

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Specialistica in Matematica

**PROCESSI STAZIONARI E
ANALISI DELLE SERIE STORICHE**

Tesi di Laurea in Processi stocastici

Presentata da: **GIANLUCA VASILE**

Relatore:
Chiar.ma Prof.ssa
IRENE CRIMALDI

Correlatore:
Dott.ssa **F. MARTA L.
DI LASCIO**

I Sessione
Anno Accademico 2010-2011

*Congetturare una certa cosa significa misurarne la probabilità.
Perciò designiamo con arte della congettura (ars coniectandi
sive stochastice), l'arte di misurare il più esattamente possibile
la probabilità delle cose allo scopo di poter costantemente
scegliere quel giudizio e seguire quella condotta, che ci sembrano
migliori, più consoni, più sicuri o più consigliabili. Solo in
ciò consiste la saggezza del filosofo e il senno dell'uomo di Stato.*

Jacob Bernoulli, "Ars coniectandi", 1713

Introduzione

Oggigiorno siamo talmente abituati ad associare l'idea di futuro alla probabilità che pianifichiamo la nostra stessa vita a partire da essa: stipuliamo assicurazioni, guardiamo le previsioni del tempo, valutiamo il rischio di operazioni finanziarie, etc.

Da qui nasce l'idea di scrivere la presente dissertazione sui *processi stocastici stazionari*, che sono uno dei modelli probabilistici più usati, e sull'*analisi delle serie storiche*, che ha come obiettivo proprio quello di individuare un opportuno processo stocastico che si adatti alle osservazioni e che possa essere usato per formulare previsioni future.

Un *processo stocastico* è una famiglia di variabili aleatorie ordinate rispetto a un parametro. Essendo un insieme finito di osservazioni relative ad un certo fenomeno, ordinate nel tempo, una *serie storica* può essere vista quindi come una realizzazione finita di un processo stocastico. L'obiettivo dell'analisi delle serie storiche è quello di usare la serie storica osservata per individuare il processo stocastico generatore da utilizzare per fare previsioni future. I processi stocastici stazionari godono di particolari proprietà probabilistiche che risultano utili per modellare molte serie storiche che si incontrano nella realtà.

La tesi è così organizzata. Nel primo capitolo introduciamo le nozioni di processo stocastico stazionario e serie storica, illustrando quella che è l'analisi classica delle serie storiche, ossia la scomposizione nelle relative componenti: trend, stagionalità e rumore. Vengono inoltre illustrati diversi metodi per ricondursi da una serie storica non stazionaria ad una stazionaria con vari

esempi. Nel capitolo successivo presentiamo e analizziamo una classe di processi stocastici stazionari molto importanti nella modellizzazione delle serie storiche: i *processi autoregressivi a media mobile* (in inglese, *autoregressive moving average processes*), detti processi ARMA. Oltre a illustrare le proprietà di questi processi, mostreremo tre diversi metodi numerici per il calcolo della funzione di autocorrelazione. Il terzo capitolo è incentrato sul Teorema di Herglotz e sulla rappresentazione spettrale di un processo stazionario. Le nozioni qui presentate vengono utilizzate nel capitolo seguente per sviluppare la previsione di un processo stazionario nel dominio delle frequenze. Nel quarto capitolo viene trattata anche la previsione di un processo stazionario nel dominio dei tempi. A questo proposito, vengono illustrati due metodi ricorsivi per calcolare il migliore predittore lineare: i cosiddetti algoritmi di Durbin-Levison e delle innovazioni. Il quinto capitolo è incentrato sulla costruzione di un modello ARMA a partire dai dati. In particolare, sono trattati gli stimatori della media, della funzione di autocovarianza e della funzione di autocorrelazione. Sono presentati, inoltre, diversi metodi di stima degli ordini e dei coefficienti di un processo ARMA. Nel sesto capitolo illustriamo l'analisi "moderna" delle serie storiche: la *procedura di Box e Jenkins* per l'identificazione e la stima di *processi autoregressivi-integrati a media mobile stagionali* (in inglese, *seasonal autoregressive-integrated moving average processes*), detti processi SARIMA o, più comunemente, ARIMA. Dapprima ne facciamo una trattazione teorica, poi applichiamo la procedura di Box e Jenkins alla serie storica dei passeggeri internazionali delle linee aeree osservati mensilmente dal gennaio del 1960 al dicembre del 1971, fornendo una previsione per l'anno 1972. In appendice D è riportato il codice dell'analisi effettuata mediante il software statistico open source R.

Sono molte le persone a cui vanno i miei ringraziamenti per l'aiuto e i suggerimenti con cui mi hanno incoraggiato durante la stesura della tesi. In particolare vorrei esprimere profonda gratitudine al mio relatore, Prof.ssa Irene Crimaldi, che mi ha seguito durante tutto il lavoro, per la disponibilità, gli

incoraggiamenti e per avermi trasmesso quella precisione indispensabile per un approccio corretto alla matematica. Vorrei inoltre ringraziare la Dott.ssa Francesca Marta Lilja Di Lascio per avermi assistito nella stesura dell'ultimo capitolo, per i suoi utili consigli e per avermi guidato in un approccio "più statistico" all'analisi delle serie storiche.

Indice

Introduzione	i
1 Le serie storiche	1
1.1 I processi stocastici	1
1.2 La stazionarietà	3
1.2.1 La funzione di autocovarianza e di autocorrelazione . .	7
1.2.2 La funzione di autocovarianza campionaria e di auto- correlazione campionaria	9
1.2.3 La funzione di autocorrelazione parziale	10
1.3 La decomposizione di una serie storica	11
1.3.1 Eliminazione del trend in assenza di stagionalità	13
1.3.2 Eliminazione del trend e della stagionalità	21
2 I processi ARMA	29
2.1 Il rumore bianco	29
2.2 I processi MA ed AR	30
2.3 I processi ARMA	35
2.4 Il calcolo della funzione di autocovarianza di un processo ARMA	44
3 La rappresentazione spettrale	49
3.1 I processi stazionari a valori complessi	49
3.2 Un esempio di rappresentazione spettrale	51
3.3 Il teorema di Herglotz	52
3.4 La densità spettrale di un processo ARMA	57

3.4.1	Approssimazione della densità spettrale	63
3.5	La rappresentazione spettrale	67
4	La previsione di un processo stazionario	75
4.1	La previsione nel dominio dei tempi	75
4.2	Alcuni metodi ricorsivi per calcolare il migliore predittore lineare	77
4.2.1	L'algoritmo di Durbin-Levinson	77
4.2.2	L'algoritmo delle innovazioni	83
4.2.3	Il calcolo della previsione a più passi	86
4.3	La previsione di un processo ARMA	87
4.3.1	La previsione a più passi di un processo ARMA	92
4.3.2	La previsione di un processo ARMA causale invertibile	96
4.4	La previsione nel dominio delle frequenze	98
5	La costruzione di un modello ARMA	103
5.1	La stima della media	103
5.2	La stima della funzione di autocovarianza e di autocorrelazione	106
5.3	Alcuni metodi di stima dei parametri di un processo ARMA .	111
5.3.1	Lo stimatore di Yule-Walker per i parametri di un processo autoregressivo	112
5.3.2	L'algoritmo di Durbin-Levinson per la stima dei para- metri di un processo autoregressivo	114
5.3.3	La stima dei parametri di un processo a media mobile usando l'algoritmo delle innovazioni	117
5.3.4	La stima dei parametri di un processo ARMA	118
6	I processi SARIMA e l'analisi di una serie storica reale	121
6.1	La procedura di Box e Jenkins	122
6.1.1	Analisi preliminare	123
6.1.2	Identificazione del modello	128
6.1.3	Stima del modello	129
6.1.4	Verifica del modello	130

6.1.5	Valutazione della capacità previsiva del modello	132
6.2	Un esempio di applicazione della procedura di Box e Jenkins .	133
A	Alcuni richiami sugli spazi di Hilbert	147
B	I processi ad incrementi ortogonali	151
C	Alcuni richiami sulla serie di Fourier	155
D	Codice in R	157
	Bibliografia	161

Elenco delle figure

1.1	Popolazione degli USA,1790-1980.	14
1.2	Popolazione degli USA,1790-1980, approssimata con i minimi quadrati.	15
1.3	Scioperi negli USA,1951-1980.	17
1.4	Media mobile sul trend m_t	18
1.5	Residui sulla serie storica degli scioperi.	18
1.6	Serie con operatore alle differenze del secondo ordine.	20
1.7	Morti accidentali mensili negli USA, 1973-1978.	22
1.8	Morti accidentali mensili dopo aver sottratto il trend.	23
1.9	Componente stagionale delle morti accidentali mensili.	24
1.10	Eliminazione del trend e della stagionalità	24
1.11	Confronto tra media mobile e media aritmetica del trend.	26
1.12	Serie differenziata $\{\nabla x_t, t = 13, \dots, 72\}$	27
1.13	Serie differenziata $\{\nabla\nabla x_t, t = 14, \dots, 72\}$	28
3.1	Densità spettrale $f_X(\lambda)$ di un processo MA(1)	60
3.2	Densità spettrale $f_X(\lambda)$ di un processo AR(1)	61
5.1	Grafico della funzione di autocorrelazione campionaria relativa ad un campione di 200 osservazioni di un processo MA(1) della forma (5.4).	111
6.1	Diagramma della procedura di Box e Jenkins.	124

6.2	Grafico della serie storica dei passeggeri internazionali sulle linee aeree rilevati dal 1960 al 1971.	134
6.3	Range mean plot della serie storica non stazionaria.	134
6.4	Grafico della funzione di autocorrelazione del logaritmo della serie storica oggetto di studio.	135
6.5	Grafico della funzione di autocorrelazione globale e della funzione di autocorrelazione parziale della serie resa stazionaria. .	137
6.6	Grafici dei residui del modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato sulla serie storica oggetto di studio.	140
6.7	Grafici relativi alla funzione di autocorrelazione dei residui del modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato.	142
6.8	Grafici relativi alla valutazione del tipo di distribuzione dei residui del modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato.	143
6.9	Grafico delle ultime 12 osservazioni della serie previste mediante il modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato.	145
6.10	Grafico della serie storica airpass con le previsioni per l'anno 1972.	146

Capitolo 1

Le serie storiche

Una serie storica è un insieme finito di osservazioni, ordinate nel tempo, di un certo fenomeno che si evolve in modo aleatorio. Ogni osservazione x_t può essere vista come una realizzazione di una variabile aleatoria X_t . Una serie storica $\{x_t, t \in T_0\}$ è quindi una realizzazione di una famiglia di variabili aleatorie $(X_t)_{t \in T_0}$. Ciò ci fa pensare a un possibile modello delle serie storiche: i processi stocastici.

In questo capitolo faremo alcuni richiami sulla nozione di processo stocastico e poi restringeremo la nostra attenzione ad una classe di processi stocastici che è utile per modellare la maggior parte delle serie storiche che si incontrano nella pratica: i processi stocastici stazionari.

1.1 I processi stocastici

Cominciamo col dare la definizione di processo stocastico.

Definizione 1.1. Un *processo stocastico reale* è una famiglia $(X_t)_{t \in T}$ di variabili aleatorie reali definite su uno stesso spazio di probabilità (Ω, \mathcal{F}, P) . L'insieme T è detto *insieme dei tempi*.

Un processo stocastico è quindi un'applicazione che associa un numero reale a una coppia (ω, t) costituita, rispettivamente, da una eventualità e da un istante temporale.

Una prima classificazione dei processi stocastici deriva dalla tipologia dell'insieme T , che per noi sarà sempre un sottoinsieme di \mathbb{R} . Si avrà un processo stocastico *discreto* quando l'insieme dei tempi T è un insieme discreto (in genere \mathbb{N} o \mathbb{Z}); mentre si avrà un processo stocastico *continuo* quando T è un intervallo di \mathbb{R} .

Definizione 1.2. Dato un processo stocastico $(X_t)_{t \in T}$, la variabile aleatoria X_t , fissato $t \in T$, è detta *stato del processo* all'istante t .

Per ogni $\omega \in \Omega$ fissato, l'applicazione $t \mapsto X_t(\omega)$ è detta *realizzazione* o *traiettoria del processo* associata all'eventualità ω e viene denotata con il simbolo $X(\omega)$.

Per descrivere un processo stocastico è necessario conoscere per ogni insieme di istanti temporali (t_1, t_2, \dots, t_n) la distribuzione congiunta di

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}).$$

In altre parole, la legge di un processo stocastico è univocamente determinata dalle sue leggi finito-dimensionali.

Definizione 1.3. Sia \mathcal{T} l'insieme di tutti i vettori

$$\{\mathbf{t} = (t_1, \dots, t_n) \in T^n : t_1 < t_2 < \dots < t_n, n = 1, 2, \dots\}.$$

Si chiamano *funzioni di ripartizione finito-dimensionali* del processo stocastico $(X_t)_{t \in T}$ le funzioni $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathcal{T}\}$ definite per $\mathbf{t} = (t_1, \dots, t_n)$ da

$$F_{\mathbf{t}}(\mathbf{x}) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n), \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Il seguente risultato ci assicura che, assegnate delle funzioni di ripartizione finito-dimensionali che soddisfano un'opportuna proprietà di consistenza, esiste un processo stocastico che le ha come proprie funzioni di ripartizione.

Teorema 1.1.1. (Teorema di Kolmogorov). *Siano date le funzioni di ripartizione $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathcal{T}\}$. Allora esse sono le funzioni di ripartizione finito-dimensionali di un processo stocastico se e solo se per ogni $n \geq 1$, per ogni $\mathbf{t} = (t_1, \dots, t_n) \in \mathcal{T}$ e $1 \leq i \leq n$, si ha*

$$\lim_{x_i \rightarrow +\infty} F_{\mathbf{t}(\mathbf{x})} = F_{\mathbf{t}^{(i)}}(\mathbf{x}^{(i)}) \quad (1.1)$$

dove $\mathbf{t}(i)$ e $\mathbf{x}(i)$ sono vettori di $(n - 1)$ componenti ottenuti cancellando rispettivamente la i -esima componente di \mathbf{t} e \mathbf{x} .

Se $\phi_{\mathbf{t}}(\cdot)$ è la funzione caratteristica relativa a $F_{\mathbf{t}}(\cdot)$, ossia

$$\phi_{\mathbf{t}}(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\mathbf{u}'\mathbf{x}} F_{\mathbf{t}}(dx_1, \dots, dx_n), \quad \mathbf{u} = (u_1, \dots, u_n)' \in \mathbb{R}^n,$$

allora la (1.1) può essere riscritta in forma equivalente come

$$\lim_{u_i \rightarrow 0} \phi_{\mathbf{t}}(\mathbf{u}) = \phi_{\mathbf{t}(i)}(\mathbf{u}(i)),$$

dove $\mathbf{u}(i)$ è un vettore di $(n - 1)$ componenti ottenuto cancellando la i -esima componente del vettore \mathbf{u} .

Definizione 1.4. Una *serie storica* è un insieme finito di osservazioni ordinate nel tempo e solitamente misurate ad intervalli di tempo equispaziati (mensile, trimestrale, annuale). È quindi definita come una parte finita di una realizzazione di un processo stocastico.

Una serie storica $\{x_t : t \in T_0\}$ risulta dunque essere una successione finita di valori, ognuno dei quali è la realizzazione di una variabile aleatoria X_t che fa parte di un processo stocastico $(X_t)_{t \in T}$, con $T_0 \subseteq T$.

1.2 La stazionarietà

L'analisi delle serie storiche ha come obiettivo principale quello di individuare un opportuno processo stocastico che abbia traiettorie che si adattino ai dati, per poter poi formulare previsioni. Per fare questo, risulta necessario restringere la nostra attenzione ad una classe di processi stocastici che ci consenta di:

1. identificare univocamente il processo, ossia il modello;
2. fare inferenza sui momenti del processo stesso, cioè di ottenere delle stime corrette e consistenti dei momenti del processo.

Utilizzando strutture probabilistiche dobbiamo infatti tener conto che ciò che osserviamo non è che una sola realizzazione e peraltro relativa ad un intervallo di tempo limitato. D'altra parte, se ciò che osserviamo viene interpretata come una realizzazione finita di un processo stocastico che gode di particolari proprietà, allora è possibile trovare un unico modello adatto a rappresentare l'evoluzione temporale del fenomeno oggetto di studio.

Da qui in avanti ci restringiamo a prendere in esame solo i processi stocastici *stazionari*. Intuitivamente, un processo stocastico è stazionario se la sua struttura probabilistica è invariante nel tempo.

Prima di definire in maniera formale la proprietà di stazionarietà abbiamo bisogno della seguente definizione.

Definizione 1.5. Dato un processo stocastico $(X_t)_{t \in T}$ tale che ogni X_t sia di quadrato integrabile, la funzione di *autocovarianza* $\gamma_X(\cdot, \cdot)$ di $(X_t)_{t \in T}$ è data da

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - E[X_r])(X_s - E[X_s])], \quad r, s \in T. \quad (1.2)$$

L'autocovarianza indica quindi il modo in cui due variabili del processo co-variano nel tempo.

Siamo adesso pronti per dare la proprietà di stazionarietà .

Definizione 1.6. Un processo stocastico $(X_t)_{t \in \mathbb{Z}}$ è detto *stazionario* o *debolmente stazionario* se soddisfa le seguenti condizioni:

1. $E[|X_t|^2] < \infty$,
2. $E[X_t] = m \forall t \in \mathbb{Z}$,
3. $\gamma_X(r, s) = \gamma_X(r + t, s + t) \forall r, s, t \in \mathbb{Z}$,

ossia se ha (1) momenti secondi finiti, (2) valore atteso costante e (3) autocovarianza non dipendente dal tempo.

Se $(X_t)_{t \in \mathbb{Z}}$ è un processo stazionario allora

$$\gamma_X(r, s) = \gamma_X(r - s, 0) \quad \forall r, s \in \mathbb{Z}.$$

Risulta quindi conveniente definire l'autocovarianza di un processo stocastico stazionario in funzione di una sola variabile:

$$\gamma_X(h) = \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t) \quad \forall t, h \in \mathbb{Z}$$

dove la quantità h viene chiamata *lag* o *passo*. Ponendo $h = 0$ nella (1.2), si ha

$$\gamma_X(0) = \text{Var}(X_t) = E[(X_t - E[X_t])^2]. \quad (1.3)$$

Per ogni h , $\gamma_X(h)$ misura la covarianza tra due variabili del processo, che risultano separate da un intervallo di tempo di lunghezza h .

Osservazione 1.1. Abbiamo definito la stazionarietà solo nel caso $T = \mathbb{Z}$. Per la nostra trattazione ci interessa solo questo caso, ma non è difficile estendere questa nozione a casi più generali.

Un'altra importante nozione di stazionarietà è data dalla seguente definizione.

Definizione 1.7. Un processo stocastico $(X_t)_{t \in \mathbb{Z}}$ è detto *strettamente o fortemente stazionario* se presi $t_1 < t_2 < \dots < t_k$, le distribuzioni congiunte di $(X_{t_1}, \dots, X_{t_k})$ e $(X_{t_1+h}, \dots, X_{t_k+h})$ sono uguali per tutti gli interi positivi k e per ogni t_1, \dots, t_k, h in \mathbb{Z} .

Per un processo fortemente stazionario la distribuzione congiunta è quindi invariante per traslazioni temporali di una quantità arbitraria h .

Se $(X_t)_{t \in \mathbb{Z}}$ è fortemente stazionario, prendendo $k = 1$ nella definizione precedente, si ottiene che tutte le variabili X_t sono identicamente distribuite. Inoltre, un processo stocastico fortemente stazionario con momenti secondi finiti, è anche stazionario. Il viceversa, invece, non è sempre vero: un processo stazionario non è detto sia anche fortemente stazionario. La stazionarietà in senso forte coinvolge infatti tutti i momenti della distribuzione di probabilità del processo e non solo i primi due. Vediamo un esempio di serie storica stazionaria che non è fortemente stazionaria.

Esempio 1. Sia $(X_t)_{t \in \mathbb{Z}}$ una sequenza di variabili aleatorie indipendenti tale che X_t ha distribuzione esponenziale di parametro 1 quando t è dispari, mentre ha distribuzione normale con media 1 e varianza 1 quando t è pari. Allora $(X_t)_{t \in \mathbb{Z}}$ è un processo stazionario con $\gamma_X(0) = 1$ e $\gamma_X(h) = 0$ per $h \neq 0$. Dal fatto che X_1 e X_2 hanno distribuzione diversa, possiamo concludere che il processo $(X_t)_{t \in \mathbb{Z}}$ non è fortemente stazionario.

Esempio 2 (Passeggiata aleatoria). Sia $(X_t)_{t \in \mathbb{N}}$ un processo stocastico tale che

$$X_0 = 0, \quad X_t = \sum_{i=1}^t \epsilon_i \quad \text{per } t \geq 1, \quad \text{con } (\epsilon_t)_{t \geq 1} \sim IID(0, \sigma^2)^1.$$

Il processo $(X_t)_{t \in \mathbb{N}}$ è detto *passeggiata aleatoria* o, in inglese, *random walk*. Questo processo gode delle seguenti proprietà:

$$E[X_t] = E[\epsilon_1] + \dots + E[\epsilon_t] = 0 \quad \forall t \geq 1,$$

$$\text{Var}(X_t) = \sum_{j=1}^t \text{Var}(\epsilon_j) = t\sigma^2 \quad \forall t \geq 1,$$

$$\text{Cov}(X_t, X_{t-h}) = E[X_t X_{t-h}] = (t-h)\sigma^2 \quad \forall t \geq 1 \text{ e } \forall h > 0.$$

La passeggiata aleatoria è quindi un processo non stazionario ma con incrementi $X_t - X_{t-1} = \epsilon_t$ stazionari.

Una tipologia di processi stocastici stazionari che risultano essere anche fortemente stazionari sono i processi gaussiani.

Definizione 1.8. Un processo stocastico $(X_t)_{t \in \mathbb{Z}}$ si dice *gaussiano* se per ogni n in \mathbb{N}^+ e per ogni (t_1, t_2, \dots, t_n) in \mathbb{Z}^n il vettore aleatorio n -dimensionale $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ ha distribuzione gaussiana multivariata, cioè la sua densità di probabilità è:

$$f_{(X_{t_1}, X_{t_2}, \dots, X_{t_n})}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{K}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{b})^T \mathbf{K}^{-1}(\mathbf{x} - \mathbf{b})\right)$$

dove $b_i = E[X_{t_i}]$ per ogni $i = 1, \dots, n$ e $\mathbf{K} = (\text{Cov}(X_{t_i}, X_{t_j}))_{i,j=1, \dots, n}$ è la matrice di covarianza.

¹Variabili aleatorie indipendenti e identicamente distribuite con media zero e varianza σ^2 .

Se $(X_t)_{t \in \mathbb{Z}}$ è un processo gaussiano stazionario allora $(X_t)_{t \in \mathbb{Z}}$ è fortemente stazionario: infatti, per ogni $n \geq 1$ e ogni h, t_1, t_2, \dots, t_n in \mathbb{Z} , i vettori aleatori $(X_{t_1}, \dots, X_{t_n})$ e $(X_{t_1+h}, \dots, X_{t_n+h})$ sono gaussiani con la stessa media e la stessa matrice di covarianza, quindi hanno la stessa distribuzione.

1.2.1 La funzione di autocovarianza e di autocorrelazione

In questo paragrafo analizziamo le proprietà delle funzioni di autocovarianza e di autocorrelazione.

Definizione 1.9. La funzione di *autocorrelazione* di $(X_t)_{t \in \mathbb{Z}}$ è definita come

$$\rho_X(h) = \frac{\text{Cov}(X_t, X_{t+h})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+h})}}.$$

Nel caso di un processo stazionario, grazie a (1.3), possiamo scrivere

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad \forall h \in \mathbb{Z}.$$

Per ogni h , $\rho_X(h)$ rappresenta il coefficiente di correlazione tra due variabili del processo, che risultano separate da un intervallo di tempo di lunghezza h . L'autocorrelazione esprime quindi la dipendenza lineare che esiste tra il processo al tempo t e se stesso al tempo $t + h$.

