il comando `matplot` nel quale occorre specificare sia la variabile in ascissa (nel nostro caso `h`) che la variabile in ordinata (nel nostro caso `y1`). Per sovrapporre i grafici delle serie successive si usa il comando `lines()`. La linea continua in basso nel grafico della Figura 9 si riferisce

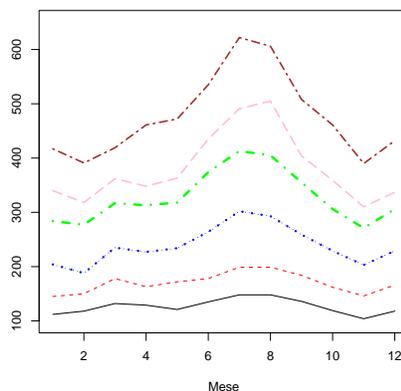


Figure 9: Seasonal plot della serie `AirPassengers`, relativamente agli anni 1949, 1951, 1954, 1956, 1958 e 1960.

ai 12 valori della serie nell'anno 1949 e gli anni successivi si trovano su livelli sempre più elevati. Il grafico suggerisce due conclusioni: (a) la serie presenta una significativa componente

stagionale dal momento che il seasonal plot presenta, in tutti gli anni, picchi significativi in corrispondenza dei mesi di luglio e di agosto e l'andamento nei diversi anni è pressoché simile; (b) i valori numerici della serie tendono a crescere nel tempo (le curve tendono a collocarsi su livelli più elevati con il passare degli anni) ed aumenta l'oscillazione tra i valori minimi e quelli massimi della serie nel corso dell'anno: ciò è spiegabile con la presenza di una componente di trend nella stagionalità. Queste considerazioni, che verranno ulteriormente supportate dall'analisi della funzione di autocorrelazione, ci inducono a differenziare la serie applicando gli operatori ∇ e ∇_{12} mediante la sequenza di istruzioni

```
> yt <- diff(AirPassengers, lag=1, differences=1)
```

```
> zt <- diff(yt, lag=12, differences=1)
```

si genera la serie

$$Z_t = \nabla \nabla_{12} X_t = (1 - B)(1 - B^{12}) X_t$$

il cui grafico e la cui funzione di autocorrelazione sono rappresentati in Figura 10. Il grafico

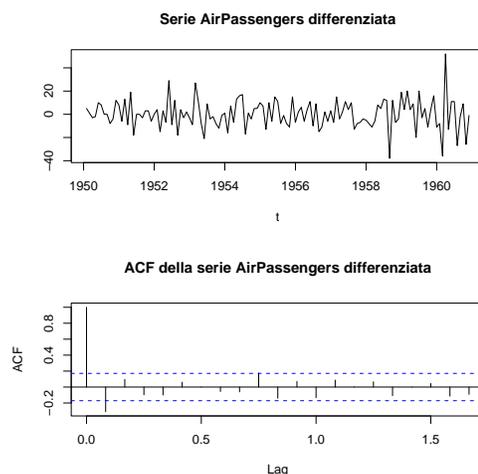


Figure 10: Grafici della serie $Z_t = \nabla \nabla_{12} X_t$, essendo X_t la serie `AirPassengers`.

della serie lascia presumere che $\{Z_t\}$ sia stazionario. Si uò, quindi, procedere all'identificazione del modello che ha generato $\{Z_t\}$ e che sarà utilizzato per fare previsioni.

Esercizio. Si provi ad applicare le medesime operazioni alla serie `JohnsonJohnson` memorizzata in R. Si tenga conto del fatto che questi dati sono trimestrali.