Vediamo alcune proprietà della funzione di autocovarianza.

Proposizione 1.2.1. Sia $\gamma_X(\cdot)$ la funzione di autocovarianza di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$, allora valgono:

1. $\gamma_X(0) \geq 0$,
2. per ogni n e ogni $(t_1, t_2, \dots, t_n) \in \mathbb{Z}^n$, la matrice $\Gamma_n = \{\gamma_X(t_i - t_j)\}_{i,j=1,\dots,n}$ è semidefinita positiva,
3. $|\gamma_X(h)| \leq \gamma_X(0) \quad \forall h \in \mathbb{Z}$,

$$4. \quad \gamma_X(h) = \gamma_X(-h) \quad \forall h \in \mathbb{Z}.$$

Dimostrazione. Per semplicità supponiamo $E[X_t] = 0$ per ogni t . Allora $\gamma_X(0) = E[X_t^2] \geq 0$. Inoltre, per ogni $j = 1, \dots, n$, ogni t_j in \mathbb{Z} , e a_j in \mathbb{R} vale

$$\begin{aligned} 0 &\leq \text{Var} \left(\sum_{j=1}^n a_j X_{t_j} \right) = E \left[\sum_{j=1}^n \sum_{k=1}^n a_j a_k X_{t_j} X_{t_k} \right] \\ &= \sum_{j=1}^n \sum_{k=1}^n a_j a_k E[X_{t_j} X_{t_k}] = \sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_X(t_j - t_k). \end{aligned}$$

Ne segue la proprietà (3) prendendo $n = 2$. Si ottiene infatti

$$0 \leq a_1^2 \gamma_X(0) + a_2^2 \gamma_X(0) + 2a_1 a_2 \gamma_X(t_1 - t_2),$$

da cui si ricava

$$\frac{1}{2}(a_1^2 + a_2^2) \geq -a_1 a_2 \frac{\gamma_X(t_1 - t_2)}{\gamma_X(0)}.$$

Ponendo $h = t_1 - t_2$ e scegliendo $-a_1 = a_2 = 1$ otteniamo $\gamma_X(h) \leq \gamma_X(0)$.

Scegliendo poi $a_1 = a_2 = 1$ otteniamo $\gamma_X(h) \geq -\gamma_X(0)$ e dunque la tesi.

La proprietà (4) ci dice che la funzione di autocovarianza di un processo stazionario è una funzione pari di h , cioè è simmetrica rispetto all'origine.

Abbiamo infatti

$$\begin{aligned} \gamma_X(h) &= \text{Cov}(X_t, X_{t+h}) = E[X_t X_{t+h}] \\ &= E[X_{t-h} X_t] = \text{Cov}(X_{t-h}, X_t) = \gamma_X(-h). \end{aligned}$$

□

Proposizione 1.2.2. *Sia $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ una funzione pari tale che, per ogni n e per ogni (t_1, t_2, \dots, t_n) in \mathbb{Z}^n , la matrice $\Gamma_n = \{\gamma(t_i - t_j)\}_{i,j=1,\dots,n}$ sia semi-definita positiva. Allora esiste un processo stocastico gaussiano stazionario che ha γ come funzione di autocovarianza.*

Dimostrazione. Utilizziamo il teorema di Kolmogorov (1.1.1). Per ogni intero positivo n e per ogni vettore $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{Z}^n$ con $t_1 < t_2 < \dots < t_n$, sia $F_{\mathbf{t}}$ la funzione di ripartizione su \mathbb{R}^n con funzione caratteristica

$$\phi_{\mathbf{t}}(\mathbf{u}) = \exp \left(\frac{-\mathbf{u}' K \mathbf{u}}{2} \right),$$

dove $\mathbf{u} = (u_1, \dots, u_n)' \in \mathbb{R}^n$ e $K = \{\gamma(t_i - t_j)\}_{i,j=1,\dots,n}$.

Per ipotesi la matrice K è semidefinita positiva, quindi $\phi_{\mathbf{t}}$ risulterà essere la funzione caratteristica di una distribuzione normale n -variata, con media zero e matrice di covarianza K . Grazie al teorema di Kolmogorov (1.1.1), poiché vale

$$\phi_{\mathbf{t}(i)}(\mathbf{u}(i)) = \lim_{u_i \rightarrow 0} \phi_{\mathbf{t}}(\mathbf{u}), \quad \forall t \in \mathcal{T},$$

esiste un processo stocastico $(X_t)_{t \in \mathbb{Z}}$ con funzioni di ripartizione finito-dimensionali $F_{\mathbf{t}}$. In particolare, la distribuzione congiunta di X_i e X_j risulta una normale bivariata, con media zero e matrice di covarianza

$$\begin{pmatrix} \gamma(0) & \gamma(i-j) \\ \gamma(i-j) & \gamma(0) \end{pmatrix},$$

da cui segue che $Cov(X_i, X_j) = \gamma(i-j)$. □

Osservazione 1.2. Osserviamo che dalla Proposizione (1.2.2) otteniamo che per ogni funzione di autocovarianza $\gamma(\cdot)$, esiste un processo gaussiano stazionario che ha come funzione di autocovarianza proprio $\gamma(\cdot)$.

Le proprietà della funzione di autocorrelazione sono:

1. $\rho_X(0) = 1$;
2. $|\rho_X(h)| \leq 1 \quad \forall h \in \mathbb{Z}$;
3. $\rho_X(h) = \rho_X(-h) \quad \forall h \in \mathbb{Z}$;
4. $\rho_X(h) \geq 0$.

1.2.2 La funzione di autocovarianza campionaria e di autocorrelazione campionaria

Partendo da una realizzazione finita di ampiezza n , $\{x_1, x_2, \dots, x_n\}$, di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$, vogliamo poter dare una stima della funzione di autocovarianza $\gamma_X(\cdot)$ al fine di costruire un appropriato modello matematico per l'insieme dei dati. A questo riguardo diamo la seguente definizione.

Definizione 1.10. Definiamo la funzione di *autocovarianza campionaria* di $\{x_1, x_2, \dots, x_n\}$ come

$$\widehat{\gamma}_n(h) := \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x}_n)(x_j - \bar{x}_n), \quad 0 \leq h < n,$$

e $\widehat{\gamma}_n(h) = \widehat{\gamma}_n(-h)$, $-n < h \leq 0$, dove \bar{x}_n è la *media campionaria* $\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j$.

In modo analogo, definiamo la funzione di *autocorrelazione campionaria* come

$$\widehat{\rho}_n(h) := \frac{\widehat{\gamma}_n(h)}{\widehat{\gamma}_n(0)}, \quad |h| < n.$$

1.2.3 La funzione di autocorrelazione parziale

La funzione di autocorrelazione parziale contiene importanti informazioni sulla dipendenza fra le variabili aleatorie di un processo stazionario. Come la funzione di autocorrelazione, anche la funzione di autocorrelazione parziale dipende solo dalle proprietà del secondo ordine del processo. L'autocorrelazione parziale $\alpha(k)$ al passo k può essere considerata come la correlazione tra X_1 e X_{k+1} tenendo conto delle osservazioni intermedie X_2, \dots, X_k . Più precisamente, abbiamo la seguente definizione.

Definizione 1.11. La *funzione di autocorrelazione parziale* $\alpha_X(\cdot)$ di un processo stazionario $X = (X_t)_{t \in \mathbb{Z}}$ è definita come

$$\alpha_X(1) = \text{Corr}(X_2, X_1) = \rho_X(1),$$

$$\alpha_X(k) = \text{Corr}(X_{k+1} - P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_1), \quad k \geq 2,$$

dove con $P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_{k+1}$ e $P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_1$ denotiamo la proiezione ortogonale di X_{k+1} e X_1 sul sottospazio chiuso generato da $\{1, X_2, \dots, X_k\}$. Il valore $\alpha_X(k)$ è detto *autocorrelazione parziale al passo k* .

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero e funzione di autocovarianza $\gamma_X(\cdot)$ tale che $\gamma_X(h) \rightarrow 0$ quando $h \rightarrow \infty$. Supponiamo che le costanti

ϕ_{nj} , con $j = 1, \dots, n$ e $n \geq 1$, siano i coefficienti della rappresentazione

$$P_{\overline{sp}\{X_1, \dots, X_n\}} X_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}.$$

Allora dalle equazioni

$$\langle X_{n+1} - P_{\overline{sp}\{X_1, \dots, X_n\}} X_{n+1}, X_j \rangle = 0, \quad j = n, \dots, 1,$$

otteniamo

$$\begin{pmatrix} \rho_X(0) & \rho_X(1) & \rho_X(2) & \dots & \rho_X(n-1) \\ \rho_X(1) & \rho_X(0) & \rho_X(1) & \dots & \rho_X(n-2) \\ \vdots & & & & \vdots \\ \rho_X(n-1) & \rho_X(n-2) & \rho_X(n-3) & \dots & \rho_X(0) \end{pmatrix} \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} \rho_X(1) \\ \rho_X(2) \\ \vdots \\ \rho_X(n) \end{pmatrix}, \quad (1.4)$$

con $n \geq 1$. L'autocorrelazione parziale $\alpha_X(n)$ di $(X_t)_{t \in \mathbb{Z}}$ al passo n coincide con ϕ_{nn} , per $n \geq 1$, dove ϕ_{nn} è univocamente determinato dalle equazioni (1.4).

Definizione 1.12. La funzione di *autocorrelazione parziale campionaria* $\hat{\alpha}(k)$ di $\{x_1, \dots, x_n\}$ al passo k è definita, purché $x_i \neq x_j$ per qualche i e j , da

$$\hat{\alpha}_n(k) = \hat{\phi}_{kk}^{(n)} \quad 1 \leq k < n,$$

dove $\hat{\phi}_{kk}^{(n)}$ è univocamente determinata dalla (1.4), dove ogni $\rho_X(j)$ è sostituito dalla corrispondente autocorrelazione campionaria $\hat{\rho}_n(j)$.

1.3 La decomposizione di una serie storica

Come già detto precedentemente, l'obiettivo principale dell'analisi delle serie storiche è quello di individuare un processo stocastico che fornisca un modello matematico opportuno per i dati. Il passo iniziale è quindi analizzare la serie storica per valutare la presenza o meno della stazionarietà e individuare delle opportune trasformazioni per renderla stazionaria. La non stazionarietà di una serie storica può essere ricondotta alla presenza di tre fattori:

1. di un trend;
2. del fenomeno della stagionalità;
3. di variazioni delle ampiezze e del livello delle oscillazioni al variare del tempo.

Basta che la serie presenti almeno una delle tre caratteristiche elencate per essere non stazionaria.

In questo paragrafo saranno descritti metodi per ricondursi da una serie storica non stazionaria ad una stazionaria.

Nell'analisi di una serie storica il primo passo consiste nel tracciare il grafico dell'insieme dei dati a disposizione. In questo modo possiamo dare una prima valutazione sulla presenza di salti, sulla rottura di segmenti o sulla presenza di dati periferici. L'esame del grafico dei dati di una determinata serie storica suggerisce la possibilità di rappresentare i dati come realizzazione di un processo (modello di "decomposizione classica"):

$$X_t = m_t + s_t + Y_t, \quad (1.5)$$

dove m_t è una funzione che indica di quanto i valori cambiano in media, detta 'trend'; s_t è una funzione che descrive un comportamento periodico, cioè un comportamento che si ripete ad intervalli di tempo di ampiezza d , detta 'componente stagionale'; mentre Y_t è un processo stocastico stazionario chiamato 'rumore'.

L'obiettivo è quello di stimare ed estrarre, dalla serie storica, le componenti deterministiche m_t e s_t , nella speranza che il processo rumore Y_t che rimane sia un processo stocastico stazionario. Fatto questo, è possibile usare la teoria dei processi stazionari per trovare un modello probabilistico per il processo Y_t , analizzare le sue proprietà, e usare questo in concomitanza con il trend m_t e la stagionalità s_t per fare previsioni e controllare X_t .

Un diverso approccio nell'analisi delle serie storiche, sviluppato da Box e Jenkins che nel 1970 lo presentarono nella loro opera "Time series analysis:

forecasting and control”, consiste nell’applicare più volte delle trasformazioni “alle differenze” sui dati x_t , fino ad ottenere qualcosa che somigli ad una realizzazione di un processo stazionario. Fatto questo, si utilizza la teoria dei processi stazionari per modellare, analizzare e prevedere il processo ottenuto. Nei prossimi paragrafi illustreremo i due approcci con alcuni esempi.

1.3.1 Eliminazione del trend in assenza di stagionalità

In assenza della componente stagionale, il modello (1.5) diventa

$$X_t = m_t + Y_t \quad t = 1, \dots, n \quad (1.6)$$

dove possiamo supporre $E[Y_t] = 0$, senza ledere la generalità.

Metodo 1 (Stima dei minimi quadrati di m_t). Supponiamo che

$$m_t = a_0 + a_1 t + a_2 t^2, \quad (1.7)$$

e stimiamo a_0, a_1, a_2 minimizzando $\sum_t (x_t - m_t)^2$.

Esempio 3. (Popolazione degli USA, 1790-1980)

t	x_t	t	x_t
1790	3929214	1890	62979766
1800	5308483	1900	76212168
1810	7239881	1910	92228496
1820	9638453	1920	106021537
1830	12860702	1930	123202624
1840	17063353	1940	132164569
1850	23191876	1950	151325798
1860	31443321	1960	179323175
1870	38558371	1970	203302031
1880	50189209	1980	226545805

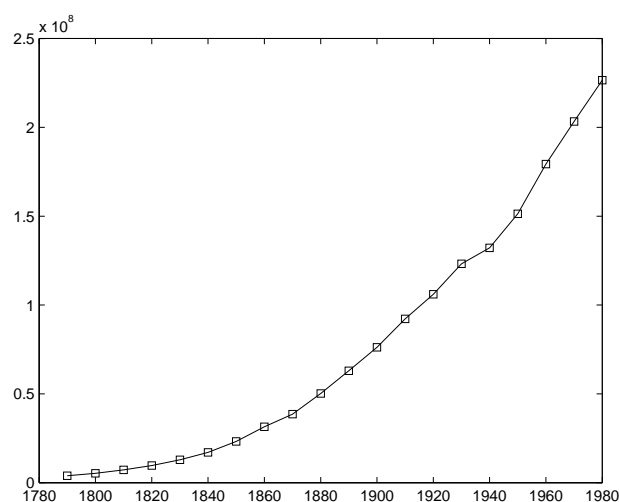


Figura 1.1: Popolazione degli USA, 1790-1980.

Adattando la funzione (1.7) ai dati relativi alla popolazione USA nel periodo 1780 – 1980, diamo una stima dei parametri

$$\hat{a}_0 = 2,097911 \times 10^{10},$$

$$\hat{a}_1 = -2,334962 \times 10^7,$$

$$\hat{a}_2 = 6.498591 \times 10^3.$$

In Figura 1.2 riportiamo il grafico dei dati originali con la parabola ottenuta approssimando m_t con il metodo dei minimi quadrati.

I valori stimati del processo rumore Y_t si ottengono sottraendo dai dati x_t le quantità

$$\hat{m}_t = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2.$$

Dal valore del trend m_t otteniamo una stima dei valori futuri di X_t . Ad esempio se stimiamo Y_{1990} con il suo valor medio, cioè zero, otteniamo la seguente stima

$$\hat{m}_{1990} = 2,484 \times 10^8,$$

per la popolazione USA nell'anno 1990.

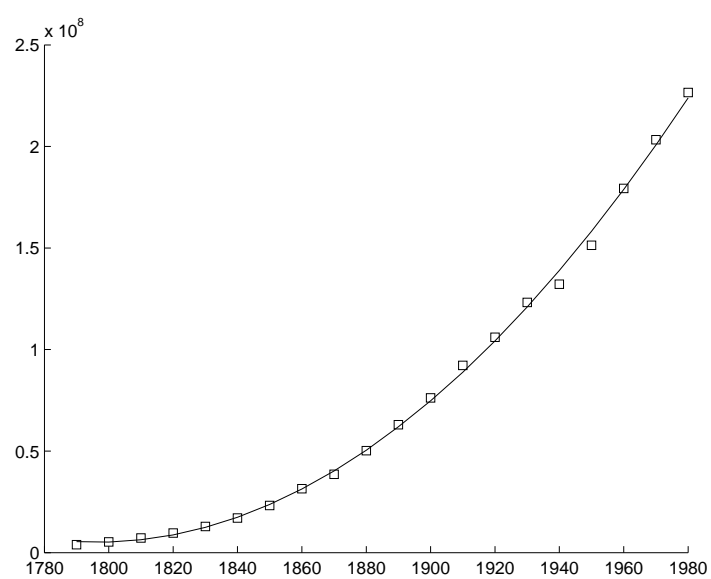


Figura 1.2: Popolazione degli USA, 1790-1980, approssimata con i minimi quadrati.

Metodo 2 (Regolarizzazione con media mobile). Sia q un intero non negativo e consideriamo la media mobile

$$W_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t+j}$$

del processo $(X_t)_{t \in \mathbb{Z}}$, definito dalla (1.6). Allora per $q + 1 \leq t \leq n - q$ otteniamo

$$W_t = (2q + 1)^{-1} \sum_{j=-q}^q m_{t+j} + (2q + 1)^{-1} \sum_{j=-q}^q Y_{t+j} \simeq m_t,$$

assumendo che m_t può essere approssimata da una funzione lineare su $[t - q, t + q]$ e che, su questo intervallo, la media del processo rumore sia vicino a zero.

In questo modo la media mobile ci permette di stimare

$$\hat{m}_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t+j}, \quad q + 1 \leq t \leq n - q. \quad (1.8)$$

Poiché il processo X_t non è osservato per $t \leq 0$ o $t > n$, non possiamo usare la (1.8) per $t \leq q$ o $t > n - q$. In questo caso possiamo calcolare \hat{m}_t come

$$\hat{m}_t = \sum_{j=0}^{n-t} \alpha(1 - \alpha)^j X_{t+j}, \quad t = 1, \dots, q,$$

e

$$\hat{m}_t = \sum_{j=0}^{t-1} \alpha(1 - \alpha)^j X_{t-j}, \quad t = n - q, \dots, n.$$

Queste ultime stime dipendono dal valore di α . Sperimentalmente è stato trovato che una stima di α tra 0,1 e 0,3 dà una stima ragionevole.

Esempio 4. Scioperi negli USA, 1951-1980

t	x_t	t	x_t	t	x_t
1951	4737	1961	3367	1971	5138
1952	5117	1962	3614	1972	5010
1953	5091	1963	3362	1973	5353
1954	3468	1964	3655	1974	6074
1955	4320	1965	3963	1975	5031
1956	3825	1966	4405	1976	5648
1957	3673	1967	4595	1977	5506
1958	3624	1968	5045	1978	4230
1959	3708	1969	5700	1979	4827
1960	3333	1970	5716	1980	3885

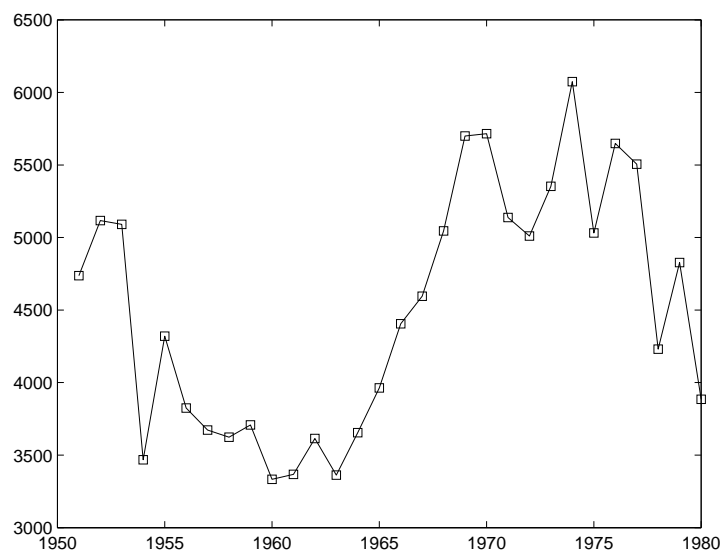


Figura 1.3: Scioperi negli USA, 1951-1980.

Applichiamo le stime per il trend a questa serie storica prendendo $q = 2$ e $\alpha = 0,3$, e otteniamo il grafico di m_t

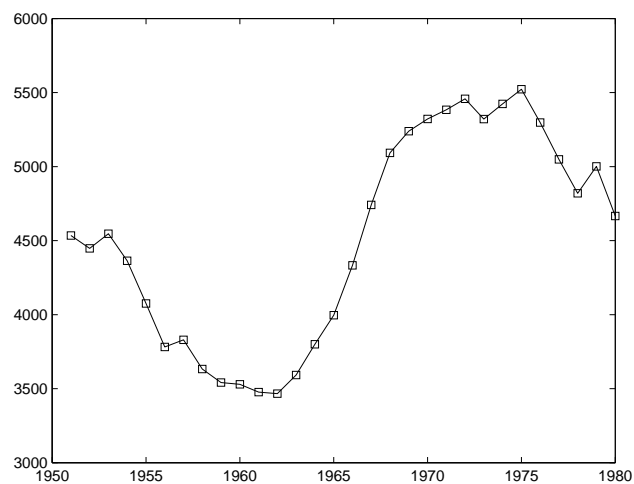


Figura 1.4: Media mobile sul trend m_t .

La stima del processo rumore è ottenuta come $\hat{Y}_t = x_t - \hat{m}_t$, ed è riportata nel seguente grafico

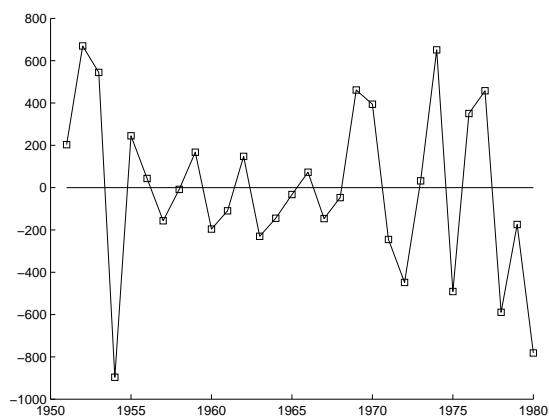


Figura 1.5: Residui sulla serie storica degli scioperi.

Metodo 3 (Differenza di ordine k per generare dati stazionari). Definiamo il primo operatore alle differenze ∇ come

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t, \quad (1.9)$$

dove B è chiamato *operatore ritardo*, in inglese “backward”, ed è definito da

$$BX_t = X_{t-1}.$$

Le potenze degli operatori B e ∇ sono definite come $B^j(X_t) = X_{t-j}$ e $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, per $j \geq 1$, con $B^0 = 1$ e $\nabla^0(X_t) = X_t$. Notiamo che polinomi negli operatori B e ∇ si comportano esattamente come polinomi a variabili reali. Per esempio:

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

Se l'operatore ∇ è applicato ad una funzione lineare che stima il trend, ossia $m_t = at + b$, allora otteniamo una funzione costante $\nabla m_t = a$. Mentre se il trend risulta stimato da un polinomio di grado k , allora possiamo applicare l'operatore ∇^k per ricondurlo ad una costante.

Partendo dal modello $X_t = m_t + Y_t$, e supponendo che $m_t = \sum_{j=0}^k a_j t^j$ e Y_t sia un processo stazionario con media zero, otteniamo

$$\nabla^k X_t = k! a_k + \nabla^k Y_t,$$

che risulta essere un processo stazionario.

Queste considerazioni suggeriscono che partendo da un insieme di dati (x_t) possiamo applicare l'operatore ∇ finché non otteniamo un insieme $(\nabla^k x_t)$ che possiamo modellare come una realizzazione di un processo stazionario.

Nella pratica l'ordine k dell'operatore alle differenze è abbastanza piccolo, molto spesso 1 o 2. Questo dipende dal fatto che molte funzioni possono essere ben approssimate, su un intervallo di lunghezza finita, da un polinomio con grado basso.

Esempio 5. (Popolazione degli USA, 1790-1980)

Applicando il metodo appena descritto ai valori x_n dell'esempio 2, troviamo che un operatore alle differenze di ordine 2 è sufficiente a produrre una serie storica senza la presenza del trend.

I dati ottenuti dopo aver applicato l'operatore alle differenze

$$\nabla^2 x_n = x_n - 2x_{n-1} + x_{n-2}$$

sono riportati in Figura 1.6.

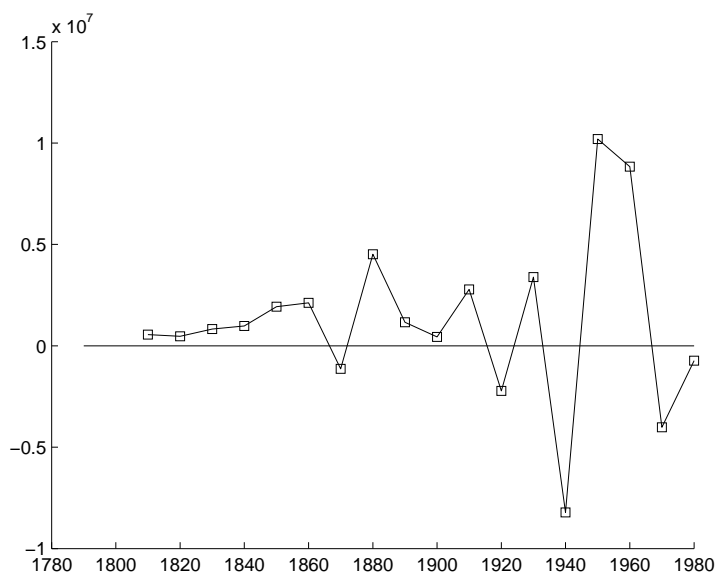


Figura 1.6: Serie con operatore alle differenze del secondo ordine.

Notiamo che l'ampiezza di fluttuazione di ∇_n^x cresce con il valore di x_n . Questo effetto può essere eliminato se applichiamo l'operatore alle differenze alla serie $y_t = \ln x_n$.

1.3.2 Eliminazione del trend e della stagionalità

I metodi descritti precedentemente per eliminare il trend, possono essere adattati per eliminare anche la componente stagionale. Supponiamo

$$X_t = m_t + s_t + Y_t$$

con $E[Y_t] = 0$, $s_{t+d} = s_t$ e $\sum_{j=1}^d s_j = 0$.

Nella descrizione dei seguenti metodi adotteremo la notazione:

- $k = 1, \dots, d$ per indicare l'indice di stagionalità, dove d è il periodo della componente stagionale;
- $j = 1, \dots, p$ dove p è l'ampiezza temporale di considerazione dei dati.

Poniamo $x_{j,k} = x_{k+d(j-1)}$.

Metodo S1 (Metodo del trend lento). Nel caso di un trend piccolo, senza ledere la generalità, possiamo supporlo costante. Partendo dall'ipotesi $\sum_{j=1}^d s_j = 0$, possiamo stimare il trend sul periodo di ampiezza j con

$$\hat{m}_j = \frac{1}{d} \sum_{k=1}^d x_{j,k}, \quad (1.10)$$

mentre possiamo stimare s_k , per $k = 1, \dots, d$, con

$$\hat{s}_k = \frac{1}{p} \sum_{j=1}^p (x_{j,k} - \hat{m}_j). \quad (1.11)$$

Osserviamo che la condizione $\sum_{k=1}^d \hat{s}_k = 0$ è verificata. Dunque otteniamo la stima

$$\hat{Y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k, \quad j = 1, \dots, p, \quad k = 1, \dots, d. \quad (1.12)$$

Esempio 6. (Morti accidentali mensili negli USA, 1973-1978)

	1973	1974	1975	1976	1977	1978
Gen.	9007	7750	8162	7717	7792	7836
Feb.	8106	6981	7306	7461	6957	6892
Mar.	8928	8038	8124	7776	7726	7791
Apr.	9137	8422	7870	7925	8106	8129
Mag.	10017	8714	9387	8634	8890	9115
Giu.	10826	9512	9556	8945	9299	9434
Lug.	11317	10120	10093	10078	10625	10484
Ago.	10744	9823	9620	9179	9302	9827
Set.	9713	8743	8285	8037	8314	9110
Ott.	9938	9129	8433	8488	8850	9070
Nov.	9161	8710	8160	7874	8265	8633
Dic.	8927	8680	8034	8647	8796	9240

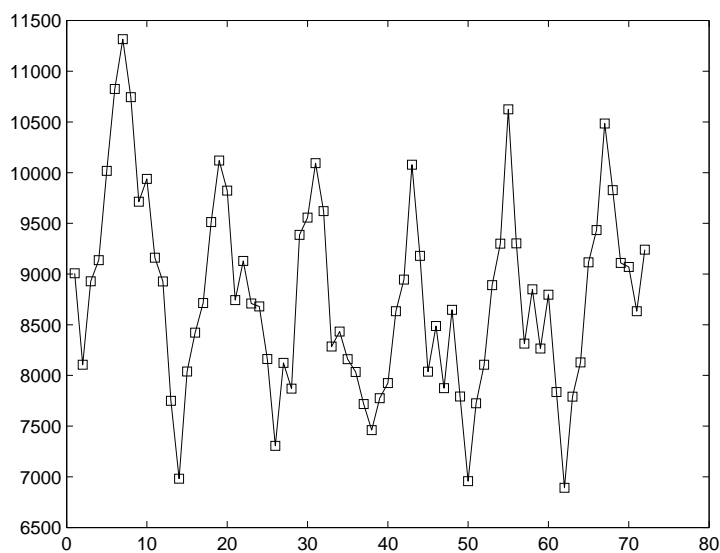


Figura 1.7: Morti accidentali mensili negli USA, 1973-1978.

In questo esempio il periodo d della componente stagionale è 12, mentre l'ampiezza temporale p è pari a 6 anni. Con $x_j, k, j = 1, \dots, 6$ e $k = 1, \dots, 12$, denotiamo il numero di morti accidentali avute nel k -esimo mese del j -esimo anno, $(1972 + j)$, cioè

$$x_{j,k} = x_{k+12(j-1)}.$$

Quindi applicando le formule (1.10)-(1.12) otteniamo l'eliminazione del trend dalle osservazioni $x_{j,k} - \hat{m}_j$, la stima della componente stagionale \hat{s}_k e l'eliminazione del trend e la destagionalizzazione dalle osservazioni

$$\hat{Y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k,$$

riportate rispettivamente nei seguenti grafici.

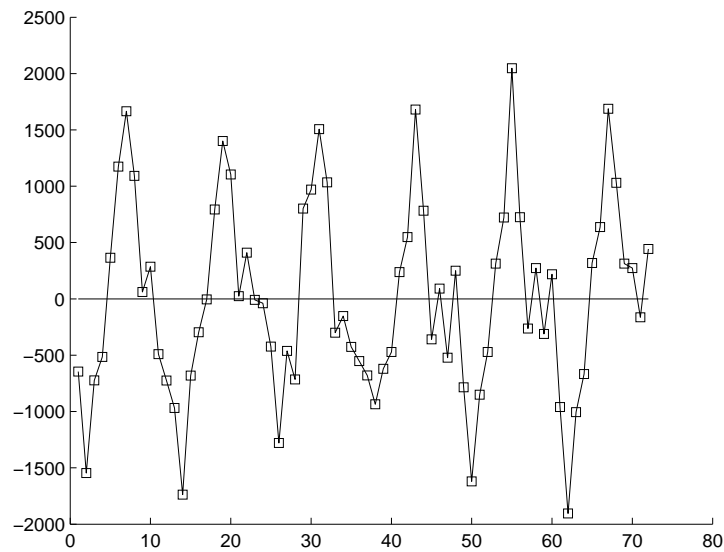


Figura 1.8: Morti accidentali mensili dopo aver sottratto il trend.

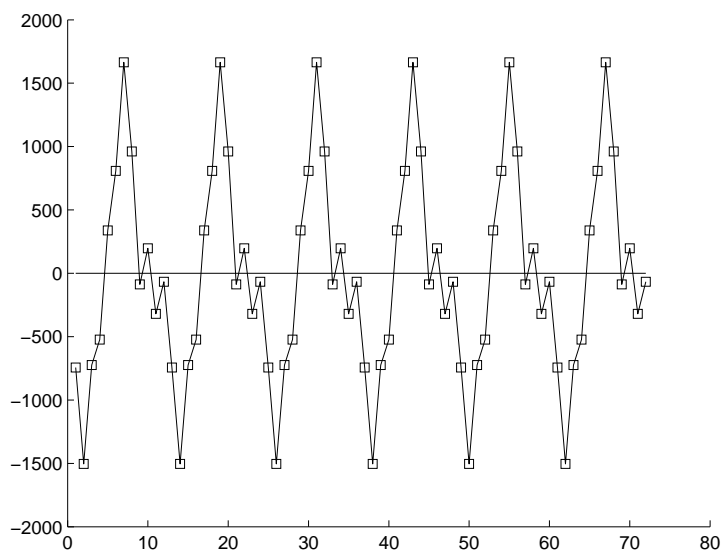


Figura 1.9: Componente stagionale delle morti accidentali mensili.

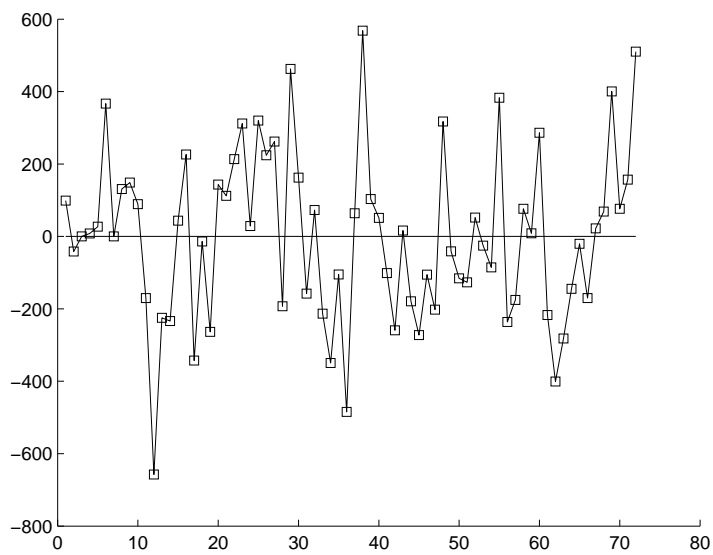


Figura 1.10: Eliminazione del trend e della stagionalità .

Metodo S2 (Stima con media mobile). L'uso di questa tecnica è preferibile rispetto a quella descritta nel metodo S1, perché non viene fatta l'ipotesi che m_t sia costante.

Supponiamo di avere un insieme di osservazioni $\{x_1, \dots, x_n\}$ e stimiamo il trend in modo da eliminare la componente stagionale. La media mobile non è altro che la media aritmetica di n osservazioni successive della serie. Se il numero dei termini è dispari, la media mobile è automaticamente centrata, nel senso che corrisponde al centro dell'intervallo; mentre se il numero dei termini è pari, la media mobile si può centrare calcolandola su $n + 1$ termini ponderati con i pesi $(\frac{1}{2}, 1, 1, 1, \dots, 1, 1, 1, \frac{1}{2})$. Se il periodo d è pari, cioè $d = 2q$, facciamo quindi la stima

$$\widehat{m}_t = \frac{1}{d}(0.5x_{t-q} + x_{t-q+1} + \dots + x_{t+q-1} + 0.5x_{t+q}), \quad q < t \leq n - q.$$

Se il periodo d è dispari, cioè $d = 2q + 1$, stimiamo il trend come nel metodo 2, cioè con la media mobile

$$\widehat{m}_t = \frac{1}{d} \sum_{j=-q}^q x_{t+j}, \quad q < t \leq n - q.$$

Esempio 7. (Morti accidentali mensili negli USA, 1973-1978)

In Figura 1.11 riportiamo la stima del trend m_t , $6 < t \leq 66$, nella serie delle morti accidentali negli USA 1972-1978, confrontandole con la stima del trend ottenuta nel metodo S1 attraverso la media aritmetica.

Il secondo passo consiste nello stimare la componente stagionale. Per ogni $k = 1, \dots, d$, calcoliamo la media delle deviazioni

$$w_k = \frac{1}{p} \sum_{j=1}^p (x_{j,k} - \widehat{m}_{j,k}).$$

Notiamo che non è detto che si abbia $\sum_{i=1}^d w_i = 0$. Stimiamo quindi la componente stagionale s_k con $\widehat{s}_k = \widehat{s}_{k-d}$ per $k > d$, e

$$\widehat{s}_k = w_k - \frac{1}{d} \sum_{i=1}^d w_i, \quad k = 1, \dots, d.$$

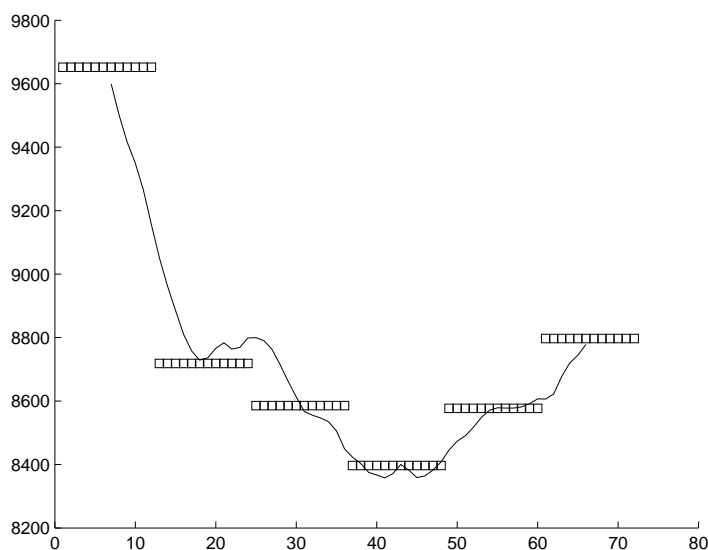


Figura 1.11: Confronto tra media mobile e media aritmetica del trend.

A questo punto facciamo un'altra stima del trend sui dati destagionalizzati

$$d_t = x_t - \hat{s}_t, \quad t = 1, \dots, n,$$

applicando uno dei metodi descritti precedentemente.

La stima del processo rumore sarà data, come nel metodo S1, da

$$\hat{Y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k, \quad t = 1, \dots, n.$$

Metodo S3 (Differenza di ordine d). La tecnica descritta nel metodo 3, per insiemi di dati non-stagionali, può essere adattata a dati che presentano una componente stagionale di periodo d , introducendo l'operatore alle differenze ∇_d di ordine d , definito come

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t. \quad (1.13)$$

Questo operatore non va confuso con l'operatore $\nabla^d = (1 - B)^d$ definito nel metodo 3. Applicando l'operatore ∇_d al modello

$$X_t = m_t + s_t + Y_t,$$

dove s_t ha periodo d , ricaviamo

$$\nabla_d X_t = (m_t - m_{t-d}) + (Y_t - Y_{t-d}).$$

Otteniamo quindi un processo destagionalizzato, dove $(m_t - m_{t-d})$ è il trend e $(Y_t - Y_{t-d})$ è il processo rumore. Notiamo che il trend di questo nuovo processo può essere eliminato usando i metodi descritti precedentemente, in modo da ottenere un processo stazionario.

Esempio 8. (Morti accidentali mensili negli USA, 1973-1978)

Nel grafico seguente riportiamo i risultati ottenuti dopo che abbiamo applicato l'operatore ∇_{12} alle serie delle morti accidentali. Notiamo che la componente stagionale è assente dal grafico di $\nabla_{12}x_t$, $t = 13, \dots, 72$. Però è ancora presente un trend crescente, che possiamo eliminare applicando un'altra volta l'operatore ∇ a $\nabla_{12}x_t$, $t = 14, \dots, 72$ e il risultato è riportato in Figura 1.13. Osserviamo che in questo caso non è presente né la componente stagionale né il trend.

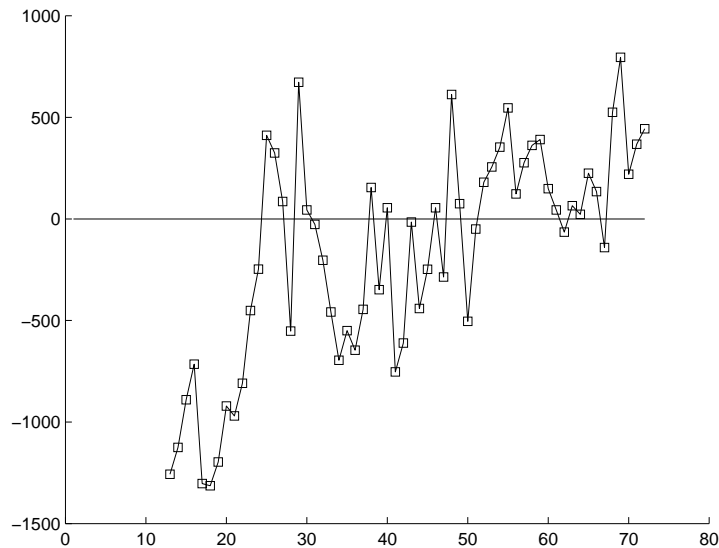


Figura 1.12: Serie differenziata $\{\nabla x_t, t = 13, \dots, 72\}$.

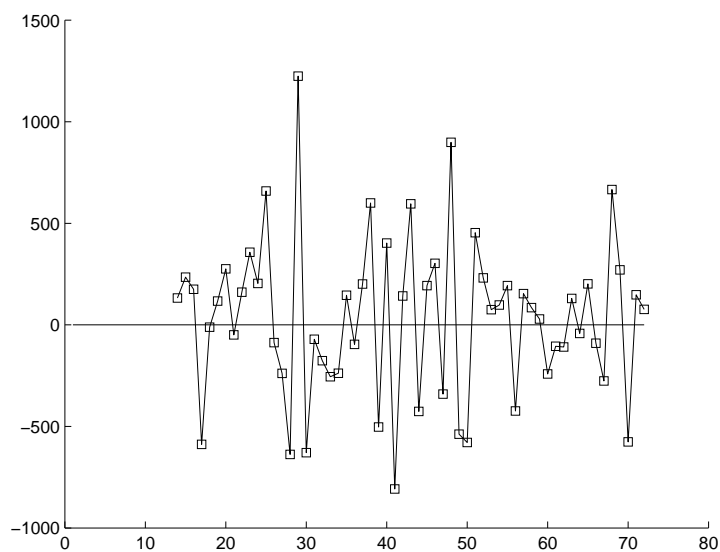


Figura 1.13: Serie differenziata $\{\nabla\nabla x_t, t = 14, \dots, 72\}$.

In questo paragrafo abbiamo dunque descritto vari metodi per stimare e/o rimuovere il trend e la stagionalità. La scelta di un metodo o di un altro dipende dalle caratteristiche presenti nella serie storica da analizzare.

Capitolo 2

I processi ARMA

In questo capitolo introduciamo una classe di processi stocastici che giocano un ruolo chiave nella modellizzazione di una serie storica: i processi ARMA. Una ragione del loro successo è che, data una qualsiasi funzione di autocovarianza $\gamma(\cdot)$ tale che $\lim_{h \rightarrow \infty} \gamma(h) = 0$ e dato un intero $k > 0$, è sempre possibile trovare un processo ARMA $X = (X_t)_{t \in \mathbb{Z}}$ con funzione di autocovarianza $\gamma_X(\cdot)$ tale che $\gamma_X(h) = \gamma(h)$ per $h = 0, 1, \dots, k$.

Prima di analizzare le caratteristiche principali di tali processi diamo alcune definizioni di base.

2.1 Il rumore bianco

Definizione 2.1. Un processo $(Z_t)_{t \in \mathbb{Z}}$ è detto *rumore bianco* (in inglese, *white noise*) con media nulla e varianza σ^2 , se è un processo stazionario con $E[Z_t] = 0$ per ogni t e funzione di autocovarianza data da

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

Nel seguito, scriveremo

$$(Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2).$$

Se $(Z_t)_{t \in \mathbb{Z}}$ è anche un processo gaussiano, allora esso prende il nome di *rumore bianco gaussiano* e scriveremo

$$(Z_t)_{t \in \mathbb{Z}} \sim GWN(0, \sigma^2).$$

Osserviamo che, se $(Z_t)_{t \in \mathbb{Z}}$ è un rumore bianco gaussiano, allora $(Z_t)_{t \in \mathbb{Z}}$ è una famiglia di variabili aleatorie indipendenti con legge normale $N(0, \sigma^2)$.

2.2 I processi MA ed AR

Denotiamo con $(Z_t)_{t \in \mathbb{Z}}$ un rumore bianco con media nulla e varianza σ^2 , ossia $(Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2)$. Richiamiamo l'operatore ritardo B definito nel paragrafo 1.3: $BX_t = X_{t-1}$.

Definizione 2.2. Un processo $(X_t)_{t \in \mathbb{Z}}$ è detto *processo a media mobile* di ordine q , con q in \mathbb{N} , più brevemente MA(q) (MA sta appunto per *Moving Average*), se è della forma

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} = \sum_{i=0}^q \theta_i Z_{t-i} = \theta(B)Z_t,$$

dove $\theta(B)$ è il polinomio di grado q , dato da

$$\theta(z) = \theta_0 + \theta_1 z + \dots + \theta_q z^q,$$

con θ_i costanti reali.

Senza ledere la generalità possiamo prendere $\theta(0) = \theta_0 = 1$.

Il suo momento primo risulta

$$E[X_t] = E \left[\sum_{i=0}^q \theta_i Z_{t-i} \right] = \sum_{i=0}^q \theta_i E[Z_{t-i}] = 0.$$

Un processo MA ha quindi media zero. Il fatto che il momento primo sia nullo ci permette di scrivere la varianza come il momento secondo, ossia

$$Var(X_t) = E[X_t^2] = E \left[\left(\sum_{i=0}^q \theta_i Z_{t-i} \right)^2 \right].$$

Sviluppando il quadrato, possiamo scomporre la somma precedente in due parti:

$$\left(\sum_{i=0}^q \theta_i Z_{t-i} \right)^2 = \sum_{i=0}^q \theta_i^2 Z_{t-i}^2 + \sum_{i=0}^q \sum_{j \neq i} \theta_i \theta_j Z_{t-i} Z_{t-j}.$$

Dalle proprietà del rumore bianco segue che il valore atteso della seconda sommatoria è nullo, cosicché

$$E[X_t^2] = E \left[\sum_{i=0}^q \theta_i^2 Z_{t-i}^2 \right] = \sum_{i=0}^q \theta_i^2 E[Z_{t-i}^2] = \sum_{i=0}^q \theta_i^2 \sigma^2 = \sigma^2 \sum_{i=0}^q \theta_i^2 < +\infty.$$

Calcoliamo ora la funzione di autocovarianza di ordine k :

$$\begin{aligned} \gamma_X(k) &= \text{Cov}(X_t, X_{t+k}) \\ &= \text{Cov}(Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, Z_{t+k} + \theta_1 Z_{t+k-1} + \dots + \theta_q Z_{t+k-q}) \\ &= \begin{cases} 0 & k > q, \\ \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k} & k = 0, 1, \dots, q, \\ \gamma_X(-k) & k < 0. \end{cases} \end{aligned}$$

Osserviamo che γ_X non dipende da t e quindi un processo a media mobile risulta essere stazionario. Vale la seguente proposizione.

Proposizione 2.2.1. *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero e con funzione di autocovarianza $\gamma_X(h)$ tale che $\gamma_X(h) = 0$ per $|h| > q$ e $\gamma_X(q) \neq 0$. Allora $(X_t)_{t \in \mathbb{Z}}$ è un processo MA(q), ossia esiste un rumore bianco $(Z_t)_{t \in \mathbb{Z}}$ tale che*

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}.$$

Dimostrazione. Per ogni t , definiamo il sottospazio chiuso \mathcal{M}_t di L^2 generato da $\{X_s, -\infty < s \leq t\}$ e poniamo

$$Z_t = X_t - P_{\mathcal{M}_{t-1}} X_t, \tag{2.1}$$

dove denotiamo con $P_{\mathcal{M}_{t-1}} X_t$ la proiezione ortogonale di X_t sul sottospazio \mathcal{M}_{t-1} . Osserviamo che $Z_t \in \mathcal{M}_t$ e $Z_t \in \mathcal{M}_{t-1}^\perp$. Perciò, se $s < t$, allora

$Z_s \in \mathcal{M}_s \subset \mathcal{M}_{t-1}$ e quindi $E[Z_s Z_t] = 0$.

Inoltre, si ha

$$P_{\overline{\text{sp}}\{X_s, s=t-n, \dots, t-1\}} X_t \xrightarrow{L^2} P_{\mathcal{M}_{t-1}} X_t, \quad n \rightarrow \infty.$$

Di conseguenza grazie alla stazionarietà e alla continuità della norma in L^2 , vale

$$\begin{aligned} \|Z_{t+1}\| &= \|X_{t+1} - P_{\mathcal{M}_t} X_{t+1}\| \\ &= \lim_{n \rightarrow \infty} \|X_{t+1} - P_{\overline{\text{sp}}\{X_s, s=t-n, \dots, t\}} X_{t+1}\| \\ &= \lim_{n \rightarrow \infty} \|X_t - P_{\overline{\text{sp}}\{X_s, s=t-n, \dots, t-1\}} X_t\| \\ &= \|X_t - P_{\mathcal{M}_{t-1}} X_t\| = \|Z_t\|. \end{aligned}$$

Ponendo $\sigma^2 = \|Z_t\|^2$, possiamo concludere che $Z_t \sim WN(0, \sigma^2)$.

Dall'uguaglianza (2.1), otteniamo

$$\begin{aligned} \mathcal{M}_{t-1} &= \overline{\text{sp}}\{X_s, s < t-1 < Z_{t-1}\} \\ &= \overline{\text{sp}}\{X_s, s < t-q, Z_{t-q}, \dots, Z_{t-1}\}. \end{aligned}$$

Possiamo quindi decomporre \mathcal{M}_{t-1} in due sottospazi ortogonali: \mathcal{M}_{t-q-1} e $\overline{\text{sp}}\{Z_{t-q}, \dots, Z_{t-1}\}$. Dall'ipotesi $\gamma_X(h) = 0$ per $|h| > q$, segue che $X_t \perp \mathcal{M}_{t-q-1}$. Dalle proprietà della proiezione ortogonale (vedi Appendice A), abbiamo che

$$\begin{aligned} P_{\mathcal{M}_{t-1}} X_t &= P_{\mathcal{M}_{t-q-1}} X_t + P_{\overline{\text{sp}}\{Z_{t-q}, \dots, Z_{t-1}\}} X_t \\ &= 0 + \sigma^{-2} E[X_t Z_{t-1}] Z_{t-1} + \dots + \sigma^{-2} E[X_t Z_{t-q}] Z_{t-q} \\ &= \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \end{aligned}$$

dove $\theta_j := \sigma^{-2} E[X_t Z_{t-j}]$. Sostituendo $P_{\mathcal{M}_{t-1}} X_t$ nella (2.1), otteniamo la tesi. \square

La seguente definizione estende quella precedente al caso $q = +\infty$.

Definizione 2.3. Il processo $(X_t)_{t \in \mathbb{Z}}$ è detto *processo a media mobile di ordine ∞ o lineare*, più brevemente $\text{MA}(\infty)$, se esistono delle costanti ψ_j con $\sum_{j=0}^{\infty} |\psi_j| < \infty$ tali che

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}. \quad (2.2)$$

Osserviamo che $(X_t)_{t \in \mathbb{Z}}$ è un processo $\text{MA}(\infty)$ solo se ogni X_t può essere scritta in funzione del solo rumore bianco $(Z_s)_{s \in \mathbb{Z}}$ con $s \leq t$, ossia se possiamo scrivere il processo come una combinazione lineare infinita di variabili aleatorie del processo rumore bianco.

La condizione $\sum_{j=0}^{\infty} |\psi_j| < \infty$ assicura che la serie nell'equazione (2.2) converge con probabilità uno e in media quadratica. Inoltre, dalla Proposizione 2.3.2 del prossimo paragrafo, segue che, un processo $\text{MA}(\infty)$ risulta essere stazionario con media zero e funzione di autocovarianza

$$\gamma_X(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}.$$

Un'altra importante classe di processi è data dai processi *autoregressivi*.

Definizione 2.4. Un processo $(X_t)_{t \in \mathbb{Z}}$ è un *processo autoregressivo* di ordine p , con p in \mathbb{N} , più brevemente $\text{AR}(p)$ (AR sta appunto per *AutoRegressive*), se vale

$$\phi_0 X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t$$

ossia

$$Z_t = \phi_0 X_t - \sum_{i=1}^p \phi_i X_{t-i} = \phi(B) X_t,$$

dove $\phi(B)$ è il polinomio di grado p , dato da

$$\phi(z) = \phi_0 - \phi_1 z - \dots - \phi_p z^p,$$

con ϕ_i costanti reali.

Senza ledere la generalità assumiamo $\phi(0) = \phi_0 = 1$.

Notiamo che i processi AR sono in un certo senso complementari ai processi MA. Infatti un processo MA è definito dall'applicazione di un polinomio nell'operatore B ad un rumore bianco; mentre un processo AR è caratterizzato dalla proprietà che l'applicazione di un polinomio nell'operatore B ad esso produce un rumore bianco.

Per analizzare le caratteristiche dei processi autoregressivi, consideriamo il caso più semplice: quello in cui $p = 1$ e quindi vale la relazione

$$X_t = \phi_1 X_{t-1} + Z_t,$$

da cui, utilizzando l'operatore ritardo,

$$(1 - \phi_1 B)X_t = Z_t.$$

Supponiamo che il processo AR(1) sia stazionario con media pari a μ .

Innanzitutto osserviamo che

$$\mu = E[X_t] = \phi_1 E[X_{t-1}] + E[Z_t] = \phi_1 \mu.$$

Quest'uguaglianza può essere vera in due casi: o $\mu = 0$ e in tal caso è vera per qualsiasi valore di ϕ_1 ; oppure nel caso $\phi_1 = 1$ e in tal caso l'uguaglianza è vera per qualsiasi valore di μ . In quest'ultimo caso si dice che il processo presenta una *radice unitaria*, perché il valore di z per cui $\phi(z) = 0$ è $z = 1$. Per quanto riguarda i momenti secondi, se supponiamo $\mu = 0$ abbiamo

$$Var(X_t) = E[X_t^2] = E[(\phi_1 X_{t-1} + Z_t)^2] = \phi_1^2 Var(X_t) + \sigma^2 + 2\phi_1 E[X_{t-1}Z_t].$$

L'ultimo elemento della somma è zero in quanto X_{t-1} è della forma $\theta(B)Z_{t-1}$ e quindi $E[X_{t-1}Z_t]$ è una combinazione lineare di autocovarianze di un rumore bianco, che sono nulle per definizione. Se ne deduce che

$$Var(X_t) = \phi_1^2 Var(X_t) + \sigma^2, \text{ da cui } Var(X_t) = \frac{\sigma^2}{1 - \phi_1^2}.$$

Questo risultato è importante perché esclude dai processi AR(1) stazionari quelli con radice unitaria.

Adesso analizziamo le autocovarianze:

$$\gamma_X(0) = \text{Var}(X_t),$$

$$\gamma_X(1) = E[X_t X_{t-1}] = E[(\phi_1 X_{t-1} + Z_t) X_{t-1}] = \phi_1 \text{Var}(X_t),$$

in generale,

$$\gamma_X(h) = E[X_t X_{t-h}] = E[(\phi_1 X_{t-1} + Z_t) X_{t-h}] = \phi_1 \gamma(h-1),$$

da cui

$$\gamma_X(h) = \phi_1^h \frac{\sigma^2}{1 - \phi_1^2}.$$

Osservando l'espressione finale di γ_X vediamo che

- per $\phi_1 > 0$, $\gamma_X(h)$ è monotona decrescente;
- per $\phi_1 < 0$, $\gamma_X(h)$ è altalenante.

Se dobbiamo modellizzare dati la cui funzione di autocovarianza è decrescente, possiamo quindi utilizzare un modello AR(1) con $\phi_1 > 0$. Viceversa, se la funzione di autocovarianza è altalenante, possiamo utilizzare un modello AR(1) con $\phi_1 < 0$.

Le funzioni di autocorrelazioni sono date da

$$\rho_X(h) = \phi_1^h.$$

Anche la definizione di AR(p) può essere estesa al caso $p = +\infty$.

2.3 I processi ARMA

La classe dei processi ARMA comprende sia i processi AR che i processi MA come caso particolare.

Definizione 2.5. Un processo stazionario $(X_t)_{t \in \mathbb{Z}}$ è detto *autoregressivo a media mobile*, più brevemente ARMA(p,q), se esistono dei coefficienti $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ tali che per ogni t in \mathbb{Z} , vale

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (2.3)$$

dove $(Z_t)_{t \in \mathbb{Z}}$ è un rumore bianco di media zero e varianza σ^2 .

Diciamo che $(X_t)_{t \in \mathbb{Z}}$ è un processo ARMA(p,q) con media μ se $(X_t - \mu)_{t \in \mathbb{Z}}$ è un processo ARMA(p,q) .

Possiamo scrivere l'equazione (2.3) in forma compatta come

$$\phi(B)X_t = \theta(B)Z_t, \quad t \in \mathbb{Z}$$

dove ϕ e θ sono polinomi di grado p e q rispettivamente, dati da

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

e

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

e B è l'operatore ritardo.

Diamo adesso la definizione di processo ARMA(p,q) causale.

Definizione 2.6. Un processo ARMA(p,q) definito dall'equazione

$$\phi(B)X_t = \theta(B)Z_t$$

si dice *causale*, o *funzione causale* di $(Z_t)_{t \in \mathbb{Z}}$, se esistono delle costanti ψ_j tali che $\sum_{j=0}^{+\infty} |\psi_j| < \infty$ e

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}. \quad (2.4)$$

La proprietà di causalità non riguarda solo il processo $(X_t)_{t \in \mathbb{Z}}$ ma interessa, piuttosto, la relazione tra i processi $(X_t)_{t \in \mathbb{Z}}$ e $(Z_t)_{t \in \mathbb{Z}}$ coinvolti nella definizione di processo ARMA(p,q) .

La proposizione seguente chiarifica il senso dell'uguaglianza (2.4).

Proposizione 2.3.1. Sia $(X_t)_{t \in \mathbb{Z}}$ una successione di variabili aleatorie tali che $\sup_t E[|X_t|] < \infty$ e sia $(\psi_j)_{j \in \mathbb{Z}}$ una successione di costanti tale che $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Allora la serie

$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (2.5)$$

è assolutamente convergente con probabilità uno.

Inoltre, se $\sup_t E[|X_t|^2] < \infty$, allora la serie converge in media quadratica.

Dimostrazione. Dal teorema di convergenza monotona di Beppo Levi e dal fatto che $\sup_t E[|X_t|]$ è finito, otteniamo che

$$\begin{aligned} E \left[\sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}| \right] &= \lim_{n \rightarrow \infty} E \left[\sum_{j=-n}^n |\psi_j| |X_{t-j}| \right] \\ &\leq \lim_{n \rightarrow \infty} \left(\sum_{j=-n}^n |\psi_j| \right) \sup_t E[|X_t|] < \infty. \end{aligned}$$

Da questo segue che $\sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}|$ e $\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$ sono entrambe finite e convergono con probabilità uno.

Se $\sup_t E[|X_t|^2] < \infty$ e $n > m > 0$, allora abbiamo che

$$\begin{aligned} E \left[\left| \sum_{m < |j| \leq n} \psi_j X_{t-j} \right|^2 \right] &= \sum_{m < |j| \leq n} \sum_{m < |k| \leq n} \psi_j \bar{\psi}_k E[X_t - \bar{X}_{t-k}] \\ &\leq \sup_t E[|X_t|^2] \left(\sum_{m < |j| \leq n} |\psi_j| \right)^2 \rightarrow 0 \quad \text{per } m, n \rightarrow \infty, \end{aligned}$$

per il criterio di Cauchy la serie (2.5) converge in media quadratica.

Se con S denotiamo il limite della media quadratica, allora per il lemma di Fatou, vale

$$\begin{aligned} E[|S - \psi(B)X_t|^2] &= E \left[\liminf_{n \rightarrow \infty} \left| S - \sum_{j=-n}^n \psi_j X_{t-j} \right|^2 \right] \\ &\leq \liminf_{n \rightarrow \infty} E \left[\left| S - \sum_{j=-n}^n \psi_j X_{t-j} \right|^2 \right] = 0, \end{aligned}$$

che dimostra che il limite S e $\psi(B)X_t$ sono uguali quasi certamente. \square

Proposizione 2.3.2. Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con funzione di autocovarianza $\gamma_X(\cdot)$ e sia $(\psi_j)_{j \in \mathbb{Z}}$ una successione di costanti tale che

$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Allora, per ogni t in \mathbb{Z} , la serie

$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (2.6)$$

converge assolutamente con probabilità uno e in media quadratica.

Inoltre, posto

$$Y_t = \psi(B)X_t,$$

allora il processo $(Y_t)_{t \in \mathbb{Z}}$ è stazionario con funzione di autocovarianza

$$\gamma_Y(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h - j + k).$$

Dimostrazione. La convergenza della serie segue dalla Proposizione (2.3.1). Infatti, siccome per ipotesi $(X_t)_{t \in \mathbb{Z}}$ è stazionario, abbiamo

$$E[|X_t|] \leq (E[|X_t|^2])^{1/2} = c,$$

dove c è finito e non dipendente da t .

Per dimostrare la stazionarietà del processo $(Y_t)_{t \in \mathbb{Z}}$, osserviamo che, usando la convergenza in media quadratica della serie (2.6) e la continuità del prodotto scalare, abbiamo

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

e

$$\begin{aligned} E[Y_{t+h}Y_t] &= \lim_{n \rightarrow \infty} E \left[\left(\sum_{j=-n}^n \psi_j X_{t+h-j} \right) \left(\sum_{k=-n}^n \psi_k X_{t-k} \right) \right] \\ &= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k (\gamma_X(h - j + k) + (E[X_t])^2). \end{aligned}$$

Ne segue che $E[Y_t]$ e $E[Y_{t+h}Y_t]$ sono entrambi finiti e non dipendenti da t .

La funzione di autocovarianza $\gamma_Y(\cdot)$ di $(Y_t)_{t \in \mathbb{Z}}$ è data da

$$\gamma_Y(h) = E[Y_{t+h}Y_t] - E[Y_{t+h}]E[Y_t] = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h - j + k).$$

□

Il seguente teorema fornisce condizioni necessarie e sufficienti affinché un processo ARMA(p,q) sia causale.

Teorema 2.3.3. *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo ARMA(p,q) tale che i polinomi $\phi(\cdot)$ e $\theta(\cdot)$ non abbiano zeri in comune. Allora $(X_t)_{t \in \mathbb{Z}}$ è causale se e solo se $\phi(z) \neq 0$ per ogni $z \in \mathbb{C}$ con $|z| \leq 1$. I coefficienti ψ_j della (2.4) sono determinati dalla relazione*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1. \quad (2.7)$$

Dimostrazione. Supponiamo che $\phi(z) \neq 0$ con $|z| \leq 1$. Questo ci assicura che esiste un $\epsilon > 0$ tale che

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \xi_j z^j = \xi(z), \quad |z| < 1 + \epsilon.$$

Di conseguenza $\xi_j(1 + \epsilon/2)^j \rightarrow 0$ quando $j \rightarrow \infty$ e quindi esiste una costante $K \in (0, \infty)$ tale che

$$|\xi_j| < K(1 + \epsilon/2)^{-j}, \quad j = 0, 1, 2, \dots$$

Ne ricaviamo che $\sum_{j=0}^{\infty} |\xi_j| < \infty$ e $\xi(z)\phi(z) \equiv 1$ per $|z| \leq 1$. Grazie alla Proposizione 2.3.2, possiamo applicare l'operatore $\xi(B)$ ad entrambi i membri dell'equazione $\phi(B)X_t = \theta(B)Z_t$, ottenendo

$$X_t = \xi(B)\theta(B)Z_t.$$

Abbiamo dunque che $(X_t)_{t \in \mathbb{Z}}$ è causale, ossia

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

dove le costanti ψ_j sono determinate dalla (2.7).

Supponiamo adesso che $(X_t)_{t \in \mathbb{Z}}$ sia causale, cioè $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ con $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Allora vale

$$\theta(B)Z_t = \phi(B)X_t = \phi(B)\psi(B)Z_t.$$

Ponendo $\eta(z) = \phi(z)\psi(z) = \sum_{j=0}^{\infty} \eta_j z^j$ per $|z| \leq 1$, possiamo riscrivere l'equazione precedente come

$$\sum_{j=0}^q \theta_j Z_{t-j} = \sum_{j=0}^{\infty} \eta_j Z_{t-j}.$$

Moltiplicando per Z_{t-k} entrambe le parti dell'equazione e passando alle speranze, otteniamo $\eta_k = \theta_k$, $k = 0, \dots, q$ e $\eta_k = 0$, $k > q$ ossia

$$\theta(z) = \eta(z) = \phi(z)\psi(z), \quad |z| \leq 1.$$

Poiché $\theta(z)$ e $\phi(z)$ non hanno zeri in comune e $|\psi(z)| < \infty$ per $|z| \leq 1$, concludiamo che $\phi(z)$ non può essere zero per $|z| \leq 1$. \square

Osservazione 2.1. Se $(X_t)_{t \in \mathbb{Z}}$ è un processo ARMA per il quale i polinomi $\phi(\cdot)$ e $\theta(\cdot)$ hanno zeri in comune, allora ci sono due possibilità:

1. nessun zero sta sul cerchio unitario, in questo caso $(X_t)_{t \in \mathbb{Z}}$ è l'unica soluzione stazionaria dell'equazione ARMA.
2. almeno uno zero in comune sta sul cerchio unitario, in questo caso l'equazione ARMA può avere più di una soluzione stazionaria.

I processi ARMA per i quali i polinomi $\phi(\cdot)$ e $\theta(\cdot)$ hanno zeri in comune sono dunque poco considerati.

Osservazione 2.2. Nella prima parte della dimostrazione del teorema precedente abbiamo ottenuto che, se X_t è una soluzione stazionaria dell'equazione ARMA con $\phi(z) \neq 0$ per $|z| \leq 1$, allora deve essere della forma $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ dove i coefficienti ψ_j sono definiti dalla (2.7).

Viceversa, se $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, allora $\phi(B)X_t = \phi(B)\psi(B)Z_t = \theta(B)Z_t$. Il processo $(\psi(B)Z_t)_{t \in \mathbb{Z}}$ è dunque l'unico processo stazionario che è soluzione dell'equazione ARMA se $\phi(z) \neq 0$ per $|z| \leq 1$.

Osservazione 2.3. Se i polinomi $\phi(\cdot)$ e $\theta(\cdot)$ non hanno zeri in comune e se $\phi(z) = 0$ per un qualche $z \in \mathbb{C}$ con $|z| = 1$, allora non esiste una soluzione stazionaria dell'equazione $\phi(B)X_t = \theta(B)Z_t$.

Introduciamo un altro concetto, il quale è legato alla causalità .

Definizione 2.7. Un processo ARMA(p,q) definito dall'equazione

$$\phi(B)X_t = \theta(B)Z_t$$

è detto *invertibile* se esistono delle costanti π_j tali che $\sum_{j=0}^{\infty} |\pi_j| < \infty$ e

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t \in \mathbb{Z}, \quad (2.8)$$

ossia

$$\pi_0 X_t = Z_t - \sum_{j=1}^{\infty} \pi_j X_{t-j}.$$

Come per la causalità , la proprietà di invertibilità riguarda la relazione tra i processi $(X_t)_{t \in \mathbb{Z}}$ e $(Z_t)_{t \in \mathbb{Z}}$. Infatti un processo ARMA $(X_t)_{t \in \mathbb{Z}}$ risulta invertibile se può essere espresso come una funzione dei suoi valori precedenti, rispetto al tempo t , più un processo rumore bianco $(Z_t)_{t \in \mathbb{Z}}$.

Il seguente teorema fornisce condizioni necessarie e sufficienti per l'invertibilità .

Teorema 2.3.4. *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo ARMA(p,q) tale che i polinomi $\phi(\cdot)$ e $\theta(\cdot)$ non abbiano zeri in comune. Allora $(X_t)_{t \in \mathbb{Z}}$ è invertibile se e solo se $\theta(z) \neq 0$ per ogni $z \in \mathbb{C}$ con $|z| \leq 1$. I coefficienti π_j della (2.8) sono determinati dalla relazione*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1. \quad (2.9)$$

Dimostrazione. Supponiamo che $\theta(z) \neq 0$ se $|z| \leq 1$. Procedendo come nella dimostrazione del Teorema 2.3.3, possiamo scrivere $1/\theta(z)$ come serie di potenze:

$$\frac{1}{\theta(z)} = \sum_{j=0}^{\infty} \eta_j z^j = \eta(z), \quad |z| < 1 + \epsilon, \text{ con } \epsilon > 0.$$

Poiché $\sum_{j=0}^{\infty} |\eta_j| < \infty$, grazie alla Proposizione 2.3.1, possiamo applicare $\eta(B)$ ad entrambi i membri dell'equazione $\phi(B)X_t = \theta(B)Z_t$, ottenendo

$$\eta(B)\phi(B)X_t = \eta(B)\theta(B)Z_t = Z_t.$$

Concludiamo dunque che $(X_t)_{t \in \mathbb{Z}}$ è invertibile, ossia

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

dove le costanti π_j sono determinate dalla (2.9).

Viceversa, se $(X_t)_{t \in \mathbb{Z}}$ è invertibile, allora $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ con $\sum_{j=0}^{\infty} |\pi_j| < \infty$. Allora vale

$$\phi(B)Z_t = \pi(B)\phi(B)X_t = \pi(B)\theta(B)Z_t.$$

Ponendo $\xi(z) = \pi(z)\theta(z) = \sum_{j=0}^{\infty} \xi_j z^j$ per $|z| \leq 1$, possiamo riscrivere l'equazione precedente come

$$\sum_{j=0}^p \phi_j Z_{t-j} = \sum_{j=0}^{\infty} \xi_j Z_{t-j}.$$

Moltiplicando per Z_{t-k} entrambi le parti dell'equazione precedente e passando alle speranze, otteniamo $\xi_k = \phi_k$, per $k = 0, \dots, p$ e $\xi_k = 0$, per $k > p$. Dunque

$$\phi(z) = \xi(z) = \pi(z)\theta(z), \quad |z| \leq 1.$$

Poiché $\theta(z)$ e $\phi(z)$ non hanno zeri in comune e $|\pi(z)| < \infty$ per $|z| \leq 1$, possiamo concludere che $\theta(z)$ non può essere zero per $|z| \leq 1$. \square

Osservazione 2.4. Se $(X_t)_{t \in \mathbb{Z}}$ è una soluzione stazionaria dell'equazione

$$\phi(B)X_t = \theta(B)Z_t, \quad \text{con } Z_t \sim WN(0, \sigma^2)$$

e se $\phi(z)\theta(z) \neq 0$ per $|z| \leq 1$, allora

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

e

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

dove $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$ e $\sum_{j=0}^{\infty} \pi_j z^j = \phi(z)/\theta(z)$, $|z| \leq 1$.

Osservazione 2.5. Sia $(X_t)_{t \in \mathbb{Z}}$ un processo ARMA definito da $\phi(B)X_t = \theta(B)Z_t$, con $\phi(z)$ diverso da zero per tutti i valori z tali che $|z| = 1$, allora è possibile trovare dei polinomi $\tilde{\phi}(\cdot)$, $\tilde{\theta}(\cdot)$ e un rumore bianco $(\tilde{Z}_t)_{t \in \mathbb{Z}}$ tali che $\tilde{\phi}(B)X_t = \tilde{\theta}(B)\tilde{Z}_t$, in modo che $(X_t)_{t \in \mathbb{Z}}$ sia una funzione causale di $(\tilde{Z}_t)_{t \in \mathbb{Z}}$. Se inoltre $\theta(z)$ è diverso da zero quando $|z| = 1$, allora è possibile scegliere $\tilde{\theta}(\cdot)$ in modo che il processo $(X_t)_{t \in \mathbb{Z}}$ sia anche funzione invertibile di $(\tilde{Z}_t)_{t \in \mathbb{Z}}$, cioè tale che $\tilde{\theta}(z)$ è diverso da zero per $|z| \leq 1$. Se $(Z_t)_{t \in \mathbb{Z}}$ è indipendente, non è vero in generale che anche $(\tilde{Z}_t)_{t \in \mathbb{Z}}$ lo sia. Ciò è vero, però, se $(Z_t)_{t \in \mathbb{Z}}$ è gaussiano.

Se $\phi(\cdot)$ e $\theta(\cdot)$ non hanno zeri in comune e se $\phi(z) = 0$ per $z \in \mathbb{C}$ con $|z| = 1$, allora non esiste una soluzione stazionaria dell'equazione $\phi(B)X_t = \theta(B)Z_t$. D'altra parte, se $\phi(z) \neq 0$ per ogni $z \in \mathbb{C}$ tale che $|z| = 1$, allora esiste un $r > 1$ tale che la serie di Laurent

$$\theta(z)\phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j z^j = \psi(z), \quad r^{-1} < |z| < r, \quad (2.10)$$

è assolutamente convergente nella corona circolare. Questo risultato è fondamentale per dimostrare il seguente teorema.

Teorema 2.3.5. *Se $\phi(z) \neq 0$ per ogni $z \in \mathbb{C}$ con $|z| = 1$, allora il processo ARMA(p, q) definito dall'equazione $\phi(B)X_t = \theta(B)Z_t$ ha un'unica soluzione stazionaria, data da*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (2.11)$$

dove i coefficienti ψ_j sono determinati dall'equazione (2.10).

Dimostrazione. Dalla Proposizione 2.3.2, il processo $(X_t)_{t \in \mathbb{Z}}$ definito nell'equazione (2.11) risulta essere stazionario. Sempre grazie alla Proposizione

2.3.2, sappiamo che $\phi(B)\psi(B)Z_t = \theta(B)Z_t$ e quindi, applicando $\phi(B)$ ad entrambi i membri della (2.11), otteniamo

$$\phi(B)X_t = \theta(B)Z_t. \quad (2.12)$$

Il processo $(X_t)_{t \in \mathbb{Z}}$ risulta quindi una soluzione stazionaria del processo ARMA.

Viceversa, sia $(X_t)_{t \in \mathbb{Z}}$ una soluzione stazionaria dell'equazione (2.12). Poiché $\phi(z) \neq 0$ per ogni $z \in \mathbb{C}$ con $|z| = 1$, esiste $\delta > 1$ tale che la serie $\sum_{j=-\infty}^{\infty} \xi_j z^j = \phi(z)^{-1} = \xi(z)$ è assolutamente convergente per $\delta^{-1} < |z| < \delta$. Di conseguenza possiamo applicare $\xi(B)$ ad entrambi i membri della (2.12) ottenendo

$$\xi(B)\phi(B)X_t = \xi(B)\theta(B)Z_t,$$

o, equivalentemente,

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

□

In virtù delle Osservazioni 2.4 e 2.5, nel seguito concentreremo la nostra attenzione sui processi ARMA causali e invertibili.

2.4 Il calcolo della funzione di autocovarianza di un processo ARMA

Nel seguito descriveremo tre diversi metodi per calcolare la funzione di autocovarianza di un processo ARMA(p,q). Il terzo metodo è il più utile per ottenere risultati numerici; mentre il secondo è il più conveniente per ottenere soluzioni in forma chiusa.

Primo metodo. La funzione di autocovarianza γ_X di un processo ARMA(p,q) causale definito da $\phi(B)X_t = \theta(B)Z_t$ può essere scritta, per

2.4 Il calcolo della funzione di autocovarianza di un processo ARMA5

quanto visto nel paragrafo precedente, come

$$\gamma_X(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}, \quad (2.13)$$

dove

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1, \quad (2.14)$$

$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ e $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$. Per determinare i coefficienti ψ_j , possiamo scrivere la (2.14) nella forma $\psi(z)\phi(z) = \theta(z)$. Uguagliando i coefficienti di z^j e ponendo $\theta_0 = 1$, $\theta_j = 0$ per $j > q$ e $\phi_j = 0$ per $j > p$, otteniamo

$$\psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j < \max(p, q + 1), \quad (2.15)$$

e

$$\psi_j - \sum_{0 < k \leq p} \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q + 1). \quad (2.16)$$

Da queste equazioni ricaviamo ricorsivamente $\psi_0, \psi_1, \psi_2, \dots$, ottenendo

$$\psi_0 = \theta_0 = 1,$$

$$\psi_1 = \theta_1 + \psi_0 \phi_1 = \theta_1 + \phi_1,$$

$$\psi_2 = \theta_2 + \psi_0 \phi_2 + \psi_1 \phi_1 = \theta_2 + \phi_2 + \theta_1 \phi_1 + \phi_1^2,$$

...

Un altro modo consiste nello scrivere la soluzione generale della (2.16) (si veda Paragrafo 3.6 di [7]) come

$$\psi_n = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \alpha_{ij} n^j \xi_i^{-n}, \quad n \geq \max(p, q + 1) - p,$$

dove ξ_i , con $i = 1, \dots, k$ sono gli zeri distinti di $\phi(z)$ e r_i è la molteplicità di ξ_i , in modo da avere $\sum_{i=1}^k r_i = p$. Le p costanti α_{ij} e i coefficienti ψ_j , $0 \leq j < \max(p, q + 1) - p$, sono unicamente determinati dalle condizioni (2.15). Riusciamo a calcolare quindi tutti i coefficienti ψ_j e, di conseguenza, la funzione di autocovarianza γ .

Esempio 9. Sia dato il processo ARMA(p,q)

$$(1 - B + \frac{1}{4}B^2)X_t = (1 + B)Z_t.$$

In questo caso abbiamo che i coefficienti valgono $\phi_0 = \phi_1 = 1$, $\phi_2 = -1/4$, $\theta_0 = \theta_1 = 1$. Le equazioni (2.15) diventano

$$\psi_0 = \theta_0 = 1$$

$$\psi_1 = \theta_1 + \psi_0\phi_1 = \theta_1 + \phi_1 = 2.$$

Dalla (2.16) otteniamo

$$\psi_j - \psi_{j-1} + \frac{1}{4}\psi_{j-2} = 0, \quad \text{per } j \geq 2.$$

Osservando che

$$0 = \phi(z) = (1 - z + 1/4z^2) = (z - 2)^2 \Rightarrow z = 2 \Rightarrow \xi = 2,$$

abbiamo che la soluzione generale della (2.16) è data da

$$\psi_n = (\alpha_{10} + n\alpha_{11})2^{-n}, \quad \text{con } n \geq 0,$$

dove le costanti α_{10} e α_{11} sono individuate dalle condizioni

$$\psi_0 = 2^0\alpha_{10} \Rightarrow \alpha_{10} = 1,$$

$$\psi_1 = \frac{1}{2}(1 + \alpha_{11}) \Rightarrow \alpha_{11} = 3.$$

Otteniamo dunque

$$\psi_n = (1 + 3n)2^{-n}, \quad \text{per } n = 0, 1, 2, \dots$$

Infine, sostituendo nella (2.13), otteniamo il valore di $\gamma_X(k)$ per $k \geq 0$

$$\begin{aligned} \gamma_X(k) &= \sigma^2 \sum_{j=0}^{\infty} (1 + 3j)(1 + 3j + 3k)2^{-2j-k} \\ &= \sigma^2 2^{-k} \sum_{j=0}^{\infty} [(3k + 1)4^{-j} + 3(3k + 2)j4^{-j} + 9j^2 4^{-j}] \\ &= \sigma^2 2^{-k} \left[\frac{4}{3}(3k + 1) + \frac{12}{9}(3k + 2) + \frac{180}{27} \right] \\ &= \sigma^2 2^{-k} \left[\frac{32}{3} + 8k \right]. \end{aligned}$$

2.4 Il calcolo della funzione di autocovarianza di un processo ARMA 7

Secondo metodo. Un altro metodo per calcolare la funzione di autocovarianza $\gamma_X(\cdot)$ di un processo ARMA(p,q) causale definito da

$$\phi(B)X_t = \theta(B)Z_t, \quad (2.17)$$

è basato sulle equazioni alle differenze per $\gamma_X(k)$, con $k = 0, 1, 2, \dots$, ottenute moltiplicando entrambi i membri dell'equazione (2.17) per X_{t-k} e prendendone il valore atteso:

$$\gamma_X(k) - \phi_1\gamma_X(k-1) - \dots - \phi_p\gamma_X(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k}, \quad (2.18)$$

per $0 \leq k < \max(p, q+1)$ e

$$\gamma_X(k) - \phi_1\gamma_X(k-1) - \dots - \phi_p\gamma_X(k-p) = 0, \quad k \geq \max(p, q+1), \quad (2.19)$$

(dove abbiamo usato la rappresentazione $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$).

Come nel primo metodo, possiamo scrivere la soluzione della (2.19) come

$$\gamma_X(h) = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \beta_{ij} h^j \xi_i^{-h}, \quad n \geq \max(p, q+1) - p,$$

dove le p costanti β_{ij} e le covarianze $\gamma_X(j)$, con $0 \leq j < \max(p, q+1) - p$, sono univocamente determinate dalla (2.18), dopo aver calcolato $\psi_0, \psi_1, \dots, \psi_q$ come nel metodo precedente.

Esempio 10. Allo stesso processo ARMA(p,q) dell'esempio 9, applichiamo il secondo metodo. L'equazione (2.19) diventa

$$\gamma_X(k) - \gamma_X(k-1) + \frac{1}{4}\gamma_X(k-2) = 0, \quad k \geq 2,$$

che ha come soluzione generale

$$\gamma_X(n) = (\beta_{10} + \beta_{11}n)2^{-n}, \quad n \geq 0. \quad (2.20)$$

Utilizzando il fatto che la funzione di autocovarianza è pari, le condizioni (2.18) diventano

$$\gamma_X(0) - \gamma_X(1) + \frac{1}{4}\gamma_X(2) = \sigma^2(\psi_0 + \psi_1),$$

$$\gamma_X(1) - \gamma_X(0) + \frac{1}{4}\gamma_X(1) = \sigma^2\psi_0.$$

Come nell'esempio 9, $\psi_0 = 1$ e $\psi_1 = 2$.

Sostituendo $\gamma_X(0) = \beta_{10}$, $\gamma_X(1) = 2^{-1}(\beta_{10} + \beta_{11})$ e $\gamma_X(2) = 2^{-2}(\beta_{10} + 2\beta_{11})$ nella (2.20), otteniamo

$$3\beta_{10} - 2\beta_{11} = 16\sigma^2,$$

$$-3\beta_{10} + 5\beta_{11} = 8\sigma^2,$$

da cui $\beta_{11} = 8\sigma^2$ e $\beta_{10} = 32\sigma^2/3$. Ricaviamo quindi la funzione di autocovarianza

$$\gamma_X(k) = \sigma^2 2^{-k} \left[\frac{32}{3} + 8k \right],$$

ottenendo lo stesso risultato dell'esempio 9.

Terzo metodo. Calcoliamo innanzitutto $\gamma_X(0), \dots, \gamma_X(p)$ dalle equazioni (2.18) e (2.19) con $k = 0, 1, \dots, p$, e dopo usiamo le successive equazioni per determinare $\gamma_X(p+1), \gamma_X(p+2), \dots$ ricorsivamente.

Esempio 11. Applichiamo il terzo metodo al processo considerato negli esempi 9 e 10, per il quale le equazioni (2.18) e (2.19) con $k = 0, 1, 2$ sono

$$\gamma_X(0) - \gamma_X(1) + \frac{1}{4}\gamma_X(2) = 3\sigma^2,$$

$$\gamma_X(1) - \gamma_X(0) + \frac{1}{4}\gamma_X(1) = \sigma^2,$$

$$\gamma_X(2) - \gamma_X(1) + \frac{1}{4}\gamma_X(0) = 0,$$

da cui otteniamo $\gamma_X(0) = 32\sigma^2/3$, $\gamma_X(1) = 28\sigma^2/3$, $\gamma_X(2) = 20\sigma^2/3$. Possiamo calcolare la funzione di autocovarianza per un generico $k \geq 3$ come

$$\gamma_X(k) = \gamma_X(k-1) - \frac{1}{4}\gamma_X(k-2), \quad k = 3, 4, \dots$$

Capitolo 3

La rappresentazione spettrale

La rappresentazione spettrale di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$ consiste essenzialmente nella decomposizione del processo in una somma di componenti sinusoidali con coefficienti non correlati. A questa decomposizione corrisponde una scomposizione, sempre in componenti periodiche, della funzione di autocovarianza del processo. La rappresentazione spettrale di un processo stazionario è dunque analoga alla rappresentazione di Fourier delle funzioni deterministiche. L'analisi dei processi stazionari per mezzo della loro rappresentazione spettrale è spesso chiamata *analisi del dominio delle frequenze* della serie storica. Essa è equivalente all'analisi del *dominio dei tempi*, basata sulla funzione di autocovarianza.

3.1 I processi stazionari a valori complessi

Fino a questo punto abbiamo considerato processi stazionari a valori reali, perché sono quelli che si incontrano nella realtà. Nell'analisi spettrale è però più semplice utilizzare processi a valori complessi. Estendiamo quindi le definizioni date nei capitoli precedenti al caso dei processi a valori complessi.

Definizione 3.1. Un processo $(X_t)_{t \in \mathbb{Z}}$ a valori complessi è un processo stazionario se

- $E[|X_t|^2] < \infty$;
- $E[X_t]$ non dipende da t ;
- $E[X_{t+h}\bar{X}_t]$ non dipende da t .

Definizione 3.2. La funzione di autocovarianza $\gamma_X(\cdot)$ di un processo stazionario a valori complessi $(X_t)_{t \in \mathbb{Z}}$ è

$$\gamma_X(h) = E[X_{t+h}\bar{X}_t] - E[X_{t+h}]E[\bar{X}_t].$$

Le proprietà della funzione di autocovarianza, viste nel caso reale, possono essere riscritte nel caso complesso come segue:

- $\gamma_X(0) \geq 0$;
- $|\gamma_X(h)| \leq \gamma_X(0)$ per ogni h ;
- $\gamma_X(\cdot)$ è una funzione Hermitiana, cioè $\gamma_X(h) = \overline{\gamma_X(-h)}$.

Ricordiamo inoltre che una funzione definita sugli interi a valori complessi, $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$, è semidefinita positiva se e solo se

$$\sum_{i,j=1}^n a_i \gamma(i-j) \bar{a}_j \geq 0,$$

per tutti gli interi positivi n e per tutti i vettori $\mathbf{a} = (a_1, \dots, a_n)' \in \mathbb{C}^n$.

Il seguente risultato è l'analogo delle Proposizioni 1.2.1 e 1.2.2 nel caso reale.

Teorema 3.1.1. *Una funzione γ sugli interi è la funzione di autocovarianza di un processo stazionario a valori complessi se e solo se γ è Hermitiana e semidefinita positiva.*

La dimostrazione è analoga a quella delle Proposizioni 1.2.1 e 1.2.2.

3.2 Un esempio di rappresentazione spettrale

Consideriamo il processo

$$X_t = \sum_{j=1}^n A(\lambda_j) e^{it\lambda_j} \quad (3.1)$$

dove $-\pi < \lambda_1 < \lambda_2 < \dots < \lambda_n \leq \pi$ e $A(\lambda_1), \dots, A(\lambda_n)$ sono variabili aleatorie a valori complessi non correlate e tali che

$$E[A(\lambda_j)] = 0, \quad j = 1, \dots, n,$$

e

$$E[A(\lambda_j) \overline{A(\lambda_j)}] = \sigma_j^2, \quad j = 1, \dots, n.$$

Il processo (3.1) risulta essere stazionario: infatti abbiamo

$$E[X_t] = 0$$

e

$$\begin{aligned} \gamma_X(h) &= E[X_{t+h} \overline{X_t}] = \sum_{j,k=1}^n E[A(\lambda_j) \overline{A(\lambda_k)}] e^{i(\lambda_j(t+h) - \lambda_k t)} \\ &= \sum_{j=1}^n E[A(\lambda_j) \overline{A(\lambda_j)}] e^{i(\lambda_j(t+h) - \lambda_j t)} = \sum_{j=1}^n E[|A(\lambda_j)|^2] e^{ih\lambda_j} = \sum_{j=1}^n \sigma_j^2 e^{ih\lambda_j}. \end{aligned}$$

Utilizzando l'integrale di Riemann-Stieltjes, possiamo scrivere la funzione di autocovarianza come

$$\gamma_X(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu), \quad (3.2)$$

dove F è la funzione di ripartizione

$$F(\lambda) = \sum_{j: \lambda_j \leq \lambda} \sigma_j^2 = \sum_{j: \lambda_j \leq \lambda} E[|A(\lambda_j)|^2],$$

detta *funzione di ripartizione spettrale* di $(X_t)_{t \in \mathbb{Z}}$.

Il secondo membro della (3.1) fornisce la rappresentazione spettrale del processo $(X_t)_{t \in \mathbb{Z}}$ e la (3.2) è la corrispondente rappresentazione spettrale della

funzione di autocovarianza.

La (3.1) può essere vista come una rappresentazione della forma

$$X_t = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu), \quad (3.3)$$

dove Z è un processo ad incrementi ortogonali (si veda Appendice B). La corrispondente rappresentazione della funzione di autocovarianza è data da

$$\begin{aligned} \gamma_X(h) &= E[X_{t+h}\overline{X}_t] = E\left(\int_{(-\pi, \pi]} e^{i\nu(t+h)} Z(\nu) d\nu \int_{(-\pi, \pi]} e^{i\mu t} \overline{Z(\mu)} d\mu\right) \\ &= \int_{(-\pi, \pi]} e^{i\nu(t+h)-it\mu} E[Z(\nu)\overline{Z(\mu)}] d\nu d\mu \\ &= \int_{(-\pi, \pi]} e^{ih\nu} E[Z(\nu)\overline{Z(\nu)}] d\nu = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu), \end{aligned} \quad (3.4)$$

dove F è la funzione di ripartizione associata al processo $\{Z(\lambda) : -\pi \leq \lambda \leq \pi\}$. In particolare, essa è tale che

$$F(-\pi) = 0 \text{ e } F(\pi) = \gamma(0) = E[|X_t|^2].$$

3.3 Il teorema di Herglotz

Nel precedente paragrafo abbiamo visto come il Teorema 3.1.1 caratterizza le funzioni di autocovarianza definite sull'insieme degli interi come quelle funzioni che risultano essere Hermitiane e semidefinite positive. Proviamo adesso il Teorema di Herglotz che caratterizza le funzioni di autocovarianza come le funzioni che possono essere scritte nella forma (3.4), per una qualche funzione di ripartizione limitata F di una misura concentrata su $(-\pi, \pi]$.

Teorema 3.3.1. *Una funzione a valori complessi $\gamma(\cdot)$ definita sugli interi è semidefinita positiva se e solo se*

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu) \quad \forall h \in \mathbb{Z}, \quad (3.5)$$

dove $F(\cdot)$ è continua a destra, non decrescente, limitata su $[-\pi, \pi]$ e $F(-\pi) = 0$.

La funzione F è chiamata *funzione di ripartizione spettrale* di γ .

Se $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu) d\nu$, $-\pi \leq \lambda \leq \pi$, allora f è chiamata *densità spettrale* di $\gamma(\cdot)$.

Dimostrazione. Se $\gamma(\cdot)$ è della forma (3.5) allora $\gamma(\cdot)$ risulta Hermitiana, cioè $\gamma(-h) = \overline{\gamma(h)}$. Inoltre, se $a_r \in \mathbb{C}$, $r = 1, \dots, n$, allora

$$\begin{aligned} \sum_{r,s=1}^n a_r \gamma(r-s) \bar{a}_s &= \int_{-\pi}^{\pi} \sum_{r,s=1}^n a_r \bar{a}_s e^{i\nu(r-s)} dF(\nu) \\ &= \int_{-\pi}^{\pi} \left| \sum_{r=1}^n a_r e^{i\nu r} \right|^2 dF(\nu) \geq 0, \end{aligned}$$

ossia $\gamma(\cdot)$ risulta semidefinita positiva e quindi, grazie al Teorema 3.1.1, è una funzione di autocovarianza.

Viceversa, supponiamo che $\gamma(\cdot)$ sia semidefinita positiva. Definiamo

$$\begin{aligned} f_N(\nu) &= \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\nu} \gamma(r-s) e^{is\nu} \\ &= \frac{1}{2\pi N} \sum_{|m| < N} (N - |m|) e^{-im\nu} \gamma(m). \end{aligned}$$

Poiché $\gamma(\cdot)$ è semidefinita positiva, allora abbiamo

$$f_N(\nu) \geq 0 \quad \forall \nu \in (-\pi, \pi].$$

Sia $F_N(\cdot)$ la funzione di ripartizione corrispondente alla densità $f_N(\cdot)I_{(-\pi, \pi]}(\cdot)$.

In questo modo $F_N(\lambda) = 0$, $\lambda \leq -\pi$, $F_N(\lambda) = F_N(\pi)$, $\lambda \geq \pi$ e

$$F_N(\lambda) = \int_{-\pi}^{\lambda} f_N(\nu) d\nu, \quad -\pi \leq \lambda \leq \pi.$$

Allora, per qualsiasi intero h , abbiamo

$$\int_{(-\pi, \pi]} e^{ih\nu} dF_N(\nu) = \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) \gamma(m) \int_{-\pi}^{\pi} e^{i(h-m)\nu} d\nu,$$

cioè

$$\int_{(-\pi, \pi]} e^{ih\nu} dF_N(\nu) = \begin{cases} \left(1 - \frac{|h|}{N}\right) \gamma(h) & |h| < N, \\ 0 & \text{altrimenti.} \end{cases} \quad (3.6)$$

Grazie al fatto che $F_N(\pi) = \int_{(-\pi, \pi]} dF_N(\nu) = \gamma(0) < +\infty$ per tutti gli N , possiamo applicare il Teorema di Helly (si veda Teorema 4.3.2 di [8]) da cui si deduce che esiste una funzione di ripartizione F e una sottosuccessione (F_{N_k}) della successione (F_N) tale che per qualsiasi funzione continua e limitata g , vale

$$\int_{(-\pi, \pi]} g(\nu) dF_{N_k}(\nu) \rightarrow \int_{(-\pi, \pi]} g(\nu) dF(\nu) \quad \text{per } k \rightarrow +\infty.$$

Sostituendo N con N_k nella (3.6) e mandando $k \rightarrow +\infty$, otteniamo

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu)$$

che è la cercata rappresentazione spettrale di $\gamma(\cdot)$. \square

Corollario 3.3.2. *Una funzione a valori complessi $\gamma(\cdot)$ definita sugli interi è la funzione di autocovarianza di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$ se e solo se soddisfa una delle seguenti condizioni:*

1. $\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu)$ per h in \mathbb{Z} , dove F è una funzione continua a destra, non decrescente, limitata su $[-\pi, \pi]$ con $F(-\pi) = 0$;
2. $\sum_{i,j=1}^n a_i \gamma(i-j) \bar{a}_j \geq 0$ per ogni numero intero positivo n e per ogni $\mathbf{a} = (a_1, \dots, a_n)' \in \mathbb{C}^n$.

La funzione di ripartizione spettrale $F(\cdot)$ (e, se esiste, la corrispondente densità spettrale) è chiamata funzione di ripartizione spettrale (e densità spettrale) di $\gamma(\cdot)$ e di $(X_t)_{t \in \mathbb{Z}}$.

Dimostrazione. Il Teorema di Herglotz garantisce l'equivalenza tra i punti 1 e 2. Da 1 segue che $\gamma(\cdot)$ è Hermitiana. Inoltre le condizioni del Teorema 3.1.1 sono soddisfatte se $\gamma(\cdot)$ verifica o 1 o 2. \square

Osserviamo che la funzione di ripartizione $F(\cdot)$ (con $F(-\pi) = 0$) è univocamente determinata da $\gamma(n)$ con n in \mathbb{Z} . Se esistessero due funzioni di ripartizione F e G , nulle su $(-\infty, \pi]$, costanti su $[\pi, +\infty)$ e tali che

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu) = \int_{[\pi, +\infty)} e^{ih\nu} dG(\nu), \quad h \in \mathbb{Z},$$

allora avremmo

$$\int_{(-\pi, \pi]} \phi(\nu) dF(\nu) = \int_{(-\pi, \pi]} \phi(\nu) dG(\nu)$$

per ogni funzione continua $\phi(\cdot)$ su $[-\pi, \pi]$ con $\phi(\pi) = \phi(-\pi)$ e dunque $F(\lambda) = G(\lambda)$ per ogni $\lambda \in (-\infty, +\infty)$.

Il seguente teorema è utile per ricavare la funzione di ripartizione da γ in molti casi (in particolare, quando γ è la funzione di autocovarianza di un processo ARMA(p,q)).

Teorema 3.3.3. *Se $K(\cdot)$ è una funzione a valori complessi sugli interi tale che*

$$\sum_{k=-\infty}^{\infty} |K(n)| < \infty, \quad (3.7)$$

allora

$$K(h) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu \quad \text{per } h \in \mathbb{Z},$$

dove

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-in\lambda} K(n).$$

Dimostrazione.

$$\begin{aligned} \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{i(h-n)\nu} K(n) d\nu \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} K(n) \int_{-\pi}^{\pi} e^{i(h-n)\nu} d\nu = K(h), \end{aligned}$$

in quanto l'unico addendo diverso da zero si ha per $n = h$. Lo scambio dell'integrale con la sommatoria è giustificato dal teorema di Fubini grazie al fatto che $\int_{-\pi}^{\pi} (1/2\pi) \sum_{n=-\infty}^{+\infty} |e^{i(h-n)\nu} K(n)| d\nu < +\infty$ per la (3.7) \square

Corollario 3.3.4. *Una funzione γ a valori complessi assolutamente sommabile definita sugli interi risulta essere la funzione di autocovarianza di un processo stazionario se e solo se*

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-in\lambda} \gamma(n) \geq 0 \quad \forall \lambda \in [-\pi, \pi]$$

(in tal caso $f(\cdot)$ è la densità spettrale di $\gamma(\cdot)$).

Dimostrazione. Supponiamo che $\gamma(\cdot)$ sia una funzione di autocovarianza. Dal fatto che $\gamma(\cdot)$ è semidefinita positiva e assolutamente sommabile, abbiamo

$$\begin{aligned} 0 \leq f_N(\lambda) &= \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\nu} \gamma(r-s) e^{is\nu} \\ &= \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) e^{-im\nu} \gamma(m) \rightarrow f(\lambda) \quad \text{per } N \rightarrow +\infty. \end{aligned}$$

Di conseguenza $f(\lambda) \geq 0$, $-\pi \geq \lambda \geq \pi$. Dal Teorema precedente abbiamo che $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu$, $h \in \mathbb{Z}$. Perciò $f(\cdot)$ è la densità spettrale di $\gamma(\cdot)$. Viceversa, se assumiamo che $\gamma(\cdot)$ sia assolutamente sommabile, dal Teorema 3.3.3 ricaviamo che $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu$. Se $f(\lambda) \geq 0$, allora l'integrale precedente è della forma (3.5) con $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu) d\nu$. Grazie al Corollario 3.3.2, otteniamo che $\gamma(\cdot)$ è una funzione di autocovarianza con densità spettrale f . \square

Dal Corollario 3.3.2 segue che ogni processo ARMA(p,q) ha una densità spettrale.

Osservazione 3.1. Nel caso in cui il processo stazionario $(X_t)_{t \in \mathbb{Z}}$ sia a valori reali con funzione di ripartizione spettrale F , prendendo $U(\lambda) = \text{Re}\{Z(\lambda)\}$ e $V(\lambda) = -\text{Im}\{Z(\lambda)\}$, possiamo riscrivere la (3.3) come

$$X_t = \int_{(-\pi, \pi]} \cos(\nu t) dU(\nu) + \int_{(-\pi, \pi]} \sin(\nu t) dV(\nu). \quad (3.8)$$

Inoltre, essendo la sua funzione di autocovarianza $\gamma_X(h)$ reale, necessariamente F sarà simmetrica, ossia $F(\lambda) = F(\pi^-) - F(-\lambda^-)$, $-\pi < \lambda < \pi$, dove $F(\lambda^-)$ è il limite da sinistra di F in λ . La (3.4) diventa quindi

$$\gamma_X(h) = \int_{(-\pi, \pi]} \cos(\nu h) dF(\nu).$$

In particolare, se $\gamma_X(\cdot)$ ha densità spettrale $f_X(\lambda)$ per $-\pi \leq \lambda \leq \pi$, allora $f_X(\lambda) = f_X(-\lambda)$ per $-\pi \leq \lambda \leq \pi$, e dunque vale

$$\gamma_X(h) = 2 \int_0^{\pi} f_X(\nu) \cos(\nu h) d\nu.$$

La struttura della funzione di autocovarianza di un processo stazionario reale $(X_t)_{t \in \mathbb{Z}}$ è determinata da $F_X(0-)$ e $F_X(\lambda)$ per $0 \leq \lambda \leq \pi$, oppure da $f_X(\lambda)$ per $0 \leq \lambda \leq \pi$, se esiste la densità spettrale $f_X(\cdot)$.

Osservazione 3.2. Dall'osservazione precedente segue che una funzione f definita su $[-\pi, \pi]$ è la densità spettrale di un processo stazionario reale se e solo se

- $f(\lambda) = f(-\lambda)$;
- $f(\lambda) \geq 0$;
- $\int_{-\pi}^{\pi} f(\lambda) d\lambda < +\infty$.

Esempio 12 (Densità spettrale di un rumore bianco). Ricordiamo che il processo rumore bianco $(Z_t)_{t \in \mathbb{Z}}$ ha funzione di autocovarianza

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

Applicando il Corollario 3.3.4 otteniamo che la densità spettrale di un rumore bianco è data da

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-in\lambda} \gamma_Z(n) = \frac{1}{2\pi} \gamma_Z(0) = \frac{\sigma^2}{2\pi}.$$

Il processo rumore bianco ha quindi densità spettrale costante.

Dal fatto che la densità spettrale è la stessa per tutte le frequenze, usando un'analogia con lo spettro della luce bianca, che è composta da tutte le frequenze costanti, capiamo la provenienza del nome rumore bianco.

3.4 La densità spettrale di un processo ARMA

Teorema 3.4.1. *Sia $(Y_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero a valori complessi con funzione di ripartizione spettrale $F_Y(\cdot)$, e sia $(X_t)_{t \in \mathbb{Z}}$ il processo definito da*

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j Y_{t-j}, \quad (3.9)$$

con $\sum_{j=-\infty}^{+\infty} |\psi_j| < +\infty$. Allora il processo $(X_t)_{t \in \mathbb{Z}}$ risulta stazionario con funzione di ripartizione spettrale

$$F_X(\lambda) = \int_{(-\pi, \lambda]} \left| \sum_{j=-\infty}^{+\infty} \psi_j e^{-ij\nu} \right|^2 dF_Y(\nu), \quad -\pi \leq \lambda \leq \pi. \quad (3.10)$$

Se $(Y_t)_{t \in \mathbb{Z}}$ ha densità spettrale $f_Y(\cdot)$ e $(X_t)_{t \in \mathbb{Z}}$ è un processo come definito nella (3.9), allora $(X_t)_{t \in \mathbb{Z}}$ ha densità spettrale f_X data da

$$f_X(\lambda) = |\psi(e^{-i\lambda})|^2 f_Y(\lambda),$$

dove $\psi(e^{-i\lambda}) = \sum_{j=-\infty}^{+\infty} \psi_j e^{-ij\lambda}$.

Dimostrazione. Con lo stesso ragionamento usato per la Proposizione 2.3.2, si dimostra che il processo $(X_t)_{t \in \mathbb{Z}}$ è stazionario con media zero e funzione di autocovarianza data da

$$\gamma_X(h) = E[X_{t+h} \bar{X}_t] = \sum_{j,k=-\infty}^{+\infty} \psi_j \bar{\psi}_k \gamma_Y(h-j+k), \quad h \in \mathbb{Z}.$$

Usando la rappresentazione spettrale di $(\gamma_Y(h))_{h \in \mathbb{Z}}$, possiamo scrivere

$$\begin{aligned} \gamma_X(h) &= \sum_{j,k=-\infty}^{+\infty} \psi_j \bar{\psi}_k \int_{(-\pi, \pi]} e^{i(h-j+k)\nu} dF_Y(\nu) \\ &= \int_{(-\pi, \pi]} \left(\sum_{j=-\infty}^{+\infty} \psi_j e^{-ij\nu} \right) \left(\sum_{k=-\infty}^{+\infty} \bar{\psi}_k e^{ik\nu} \right) e^{ih\nu} dF_Y(\nu) \\ &= \int_{(-\pi, \pi]} e^{ih\nu} \left| \sum_{j=-\infty}^{+\infty} \psi_j e^{-ij\nu} \right|^2 dF_Y(\nu), \end{aligned}$$

da cui segue che la funzione F_X definita nella (3.10) è la funzione di ripartizione spettrale del processo $(X_t)_{t \in \mathbb{Z}}$. \square

L'operatore $\psi(B) = \sum_{j=-\infty}^{+\infty} \psi_j B^j$ applicato al processo $(Y_t)_{t \in \mathbb{Z}}$ nella (3.9) è chiamato *filtro lineare invariante* (rispetto al tempo) con pesi (ψ_j) . La funzione $\lambda \rightarrow \psi(e^{-i\lambda})$ è chiamata *funzione di trasferimento* del filtro, mentre la funzione $\lambda \rightarrow |\psi(e^{-i\lambda})|^2$ è chiamata *funzione potenza* del filtro.

Dal Teorema 3.4.1 possiamo ricavare la densità spettrale di un processo ARMA(p,q).

Teorema 3.4.2. Sia $(X_t)_{t \in \mathbb{Z}}$ un qualsiasi processo ARMA(p, q) definito da

$$\phi(B)X_t = \theta(B)Z_t, \quad \text{con } (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2), \quad (3.11)$$

tale che i polinomi $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ e $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ non abbiano zeri in comune e $\phi(z)$ sia diverso da zero sul cerchio unitario. Allora il processo $(X_t)_{t \in \mathbb{Z}}$ ha densità spettrale

$$f_X(\lambda) = \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi.$$

Poiché la densità spettrale di un processo ARMA risulta essere un rapporto di polinomi trigonometrici è spesso chiamata *densità spettrale razionale*.

Dimostrazione. Nel precedente capitolo abbiamo visto che una soluzione della (3.11) può essere scritta nella forma

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j Z_{t-j}$$

dove $\sum_{j=-\infty}^{+\infty} |\psi_j| < +\infty$. Poiché un processo rumore bianco ha densità spettrale $\sigma^2/(2\pi)$, allora dal Teorema 3.4.1 ricaviamo che il processo $(X_t)_{t \in \mathbb{Z}}$ ha densità spettrale.

Ponendo $U_t = \phi(B)X_t = \theta(B)Z_t$ e applicando il Teorema 3.9, otteniamo

$$f_U(\lambda) = |\phi(e^{-i\lambda})|^2 f_X(\lambda) = |\theta(e^{-i\lambda})|^2 f_Z(\lambda). \quad (3.12)$$

Siccome $\phi(e^{-i\lambda}) \neq 0$ per ogni $\lambda \in [-\pi, \pi]$, possiamo dividere la (3.12) per $|\phi(e^{-i\lambda})|^2$ ottenendo f_X . \square

Esempio 13 (Densità spettrale di un processo MA(1)). Se

$$X_t = Z_t + \theta Z_{t-1}, \quad \text{con } (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

allora la densità spettrale corrispondente è

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} |1 + \theta e^{-i\lambda}|^2 = \frac{\sigma^2}{2\pi} (1 + 2\theta \cos \lambda + \theta^2), \quad -\pi \leq \lambda \leq \pi.$$

Nei grafici seguenti riportiamo la densità spettrale $f_X(\lambda)$, con $\sigma^2 = 6.25$, scegliendo prima $\theta = -0.9$ e successivamente $\theta = 0.9$. Osserviamo che per $\theta = 0.9$ la densità è più alta per le basse frequenze, mentre è più bassa per le alte frequenze. Per $\theta = -0.9$ la densità spettrale è grande per le alte frequenze e piccola per le basse frequenze.

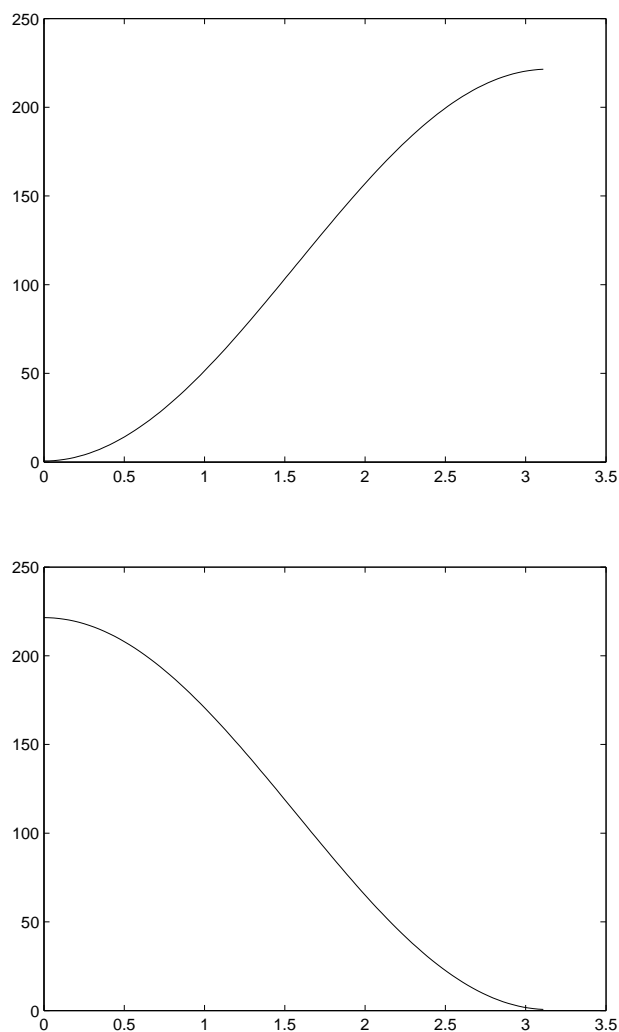


Figura 3.1: Densità spettrale $f_X(\lambda)$ di un processo MA(1)

Esempio 14 (Densità spettrale di un processo AR(1)). Se

$$X_t - \phi X_{t-1} = Z_t \quad \text{con } (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

allora, per il Teorema 3.4.2, la densità spettrale corrispondente è

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} |1 - \phi e^{-i\lambda}|^{-2} = \frac{\sigma^2}{2\pi} (1 - 2\phi \cos \lambda + \phi^2)^{-1}.$$

Nei grafici seguenti riportiamo la densità spettrale $f_X(\lambda)$, con $\sigma^2 = 6.25$, scegliendo prima $\phi = 0.7$ e successivamente $\phi = -0.7$.

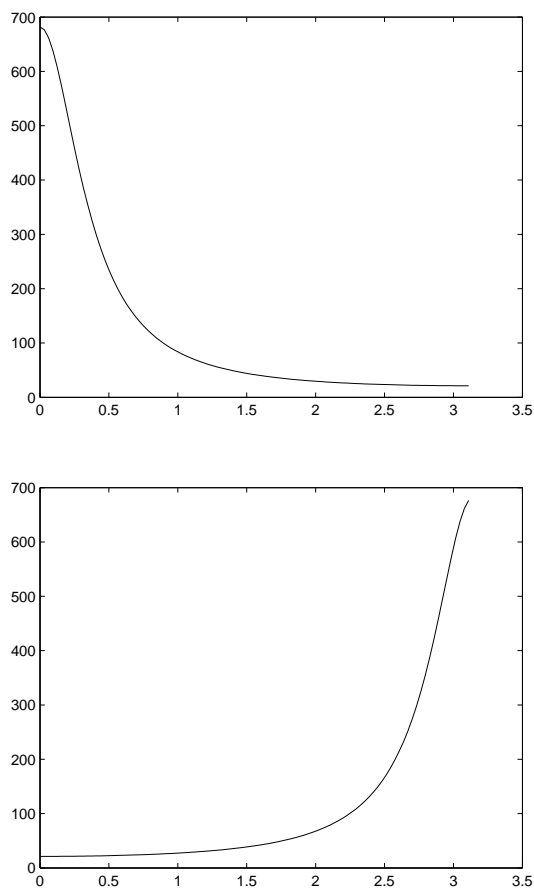


Figura 3.2: Densità spettrale $f_X(\lambda)$ di un processo AR(1)

Consideriamo il processo ARMA(p,q) $(X_t)_{t \in \mathbb{Z}}$ tale che $\phi(B)X_t = \theta(B)Z_t$, con $\phi(z)\theta(z) \neq 0$ per ogni $z \in \mathbb{C}$ tale che $|z| = 1$. Fattorizzando i polinomi $\phi(\cdot)$ e $\theta(\cdot)$ possiamo riscrivere l'equazione che definisce il processo nella forma

$$\prod_{j=1}^p (1 - a_j^{-1}B)X_t = \prod_{j=1}^q (1 - b_j^{-1}B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2), \quad (3.13)$$

dove

$$|a_j| > 1, \quad 1 \leq j \leq r, \quad |a_j| < 1, \quad r < j \leq p,$$

e

$$|b_j| > 1, \quad 1 \leq j \leq s, \quad |b_j| < 1, \quad s < j \leq q.$$

Dal Teorema 3.4.2 segue che il processo $(X_t)_{t \in \mathbb{Z}}$ ha densità spettrale

$$f_X(\lambda) = \frac{\sigma^2 \prod_{j=1}^q |1 - b_j^{-1}e^{-i\lambda}|^2}{2\pi \prod_{j=1}^p |1 - a_j^{-1}e^{-i\lambda}|^2}.$$

Definiamo

$$\tilde{\phi}(B) = \prod_{1 \leq j \leq r} (1 - a_j^{-1}B) \prod_{r < j \leq p} (1 - \bar{a}_j B)$$

e

$$\tilde{\theta}(B) = \prod_{1 \leq j \leq s} (1 - b_j^{-1}B) \prod_{s < j \leq q} (1 - \bar{b}_j B).$$

Allora il processo ARMA(p,q) $(\tilde{X}_t)_{t \in \mathbb{Z}}$ definito dall'equazione

$$\tilde{\phi}(B)\tilde{X}_t = \tilde{\theta}(B)Z_t,$$

ha densità spettrale

$$f_{\tilde{X}}(\lambda) = \frac{\sigma^2 |\tilde{\theta}(e^{-i\lambda})|^2}{2\pi |\tilde{\phi}(e^{-i\lambda})|^2}.$$

Poiché vale

$$|1 - \bar{b}_j e^{-i\lambda}| = |1 - b_j e^{i\lambda}| = |b_j| |1 - b_j^{-1} e^{-i\lambda}|,$$

possiamo riscrivere $f_{\tilde{X}}(\lambda)$ nella forma

$$f_{\tilde{X}}(\lambda) = \frac{\sigma^2 \prod_{s < j \leq q} |b_j|^2 |\theta(e^{-i\lambda})|^2}{2\pi \prod_{r < j \leq p} |a_j|^2 |\phi(e^{-i\lambda})|^2} = \frac{\prod_{s < j \leq q} |b_j|^2}{\prod_{r < j \leq p} |a_j|^2} f_X(\lambda).$$

Il processo ARMA(p,q) $(X_t^+)_{t \in \mathbb{Z}}$ definito da

$$\tilde{\phi}(B)X_t^+ = \tilde{\theta}(B)\tilde{Z}_t, \quad (\tilde{Z}_t) \sim WN \left(0, \sigma^2 \left(\prod_{r < j \leq p} |a_j| \right)^2 \left(\prod_{s < j \leq q} |b_j| \right)^{-2} \right)$$

è causale ed invertibile ed ha esattamente la stessa densità spettrale (e dunque la stessa funzione di autocovarianza) del processo ARMA (3.13). Infatti $(X_t)_{t \in \mathbb{Z}}$ stesso ha una rappresentazione causale invertibile data da

$$\tilde{\phi}(B)X_t = \tilde{\theta}(B)Z_t^*,$$

dove $(Z_t^*)_{t \in \mathbb{Z}}$ è un processo rumore bianco con la stessa varianza del processo $(\tilde{Z}_t)_{t \in \mathbb{Z}}$.

Esempio 15. Il processo ARMA

$$X_t - 2X_{t-1} = Z_t + 4Z_{t-1}, \quad \text{con } (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

non è né causale né invertibile. Introducendo i polinomi $\tilde{\phi}(z) = 1 - 0.5z$ e $\tilde{\theta}(z) = 1 + 0.25z$, otteniamo che $(X_t)_{t \in \mathbb{Z}}$ ha la rappresentazione causale ed invertibile data da

$$X_t - 0.5X_{t-1} = Z_t^* + 0.25Z_{t-1}^*, \quad \text{con } (Z_t)_{t \in \mathbb{Z}} \sim WN(0, 4\sigma^2).$$

3.4.1 Approssimazione della densità spettrale

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario a valori reali con densità spettrale continua f , allora è possibile trovare un processo causale AR(p) e un processo invertibile MA(q) con densità spettrali arbitrariamente vicine a f . Questo suggerisce che il processo $(X_t)_{t \in \mathbb{Z}}$ può essere approssimato da un processo AR(p) o MA(q). Questi risultati seguono dal prossimo teorema. Prima di enunciarlo, ricordiamo che f è la densità spettrale di un processo stazionario a valori reali se e solo se f è simmetrica, non negativa ed integrabile sull'intervallo $[-\pi, \pi]$.

Teorema 3.4.3. *Se f è una densità spettrale continua e simmetrica su $[-\pi, \pi]$, allora, per ogni $\epsilon > 0$, esiste una costante intera non negativa p e un polinomio $a(z) = \prod_{j=1}^p (1 - \eta_j^{-1}z) = 1 + a_1z + \dots + a_pz^p$ con $|\eta_j| > 1$, $j = 1, \dots, p$, tale che a_1, \dots, a_p sono coefficienti reali, per i quali vale*

$$|A|a(e^{-i\lambda})|^2 - f(\lambda)| < \epsilon \quad \forall \lambda \in [-\pi, \pi],$$

dove $A = (1 + a_1^2 + \dots + a_p^2)^{-1}(2\pi)^{-1} \int_{-\pi}^{\pi} f(\nu) d\nu$.

Dimostrazione. Se $f(\lambda) \equiv 0$, il risultato è vero con $p = 0$. Supponiamo quindi $M = \sup_{-\pi \leq \lambda \leq \pi} f(\lambda) > 0$. Per $\epsilon > 0$ prendiamo

$$\delta = \min \left\{ M, \left[4\pi M \left(\int_{-\pi}^{\pi} f(\nu) d\nu \right)^{-1} + 2 \right]^{-1} \epsilon \right\}$$

e definiamo

$$f^\delta(\lambda) = \max\{f(\lambda), \delta\}.$$

Osserviamo che f^δ è una densità spettrale continua e simmetrica con $f^\delta(\lambda) \geq \delta$ e

$$0 \leq f^\delta(\lambda) - f(\lambda) \leq \delta \quad \forall \lambda \in [-\pi, \pi]. \quad (3.14)$$

Esiste un intero r (si veda Appendice C, Teorema C.2) tale che

$$\left| r^{-1} \sum_{j=0}^{r-1} \sum_{|k| \leq j} b_k e^{-ik\lambda} - f^\delta(\lambda) \right| < \delta \quad \lambda \in [-\pi, \pi], \quad (3.15)$$

dove $b_k = (2\pi)^{-1} \int_{-\pi}^{\pi} f^\delta(\nu) e^{i\nu k} d\nu$. Scambiando l'ordine delle sommatorie e usando il fatto che f^δ è simmetrica, otteniamo

$$r^{-1} \sum_{j=0}^{r-1} \sum_{|k| \leq j} b_k e^{-ik\lambda} = \sum_{|k| < r} (1 - |k|/r) b_k e^{-ik\lambda}.$$

Questa funzione è strettamente positiva per ogni λ grazie a (3.14) e alla definizione di f^δ . Poniamo $C(z) = \sum_{|k| < r} (1 - |k|/r) b_k z^k$ e notiamo che,

se $C(m) = 0$, allora dalla proprietà di simmetria ricaviamo $C(m^{-1}) = 0$.

Prendendo $p = \max\{k : b_k \neq 0\}$, possiamo scrivere

$$z^p C(z) = K_1 \prod_{j=1}^p (1 - \eta_j^{-1} z)(1 - \eta_j z),$$

dove $K_1, \eta_1, \dots, \eta_p$ sono costanti tali che $|\eta_j| > 1$, $j = 1, \dots, p$. Questa equazione può essere riscritta come

$$K_2 a(z) a(z^{-1}) a(z^{-1}) = C(z),$$

dove $a(z)$ è il polinomio $1 + a_1 z + \dots + a_p z^p = \prod_{j=1}^p (1 - \eta_j^{-1} z)$ e $K_2 = (-1)^p \eta_1 \dots \eta_p K_1$. Eguagliando i coefficienti di z^0 relativi a ciascuna delle due parti dell'equazione precedente, otteniamo

$$K_2 = b_0 (1 + a_1^2 + \dots + a_p^2)^{-1} = (2\pi)^{-1} (1 + a_1^2 + \dots + a_p^2)^{-1} \int_{-\pi}^{\pi} f^\delta(\nu) d\nu.$$

Dalla (3.15) ricaviamo inoltre che

$$|K_2 |a(e^{-i\lambda})|^2 - f^\delta(\lambda)| < \delta \quad \forall \lambda \in [-\pi, \pi]. \quad (3.16)$$

Dalle relazioni (3.16) e (3.14) otteniamo le seguenti disequaglianze

$$\begin{aligned} (1 + a_1^2 + \dots + a_p^2)^{-1} |a(e^{-i\lambda})|^2 &\leq (f^\delta(\lambda) + \delta) 2\pi \left(\int_{-\pi}^{\pi} f^\delta(\nu) d\nu \right)^{-1} \\ &\leq 4\pi M \left(\int_{-\pi}^{\pi} f(\nu) d\nu \right)^{-1}. \end{aligned}$$

Se A è definito come nell'enunciato del teorema, allora vale

$$\begin{aligned} &|K_2 |a(e^{-i\lambda})|^2 - A |a(e^{-i\lambda})|^2| \\ &\leq (2\pi)^{-1} \left(\int_{-\pi}^{\pi} (f^\delta(\nu) - f(\nu)) d\nu \right) 4\pi M \left(\int_{-\pi}^{\pi} f(\nu) d\nu \right)^{-1} \\ &\leq 4\pi M \delta \left(\int_{-\pi}^{\pi} f(\nu) d\nu \right)^{-1}. \end{aligned} \quad (3.17)$$

Dalle disequaglianze (3.14), (3.16) e (3.17) otteniamo dunque

$$|A |a(e^{-i\lambda})|^2 - f(\lambda)| < \delta + \delta + 4\pi M \delta \left(\int_{-\pi}^{\pi} f(\nu) d\nu \right)^{-1} < \epsilon,$$

grazie alla definizione di δ . □

Corollario 3.4.4. *Sia f una densità spettrale continua e simmetrica e $\epsilon > 0$. Allora esiste un processo $AR(p)$ causale definito da*

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = Z_t \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

tale che

$$|f_X(\lambda) - f(\lambda)| < \epsilon \quad \forall \lambda \in [-\pi, \pi].$$

Dimostrazione. Sia $f^\epsilon(\lambda) = \max\{f(\lambda), \epsilon/2\}$. Allora $f^\epsilon(\lambda) \geq \epsilon/2$ e

$$0 \leq f^\epsilon(\lambda) - f(\lambda) \leq \epsilon/2 \quad \forall \lambda \in [-\pi, \pi]. \quad (3.18)$$

Poniamo $M = \max_\lambda f^\epsilon(\lambda)$ e $\delta = \min\{(2M)^{-2}\epsilon, (2M)^{-1}\}$. Applicando il Teorema 3.4.3 alla funzione $\frac{1}{f^\epsilon(\lambda)}$, otteniamo

$$|K|a(e^{-i\lambda})|^2 - \frac{1}{f^\epsilon(\lambda)}| < \delta \quad \forall \lambda \in [-\pi, \pi], \quad (3.19)$$

dove il polinomio $a(z) = 1 + a_1 z + \cdots + a_p z^p$ è diverso da zero per $|z| \leq 1$ e K è una costante positiva. Dalla definizione di δ e dalla disuguaglianza (3.19), discende la seguente maggiorazione

$$K^{-1}|a(e^{-i\lambda})|^{-2} \leq \frac{f^\epsilon(\lambda)}{1 - \delta f^\epsilon(\lambda)} \leq \frac{M}{(1 - M\delta)} \leq 2M.$$

Perciò abbiamo

$$\begin{aligned} |K^{-1}|a(e^{-i\lambda})|^{-2} - f^\epsilon(\lambda)| &= |K|a(e^{-i\lambda})|^2 - \frac{1}{f^\epsilon(\lambda)}| [K^{-1}|a(e^{-i\lambda})|^{-2} f^\epsilon(\lambda)] \\ &< 2M^2 \delta \leq \frac{\epsilon}{2}. \end{aligned} \quad (3.20)$$

Dalle disuguaglianze (3.18) e (3.20) otteniamo

$$|K^{-1}|a(e^{-i\lambda})|^{-2} - f(\lambda)| < \epsilon \quad \forall \lambda \in [-\pi, \pi]. \quad (3.21)$$

Grazie al Teorema 3.4.1, il processo $AR(p)$ definito da

$$a(B)X_t = Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, 2\pi K^{-1})$$

ha densità spettrale $K^{-1}|a(e^{-i\lambda})|^{-2}$, che grazie alla disuguaglianza (3.21) fornisce l'approssimazione richiesta per $f(\lambda)$. \square

In modo simile, possiamo ottenere il seguente corollario.

Corollario 3.4.5. *Sia f una densità spettrale continua e simmetrica e $\epsilon > 0$. Allora esiste un processo $MA(q)$ invertibile definito da*

$$X_t = Z_t + a_1 Z_{t-1} + \cdots + a_q Z_{t-q}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

tale che

$$|f_X(\lambda) - f(\lambda)| < \epsilon \quad \forall \lambda \in [-\pi, \pi],$$

dove $\sigma^2 = (1 + a_1^2 + \cdots + a_q^2)^{-1} (2\pi)^{-1} \int_{-\pi}^{\pi} f(\nu) d\nu$.

3.5 La rappresentazione spettrale

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero e con funzione di ripartizione spettrale F . Per dimostrare la rappresentazione spettrale

$$X_t = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu)$$

dobbiamo trovare, innanzitutto, un opportuno processo ad incrementi ortogonali Z . Per fare questo, definiamo un isomorfismo tra il sottospazio $\overline{\mathcal{H}} = \overline{\text{sp}}\{X_t, t \in \mathbb{Z}\}$ di $L^2(P)$ e $\overline{\mathcal{K}} = \overline{\text{sp}}\{e^{it}, t \in \mathbb{Z}\}$ di $L^2(F)$, dove con questa notazione indichiamo i sottospazi chiusi costituiti dalle combinazioni lineari finite di X_t , con $t \in \mathbb{Z}$, e di e^{it} , con $t \in \mathbb{Z}$, rispettivamente. Questo isomorfismo collega le variabili aleatorie del dominio dei tempi con le funzioni definite su $[-\pi, \pi]$ nel dominio delle frequenze.

Teorema 3.5.1. *Se F è la funzione di ripartizione spettrale di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$, allora esiste un solo isomorfismo T da $\overline{\text{sp}}\{X_t, t \in \mathbb{Z}\}$ in $L^2(F)$, tale che*

$$T(X_t) = e^{it} \quad \text{con } t \in \mathbb{Z}.$$

Dimostrazione. L'applicazione

$$T \left(\sum_{j=1}^n a_j X_{t_j} \right) = \sum_{j=1}^n a_j e^{it_j},$$

definisce un isomorfismo tra \mathcal{H} e \mathcal{K} . L'applicazione T infatti risulta ben definita: se $\|\sum_{j=1}^n a_j X_{t_j} - \sum_{k=1}^m b_k X_{t_k}\| = 0$, dalla definizione della norma in $L^2(F)$ e dal Teorema di Herglotz, otteniamo

$$\begin{aligned} & \left\| T \left(\sum_{j=1}^n a_j X_{t_j} \right) - T \left(\sum_{k=1}^m b_k X_{t_k} \right) \right\|_{L^2(F)}^2 \\ &= \int_{(-\pi, \pi]} \left| \sum_{j=1}^n a_j e^{it_j \nu} - \sum_{k=1}^m b_k e^{it_k \nu} \right|^2 dF(\nu) \\ &= E \left[\left| \sum_{j=1}^n a_j X_{t_j} - \sum_{k=1}^m b_k X_{t_k} \right|^2 \right] = 0. \end{aligned}$$

Da questo risultato ricaviamo anche la linearità di T .

Vale inoltre

$$\begin{aligned} \left\langle T \left(\sum_{j=1}^n a_j X_{t_j} \right), T \left(\sum_{k=1}^m b_k X_{s_k} \right) \right\rangle &= \sum_{j=1}^n \sum_{k=1}^m a_j \bar{b}_k \langle e^{it_j \cdot}, e^{is_k \cdot} \rangle_{L^2(F)} \\ &= \sum_{j=1}^n \sum_{k=1}^m a_j \bar{b}_k \int_{(-\pi, \pi]} e^{i(t_j - s_k) \nu} dF(\nu) = \sum_{j=1}^n \sum_{k=1}^m a_j \bar{b}_k \langle X_{t_j}, X_{s_k} \rangle \\ &= \left\langle \sum_{j=1}^n a_j X_{t_j}, \sum_{k=1}^m b_k X_{s_k} \right\rangle, \end{aligned}$$

da cui otteniamo che T è un isomorfismo tra \mathcal{H} e \mathcal{K} . L'applicazione T può essere estesa in modo unico ad un isomorfismo da $\overline{\mathcal{H}}$ a $\overline{\mathcal{K}}$. Se $Y \in \overline{\mathcal{H}}$ allora esiste una successione $(Y_n)_{n \in \mathbb{Z}} \in \mathcal{H}$ tale che $\|Y_n - Y\| \rightarrow 0$. Allora $(Y_n)_{n \in \mathbb{Z}}$ è una successione di Cauchy e, siccome T conserva la norma, anche la successione $(T(Y_n))_{n \in \mathbb{Z}}$ è di Cauchy in $L^2(F)$. La successione $(T(Y_n))_{n \in \mathbb{Z}}$ converge quindi in norma a un elemento di $\overline{\mathcal{K}}$. Se T conserva la norma su $\overline{\mathcal{H}}$, abbiamo

$$T(Y) = \lim_{n \rightarrow \infty} T(Y_n) \quad \text{in } L^2.$$

Osserviamo che, se $\|\tilde{Y}_n - Y\| \rightarrow 0$, allora la successione $T(Y_1), T(\tilde{Y}_1), T(Y_2), T(\tilde{Y}_2), \dots$ è convergente e dunque le sottosuccessioni $(T(Y_n))_{n \in \mathbb{Z}}$ e $(T(\tilde{Y}_n))_{n \in \mathbb{Z}}$

hanno lo stesso limite $T(Y)$.

Si può dimostrare che l'applicazione T estesa a $\overline{\mathcal{H}}$ risulta essere lineare e conserva il prodotto scalare.

Dal Teorema C.2 (si veda Appendice C), \mathcal{K} è denso nello spazio delle funzioni continue ϕ su $(-\pi, \pi]$ con $\phi(\pi) = \phi(-\pi)$, che a sua volta è denso nello spazio $L^2(F)$. Perciò $\overline{\mathcal{K}} = L^2(F)$. \square

Il teorema precedente serve per identificare il processo ad incrementi ortogonali $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ usato nella rappresentazione spettrale.

Proposizione 3.5.2. *Sia T l'applicazione definita nel Teorema 3.5.1, allora il processo $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ definito da*

$$Z(\lambda) = T^{-1}(I_{(-\pi, \pi]}(\cdot)), \quad \pi \leq \lambda \leq \pi,$$

è un processo ad incrementi ortogonali continuo a destra. Inoltre, la funzione di ripartizione associata a $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ è la funzione di ripartizione spettrale F del processo $(X_t)_{t \in \mathbb{Z}}$.

Dimostrazione. Dal Teorema 3.5.1 segue che, per ogni $\lambda \in [-\pi, \pi]$, il processo $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ risulta ben definito ed appartenente a $\overline{sp}\{X_t, t \in \mathbb{Z}\}$. Abbiamo dunque

$$\langle Z(\lambda), Z(\lambda) \rangle < \infty.$$

Siccome $Z(\lambda) \in \overline{sp}\{X_t, t \in \mathbb{Z}\}$, esiste una successione $(Y_n)_{n \in \mathbb{Z}}$ di elementi di $sp\{X_t, t \in \mathbb{Z}\}$ tale che $\|Y_n - Z(\lambda)\| \rightarrow 0$ per $n \rightarrow \infty$.

Dalla continuità del prodotto scalare, ricaviamo

$$\langle Z(\lambda), 1 \rangle = \lim_{n \rightarrow \infty} \langle Y_n, 1 \rangle = 0$$

(in quanto ogni X_t , e dunque ogni Y_n , ha media zero).

Se $-\pi \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \leq \pi$, allora abbiamo

$$\langle Z(\lambda_4) - Z(\lambda_3), Z(\lambda_2) - Z(\lambda_1) \rangle = \langle T(Z(\lambda_4)) - T(Z(\lambda_3)), T(Z(\lambda_2)) - T(Z(\lambda_1)) \rangle$$

$$= \langle I_{(\lambda_3, \lambda_4]}(\cdot), I_{(\lambda_1, \lambda_2]}(\cdot) \rangle_{L^2(F)} = \int_{(-\pi, \pi]} I_{(\lambda_3, \lambda_4]}(\nu) I_{(\lambda_1, \lambda_2]}(\nu) dF(\nu) = 0.$$

Il processo $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ risulta quindi essere ad incrementi ortogonali.

Con un calcolo simile al precedente, otteniamo l'uguaglianza

$$\langle Z(\mu) - Z(\lambda), Z(\mu) - Z(\lambda) \rangle = F(\mu) - F(\lambda),$$

la quale dimostra che il processo $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ è continuo a destra con funzione di ripartizione associata F . \square

A questo punto siamo pronti per provare il *Teorema di rappresentazione spettrale*.

Teorema 3.5.3. *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero e funzione di ripartizione spettrale F , allora esiste un processo ad incrementi ortogonali $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ continuo a destra, tale che*

$$(i) \ E[|Z(\lambda) - Z(-\pi)|^2] = F(\lambda) \quad \text{per } -\pi \leq \lambda \leq \pi,$$

e

$$(ii) \ X_t = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu) \quad \text{con probabilità uno.}$$

Dimostrazione. Sia $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ il processo definito nella Proposizione 3.5.2 e sia I l'isomorfismo da $L^2(F)$ su un sottospazio di $L^2(P)$, definito come

$$I(f) = \int_{(-\pi, \pi]} f(\nu) dZ(\nu).$$

(Si veda l'Appendice B.)

Se f appartiene alla classe \mathcal{D} delle funzioni in L^2 della forma

$$f(\lambda) = \sum_{i=0}^n f_i I_{(\lambda_i, \lambda_{i+1}]}(\lambda) \quad \text{con } -\pi = \lambda_0 < \lambda_1 < \dots < \lambda_{n+1} = \pi,$$

allora vale

$$I(f) = \sum_{i=0}^n f_i (Z(\lambda_{i+1}) - Z(\lambda_i)) = T^{-1}(f).$$

Questa relazione rimane valida per tutte le $f \in \overline{\mathcal{D}} = L^2(F)$ poiché sia I che T^{-1} sono isomorfismi. Deve valere dunque $I = T^{-1}$, ossia $TI(f) = f$ per ogni $f \in L^2(F)$, e quindi grazie al Teorema 3.5.1, otteniamo

$$X_t = T^{-1}(e^{it\cdot}) = I(e^{it\cdot}) = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu).$$

L'affermazione (i) del Teorema è conseguenza della Proposizione 3.5.2. \square

Corollario 3.5.4. *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero, allora esiste un processo ad incrementi ortogonali $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ continuo a destra, tale che $Z(-\pi) = 0$ e*

$$X_t = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu) \text{ con probabilità uno.}$$

Se $(Y(\lambda))_{-\pi \leq \lambda \leq \pi}$ $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ sono due processi ad incrementi ortogonali continui a destra che verificano le due precedenti condizioni, allora

$$P(Y(\lambda) = Z(\lambda)) = 1 \quad \forall \lambda \in [-\pi, \pi].$$

Dimostrazione. Denotiamo con $(Z^*(\lambda))_{-\pi \leq \lambda \leq \pi}$ il processo ad incrementi ortogonali definito dalla Proposizione 3.5.2, allora il processo

$$Z(\lambda) = Z^*(\lambda) - Z^*(-\pi), \quad -\pi \leq \lambda \leq \pi,$$

verifica la condizione $Z(-\pi) = 0$ ed ha gli stessi incrementi del processo $(Z^*(\lambda))_{-\pi \leq \lambda \leq \pi}$. Ne segue

$$X_t = \int_{(-\pi, \pi]} e^{it\nu} dZ^*(\nu) = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu).$$

Supponiamo che $(Y(\lambda))_{-\pi \leq \lambda \leq \pi}$ sia un altro processo ad incrementi ortogonali continuo a destra tale che $Y(-\pi) = 0$ e

$$X_t = \int_{(-\pi, \pi]} e^{it\nu} dY(\nu) = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu) \text{ con probabilità uno.} \quad (3.22)$$

Per ogni $f \in L^2(F)$, definiamo

$$I_Y(f) = \int_{(-\pi, \pi]} f(\nu) dY(\nu)$$

e

$$I_Z(f) = \int_{(-\pi, \pi]} f(\nu) dZ(\nu),$$

allora la (3.22) diventa

$$I_Y(e^{it\cdot}) = I_Z(e^{it\cdot}) \quad \forall t \in \mathbb{Z}.$$

Siccome I_Y e I_Z coincidono su $sp\{e^{it\cdot}, t \in \mathbb{Z}\}$, che è denso in $L^2(F)$, segue che $I_Y(f) = I_Z(f)$ per ogni f in $L^2(F)$. Scegliendo $f(\nu) = I_{(-\pi, \lambda]}(\nu)$ otteniamo

$$Y(\lambda) = \int_{(-\pi, \pi]} f(\nu) dY(\nu) = \int_{(-\pi, \pi]} f(\nu) dZ(\nu) = Z(\lambda), \quad -\pi \leq \lambda \leq \pi,$$

con probabilità uno. □

Osservazione 3.3. Nella dimostrazione del Teorema di rappresentazione spettrale, abbiamo dimostrato che $Y \in \overline{sp}\{X_t, t \in \mathbb{Z}\}$ se e solo se esiste una funzione $f \in L^2(F)$ tale che $Y = I(f) = \int_{(-\pi, \pi]} f(\nu) dZ(\nu)$. Questo significa che I è un isomorfismo di $L^2(F)$ su $\overline{sp}\{X_t, t \in \mathbb{Z}\}$, con la proprietà $I(e^{it\cdot}) = X_t$.

Proposizione 3.5.5. *Se la funzione di ripartizione spettrale F di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$ ha un punto di discontinuità in λ_0 , con $-\pi < \lambda_0 < \pi$, allora, con probabilità uno, vale che*

$$X_t = \int_{(-\pi, \pi] \setminus \lambda_0} e^{it\nu} dZ(\nu) + (Z(\lambda_0) - Z(\lambda_0^-)) e^{it\lambda_0},$$

dove i due termini del membro destro dell'eguaglianza sono non correlati e

$$\text{Var}(Z(\lambda_0) - Z(\lambda_0^-)) = F(\lambda_0) - F(\lambda_0^-).$$

Dimostrazione. Il limite sinistro $Z(\lambda_0^-)$ è definito come

$$Z(\lambda_0^-) = \lim_{n \rightarrow \infty} Z(\lambda_n) \quad \text{in } L^2,$$

dove λ_n è una successione tale che $\lambda_n \nearrow \lambda_0$. Questo limite esiste perché $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ è una successione di Cauchy, poiché

$$\|Z(\lambda_n) - Z(\lambda_m)\|^2 = |F(\lambda_n) - F(\lambda_m)| \rightarrow 0 \quad \text{per } m, n \rightarrow \infty.$$

Inoltre, se $\nu_n \nearrow \lambda_0$ per $n \rightarrow \infty$, allora

$$\|Z(\lambda_n) - Z(\nu_n)\|^2 = |F(\lambda_n) - F(\nu_n)| \rightarrow 0 \quad \text{per } n \rightarrow \infty$$

e dunque il limite precedente risulta essere uguale per ogni successione non-decrescente con limite λ_0 .

Per $\delta > 0$, poniamo $\lambda_{\pm\delta} = \lambda_0 \pm \delta$. Per $0 < \delta < \pi - |\lambda_0|$, grazie alla rappresentazione spettrale, possiamo scrivere

$$X_t = \int_{(-\pi, \pi] \setminus (\lambda_{-\delta}, \lambda_\delta]} e^{it\nu} dZ(\nu) + \int_{(\lambda_{-\delta}, \lambda_\delta]} e^{it\nu} dZ(\nu). \quad (3.23)$$

I due termini al membro destro sono non correlati in quanto le regioni di integrazione sono disgiunte. Dal fatto che

$$e^{it\cdot} I_{(-\pi, \pi] \setminus (\lambda_{-\delta}, \lambda_\delta]} \rightarrow e^{it\cdot} I_{(-\pi, \pi] \setminus \{\lambda_0\}} \text{ in } L^2(F),$$

segue che, per $\delta \rightarrow 0$, il primo termine del membro destro della (3.23) converge in media quadratica verso $\int_{(-\pi, \pi] \setminus \{\lambda_0\}} e^{it\nu} dZ(\nu)$. Per studiare il comportamento asintotico per $\delta \rightarrow 0$, del secondo termine del membro destro della (3.23), usiamo la seguente diseguaglianza

$$\begin{aligned} & \left\| \int_{(\lambda_{-\delta}, \lambda_\delta]} e^{it\nu} dZ(\nu) - e^{it\lambda_0} (Z(\lambda_0) - Z(\lambda_0^-)) \right\| \\ & \leq \left\| \int_{(\lambda_{-\delta}, \lambda_\delta]} e^{it\nu} dZ(\nu) - e^{it\lambda_0} (Z(\lambda_\delta) - Z(\lambda_{-\delta})) \right\| \\ & \quad + \|Z(\lambda_\delta) - Z(\lambda_{-\delta}) - (Z(\lambda_0) - Z(\lambda_0^-))\|. \end{aligned} \quad (3.24)$$

Per $\delta \rightarrow 0$ il secondo termine della somma che compare nella (3.24) tende verso zero per la continuità a destra di $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$ e per come è definito $Z(\lambda_0^-)$. Per quanto riguarda il primo termine della somma, osserviamo che vale

$$\begin{aligned} & \left\| \int_{(-\pi, \pi]} (e^{it\nu} - e^{it\lambda_0}) I_{(\lambda_{-\delta}, \lambda_\delta]}(\nu) dZ(\nu) \right\| = \|(e^{it\nu} - e^{it\lambda_0}) I_{(\lambda_{-\delta}, \lambda_\delta]}(\cdot)\|_{L^2(F)} \\ & \leq \left[\sup_{\lambda_{-\delta} \leq \lambda \leq \lambda_\delta} |e^{it\nu} - e^{it\lambda_0}|^2 (F(\lambda_\delta) - F(\lambda_{-\delta})) \right]^{\frac{1}{2}} \rightarrow 0 \quad \text{per } \delta \rightarrow 0 \end{aligned}$$

grazie alla continuità della funzione $e^{it\cdot}$. Dalla (3.24) otteniamo dunque

$$\int_{(\lambda-\delta, \lambda\delta]} e^{it\nu} dZ(\nu) \rightarrow e^{it\lambda_0} (Z(\lambda_0) - Z(\lambda_0^-)) \text{ in } L^2 \text{ per } \delta \rightarrow 0.$$

La continuità del prodotto scalare e l'ortogonalità dei due integrali in (3.23) garantiscono che anche i loro limiti in L^2 siano ortogonali.

Inoltre vale

$$\text{Var}(Z(\lambda_0) - Z(\lambda_0^-)) = \lim_{\lambda_n \nearrow \lambda_0} \text{Var}(Z(\lambda_0) - Z(\lambda_n)) = F(\lambda_0) - F(\lambda_0^-).$$

□

Osservazione 3.4. Il processo $Y_t = (Z(\lambda_0) - Z(\lambda_0^-))e^{it\lambda_0}$ è detto *deterministico*, in quanto, se conosciamo Y_{t_0} per qualche t_0 , allora Y_t è determinato per tutti i t . L'esistenza di una discontinuità della funzione di ripartizione spettrale in una frequenza λ_0 indica la presenza nella serie storica di una componente sinusoidale con frequenza λ_0 .

Se la funzione di ripartizione spettrale ha k punti di discontinuità $\lambda_1, \dots, \lambda_k$, allora la rappresentazione del processo $(X_t)_{t \in \mathbb{Z}}$ sarà

$$X_t = \int_{(-\pi, \pi] \setminus \{\lambda_1, \dots, \lambda_k\}} e^{it\nu} dZ(\nu) + \sum_{j=1}^k (Z(\lambda_j) - Z(\lambda_j^-)) e^{it\lambda_j},$$

dove i $k + 1$ termini del membro destro sono non correlati.

Capitolo 4

La previsione di un processo stazionario

In questo capitolo esporremo delle tecniche di previsione dei valori $\{X_t, t \geq n + 1\}$ di un processo stazionario, a partire dai valori osservati $\{X_1, \dots, X_n\}$. L'idea base è che, dato un sottospazio chiuso \mathcal{M} di $L^2(P)$, la migliore approssimazione di X_{n+h} con un elemento di \mathcal{M} si ottiene prendendo quel elemento di \mathcal{M} che minimizza l'errore quadratico medio.

4.1 La previsione nel dominio dei tempi

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media μ e funzione di autocovarianza $\gamma_X(\cdot)$. Allora il processo $(Y_t)_{t \in \mathbb{Z}} = (X_t - \mu)_{t \in \mathbb{Z}}$ è un processo stazionario con media zero e funzione di autocovarianza $\gamma_X(\cdot)$, e vale la seguente eguaglianza

$$P_{\overline{sp}\{1, X_1, \dots, X_n\}} X_{n+h} = \mu + P_{\overline{sp}\{Y_1, \dots, Y_n\}} Y_{n+h}.$$

Senza ledere la generalità, possiamo quindi supporre $\mu = 0$. In questo caso la precedente eguaglianza diventa

$$P_{\overline{sp}\{1, X_1, \dots, X_n\}} X_{n+h} = P_{\overline{sp}\{X_1, \dots, X_n\}} X_{n+h}.$$

Denotiamo con \mathcal{H}_n il sottospazio chiuso $\overline{\text{sp}}\{X_1, \dots, X_n\}$, per $n \geq 1$, e con \widehat{X}_{n+1} , per $n \geq 0$, il “predittore ad un passo” definito come

$$\widehat{X}_{n+1} = \begin{cases} 0 & \text{se } n = 0 \\ P_{\mathcal{H}_n} X_{n+1} & \text{se } n \geq 1. \end{cases} \quad (4.1)$$

Poiché \widehat{X}_{n+1} appartiene a \mathcal{H}_n , per $n \geq 1$, possiamo scrivere

$$\widehat{X}_{n+1} = \phi_{n1} X_n + \dots + \phi_{nn} X_1 = \sum_{i=1}^n \phi_{ni} X_{n-i+1}, \quad (4.2)$$

dove $\phi_{n1}, \dots, \phi_{nn}$ verificano la relazione

$$\left\langle \sum_{i=1}^n \phi_{ni} X_{n-i+1}, X_{n-j+1} \right\rangle = \langle X_{n+1}, X_{n+1-j} \rangle, \quad j = 1, \dots, n,$$

con $\langle X, Y \rangle = E[XY]$. Dalla proprietà di linearità del prodotto scalare possiamo riscrivere le precedenti equazioni come

$$\sum_{i=1}^n \phi_{ni} \gamma_X(i-j) = \gamma_X(j), \quad j = 1, \dots, n,$$

o, equivalentemente,

$$\Gamma_n \phi_n = \gamma_n \quad (4.3)$$

dove $\Gamma_n = [\gamma_X(i-j)]_{i,j=1,\dots,n}$, $\gamma_n = (\gamma_X(1), \dots, \gamma_X(n))'$ e $\phi_n = (\phi_{n1}, \dots, \phi_{nn})'$. Il Teorema della proiezione ortogonale (vedi Appendice A), assicura che l'equazione (4.3) ha almeno una soluzione. Le equazioni (4.2) e (4.3) sono conosciute come *equazioni di previsione ad un passo*.

Notiamo che l'equazione (4.3) può avere molte soluzioni, ma che, una volta sostituite nell'equazione (4.2), devono dare lo stesso predittore \widehat{X}_{n+1} . Sempre grazie al Teorema della proiezione ortogonale, sappiamo infatti che \widehat{X}_{n+1} è univocamente definito. Osserviamo inoltre che esiste una sola soluzione della (4.3) se e solo se la matrice Γ_n è invertibile e, in tal caso, la soluzione è data da

$$\phi_n = \Gamma_n^{-1} \gamma_n$$

4.2 Alcuni metodi ricorsivi per calcolare il migliore predittore lineare

e l'errore di previsione, ossia l'errore quadratico medio, risulta essere

$$v_n = E[(\widehat{X}_{n+1} - X_{n+1})^2] = \gamma_X(0) - \gamma_n' \Gamma_n^{-1} \gamma_n.$$

Si può dimostrare (si veda Proposizione 5.1.1 di [7]) che, se $\gamma_X(0) > 0$ e $\gamma_X(h) \rightarrow 0$ per $h \rightarrow \infty$, allora la matrice Γ_n è invertibile per ogni n .

Possiamo procedere con lo stesso metodo per calcolare il migliore predittore di X_{n+h} in termini di X_1, \dots, X_n per qualche $h \geq 1$:

$$P_{\mathcal{Y}_n} X_{n+h} = \phi_{n1}^{(h)} X_n + \dots + \phi_{nn}^{(h)} X_1, \quad \text{per } n, h \geq 1, \quad (4.4)$$

dove $\phi_{nn}^{(h)} = (\phi_{n1}^{(h)}, \dots, \phi_{nn}^{(h)})'$ è una soluzione, unica se Γ_n è invertibile, di

$$\Gamma_n \phi_n^{(h)} = \gamma_n^{(h)}, \quad (4.5)$$

dove $\gamma_n^{(h)} = (\gamma_X(h), \gamma_X(h+1), \dots, \gamma_X(n+h-1))'$.

4.2 Alcuni metodi ricorsivi per calcolare il migliore predittore lineare

In questo paragrafo illustreremo due diversi algoritmi per calcolare il predittore ad un passo \widehat{X}_{n+1} , per $n \geq 1$, definito nell'equazione (4.1). Successivamente mostreremo come questi possano essere utilizzati per calcolare il predittore ad h passi $P_{\mathcal{Y}_n} X_{n+h}$, con $h \geq 1$. Questi algoritmi sono di notevole importanza perché il calcolo diretto di $P_{\mathcal{Y}_n} X_{n+h}$ usando le equazioni (4.4) e (4.5) richiede, quando n è grande, la risoluzione di un sistema di equazioni lineari molto grande. Inoltre ogni volta che il numero delle osservazioni aumenta, è necessario ripetere l'intera procedura.

4.2.1 L'algoritmo di Durbin-Levinson

Precedentemente abbiamo visto che il migliore predittore di X_{n+1} a partire da X_1, \dots, X_n è dato da

$$\widehat{X}_{n+1} = \sum_{i=1}^n \phi_{ni} X_{n+1-i}, \quad n \geq 1,$$

dove l'errore di previsione è l'errore quadratico medio

$$v_n = E[(X_{n+1} - \widehat{X}_{n+1})^2], \quad n \geq 1,$$

con $v_0 = \gamma_X(0)$. Il seguente algoritmo di Durbin-Levison ci permette di calcolare ricorsivamente $\phi_n = (\phi_{n1}, \dots, \phi_{nn})'$ e v_n per $n \geq 1$.

Proposizione 4.2.1 (Algoritmo di Durbin-Levinson). *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero e con funzione di autocovarianza $\gamma_X(\cdot)$, tale che $\gamma_X(0) > 0$ e $\gamma_X(h) \rightarrow 0$ quando $h \rightarrow \infty$, allora, per calcolare ϕ_n e v_n , possiamo eseguire i seguenti passi:*

1. poniamo $\phi_{11} = \frac{\gamma_X(1)}{\gamma_X(0)}$ e $v_0 = \gamma_X(0)$,
2. calcoliamo $\phi_{nn} = \left[\gamma_X(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma_X(n-j) \right] v_{n-1}^{-1}$,
3. calcoliamo

$$\begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{nn} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$

$$\text{e } v_n = v_{n-1}[1 - \phi_{nn}^2].$$

Dimostrazione. Osserviamo che

$$\mathcal{K}_1 = \overline{\text{sp}}\{X_2, \dots, X_n\} \quad \text{e} \quad \mathcal{K}_2 = \overline{\text{sp}}\{X_1 - P_{\mathcal{K}_1} X_1\}$$

sono sottospazi ortogonali di $\mathcal{H}_n = \overline{\text{sp}}\{X_1, \dots, X_n\}$. Ne segue che, se $Y \in L^2(P)$, allora $P_{\mathcal{H}_n} Y = P_{\mathcal{K}_1} Y + P_{\mathcal{K}_2} Y$. Otteniamo quindi

$$\widehat{X}_{n+1} = P_{\mathcal{K}_1} X_{n+1} + P_{\mathcal{K}_2} X_{n+1} = P_{\mathcal{K}_1} X_{n+1} + a(X_1 - P_{\mathcal{K}_1} X_1), \quad (4.6)$$

dove

$$a = \frac{\langle X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1 \rangle}{\|X_1 - P_{\mathcal{K}_1} X_1\|^2}. \quad (4.7)$$

4.2 Alcuni metodi ricorsivi per calcolare il migliore predittore lineare

Grazie alla stazionarietà del processo, il vettore aleatorio $(X_1, \dots, X_n)'$ ha la stessa matrice di covarianza di $(X_n, X_{n-1}, \dots, X_1)'$ e di $(X_2, \dots, X_{n+1})'$, così abbiamo

$$P_{\mathcal{K}_1} X_1 = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{j+1}, \quad (4.8)$$

$$P_{\mathcal{K}_1} X_{n+1} = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{n+1-j}, \quad (4.9)$$

e

$$\|X_1 - P_{\mathcal{K}_1} X_1\|^2 = \|X_{n+1} - P_{\mathcal{K}_1} X_{n+1}\|^2 = \|X_n - \widehat{X}_n\|^2 = v_{n-1}. \quad (4.10)$$

Dalle equazioni (4.6), (4.8) e (4.9), otteniamo

$$\widehat{X}_{n+1} = aX_1 + \sum_{j=1}^{n-1} [\phi_{n-1,j} - a\phi_{n-1,n-j}] X_{n+1-j}, \quad (4.11)$$

dove, grazie alle equazioni (4.7) e (4.8), ricaviamo

$$\begin{aligned} a &= \left(\langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} \phi_{n-1,j} \langle X_{n+1}, X_{j+1} \rangle \right) v_{n-1}^{-1} \\ &= \left[\gamma_X(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma_X(n-j) \right] v_{n-1}^{-1}. \end{aligned}$$

L'ipotesi $\gamma_X(h) \rightarrow 0$ quando $h \rightarrow \infty$ garantisce l'unicità della rappresentazione

$$\widehat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}. \quad (4.12)$$

Uguagliando i coefficienti in (4.11) e in (4.12), deduciamo

$$\phi_{nn} = a \quad \text{e} \quad \phi_{nj} = \phi_{n-1,j} - a\phi_{n-1,n-j}, \quad j = 1, \dots, n-1.$$

Da ultimo dobbiamo calcolare l'errore di previsione che risulta essere

$$\begin{aligned} v_n &= \|X_{n+1} - \widehat{X}_{n+1}\|^2 = \|X_{n+1} - P_{\mathcal{K}_1} X_{n+1} - P_{\mathcal{K}_2} X_{n+1}\|^2 \\ &= \|X_{n+1} - P_{\mathcal{K}_1} X_{n+1}\|^2 + \|P_{\mathcal{K}_2} X_{n+1}\|^2 - 2 \langle X_{n+1} - P_{\mathcal{K}_1} X_{n+1}, P_{\mathcal{K}_2} X_{n+1} \rangle \end{aligned}$$

$$= v_{n-1} + a^2 v_{n-1} - 2a \langle X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1 \rangle,$$

dove abbiamo usato la (4.10), l'ortogonalità di \mathcal{K}_1 e \mathcal{K}_2 e l'uguaglianza $P_{\mathcal{K}_2} X_{n+1} = a(X_1 - P_{\mathcal{K}_1} X_1)$. Utilizzando l'equazione (4.7), otteniamo

$$v_n = v_{n-1}(1 - a^2)$$

come richiesto. □

Osservazione 4.1. Nel primo capitolo avevamo definito la funzione di autocorrelazione parziale come

$$\alpha_X(n) = \text{Corr}(X_{n+1} - P_{\overline{\text{sp}}\{X_2, \dots, X_n\}} X_{n+1}, X_1 - P_{\overline{\text{sp}}\{X_2, \dots, X_n\}} X_1)$$

e

$$\alpha_X(n) = \phi_{nn}.$$

Queste due definizioni sono effettivamente equivalenti: infatti dalla dimostrazione precedente ricaviamo che

$$\begin{aligned} \phi_{nn} &= \langle X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1 \rangle \|X_1 - P_{\mathcal{K}_1} X_1\|^2 \\ &= \langle X_{n+1} - P_{\mathcal{K}_1} X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1 \rangle \|X_1 - P_{\mathcal{K}_1} X_1\|^2 \\ &= \text{Corr}(X_{n+1} - P_{\mathcal{K}_1} X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1). \end{aligned}$$

Esempio 16 (Previsione di un processo AR(2)). Sia dato il processo stazionario AR(2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

dove ϕ_1 e ϕ_2 sono costanti note, tali che il processo risulti causale.

Per calcolare ricorsivamente la previsione di \hat{X}_{n+1} per $n = 1, 2, \dots$ dobbiamo calcolare prima la funzione di autocovarianza del processo $(X_t)_{t \in \mathbb{Z}}$, $\gamma_X(0), \gamma_X(1), \gamma_X(2), \dots$. Possiamo calcolarla moltiplicando il processo per X_{t-k} :

$$\begin{aligned} \gamma_X(0) - \phi_1 \gamma_X(1) - \phi_2 \gamma_X(2) &= \sigma^2, \\ \gamma_X(1) - \phi_1 \gamma_X(0) - \phi_2 \gamma_X(1) &= 0, \end{aligned}$$

4.2 Alcuni metodi ricorsivi per calcolare il migliore predittore lineare

$$\gamma_X(2) - \phi_1 \gamma_X(1) - \phi_2 \gamma_X(0) = 0,$$

ottenendo

$$\gamma_X(0) = \frac{1 - \phi_2}{(1 + \phi_2)(1 - (\phi_1 + \phi_2)(\phi_1 + 1 - \phi_2))} \sigma^2,$$

$$\gamma_X(1) = \frac{\phi_1}{(1 + \phi_2)(1 - (\phi_1 + \phi_2)(\phi_1 + 1 - \phi_2))} \sigma^2,$$

$$\gamma_X(2) = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{(1 + \phi_2)(1 - (\phi_1 + \phi_2)(\phi_1 + 1 - \phi_2))} \sigma^2.$$

Per $k > 2$, abbiamo

$$\gamma_X(k) = \phi_1 \gamma_X(k-1) + \phi_2 \gamma_X(k-2). \quad (4.13)$$

Adesso applichiamo l'algoritmo di Durbin-Levinson, prendendo $\sigma = 1$:

- $n = 1$

$$\phi_{00} = 0, \quad v_0 = \gamma_X(0),$$

$$\phi_{11} = \gamma_X(1)v_0^{-1} = \frac{\gamma_X(1)}{\gamma_X(0)} = \rho_X(1) = \frac{\phi_1}{1 - \phi_2},$$

$$v_1 = v_0(1 - \phi_{11}^2) = \gamma_X(0) \left[1 - \frac{\phi_1^2}{(1 - \phi_2)^2} \right].$$

Ricordiamo che nel caso di un processo stazionario, la funzione di autocorrelazione è pari a $\rho_X(h) = \gamma_X(h)/\gamma_X(0)$, con $h \in \mathbb{Z}$.

Il migliore predittore lineare di X_2 a partire da X_1 risulta quindi essere

$$\widehat{X}_2 = \phi_{11}X_1 = \frac{\phi_1}{1 - \phi_2}X_1.$$

- $n = 2$

$$\phi_{22} = (\gamma_X(2) - \phi_{11}\gamma_X(1))v_1^{-1} = \frac{\gamma_X(2) - \frac{\gamma_X(1)}{\gamma_X(0)}\gamma_X(1)}{\gamma_X(0) \left[1 - \left(\frac{\gamma_X(1)}{\gamma_X(0)} \right)^2 \right]}$$

$$\begin{aligned}
&= \frac{\rho_X(2) - \rho_X^2(1)}{1 - \rho_X^2(1)} = \frac{(\phi_1^2 + \phi_2 - \phi_2^2)(1 - \phi_2) - \phi_1^2}{(1 - \phi_2)^2 - \phi_1^2} \\
&= \frac{\phi_2(1 - 2\phi_2 - \phi_1^2 + \phi_2^2)}{1 - 2\phi_2 + \phi_2^2 - \phi_1^2} = \phi_2,
\end{aligned}$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \frac{\phi_1}{1 - \phi_2}(1 - \phi_2) = \phi_1,$$

$$v_2 = v_1(1 - \phi_{22}^2) = \gamma_X(0) \left[1 - \frac{\phi_1^2}{(1 - \phi_2)^2} \right] (1 - \phi_2^2).$$

Il migliore predittore di X_3 a partire da X_1 e X_2 è quindi

$$\hat{X}_3 = \phi_{22}X_1 + \phi_{21}X_2 = \phi_2X_1 + \phi_1X_2.$$

- $n = 3$ Continuando, otteniamo

$$\phi_{33} = \dots = \frac{\rho_X(3) - \phi_1\rho_X(2) - \phi_2\rho_X(1)}{1 - \phi_1\rho_X(1) - \phi_2\rho_X(2)},$$

ma dall'equazione (4.13) segue che il numeratore è pari a zero. Dunque abbiamo

$$\begin{pmatrix} \phi_{31} \\ \phi_{32} \end{pmatrix} = \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} - \phi_{33} \begin{pmatrix} \phi_{22} \\ \phi_{21} \end{pmatrix} = \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}.$$

Il migliore predittore di X_4 dati X_1 , X_2 , e X_3 è

$$\hat{X}_4 = \phi_{33}X_1 + \phi_{32}X_2 + \phi_{31}X_3 = \phi_2X_2 + \phi_1X_3.$$

Per tutti gli $n \geq 3$, avremo $\phi_{nn} = 0$ e

$$\hat{X}_{n+1} = \phi_1X_n + \phi_2X_{n-1}.$$

Otteniamo anche la funzione di autocorrelazione parziale del processo AR(2)

$$\alpha_X(1) = \phi_{11} = \frac{\phi_1}{1 - \phi_2}, \quad \alpha_X(2) = \phi_{22} = \phi_2, \quad \alpha_X(3) = \phi_{33} = 0,$$

e

$$\alpha_X(n) = \phi_{nn} = 0 \text{ per } n > 2.$$

4.2.2 L'algoritmo delle innovazioni

L'idea centrale dell'algoritmo di Durbin-Levison è la decomposizione dello spazio \mathcal{H}_n in due sottospazi ortogonali \mathcal{K}_1 e \mathcal{K}_2 .

Il seguente algoritmo consiste invece nella decomposizione di \mathcal{H}_n in n sottospazi ortogonali. Esso inoltre ha una maggiore applicabilità rispetto all'algoritmo di Durbin-Levison, perché non è necessario che il processo $(X_t)_{t \in \mathbb{Z}}$ sia stazionario.

Come prima definiamo $\mathcal{H}_n = \overline{\text{sp}}\{X_1, \dots, X_n\}$, \widehat{X}_{n+1} come nella (4.1), e $v_n = \|X_{n+1} - \widehat{X}_{n+1}\|^2$. Chiaramente, ponendo $\widehat{X}_1 := 0$, abbiamo per $n \geq 1$

$$\mathcal{H}_n = \overline{\text{sp}}\{X_1 - \widehat{X}_1, X_2 - \widehat{X}_2, \dots, X_n - \widehat{X}_n\}$$

e quindi

$$\widehat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}).$$

Illustriamo adesso l'algoritmo per calcolare $\{\theta_{nj}, j = 1, \dots, n; v_n\}$, per $n \geq 1$.

Proposizione 4.2.2 (Algoritmo delle innovazioni). *Sia $(X_t)_{t \in \mathbb{Z}}$ un processo con media zero e $E[X_i X_j] = \gamma_X(i, j)$, dove la matrice $[\gamma_X(i, j)]_{i,j=1}^n$ è invertibile per ogni $n \geq 1$, allora i predittori ad un passo \widehat{X}_{n+1} , per $n \geq 0$, e i relativi errori di previsione v_n , per $n \geq 1$, sono dati da*

$$\widehat{X}_{n+1} = \begin{cases} 0 & \text{se } n = 0 \\ \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}) & \text{se } n \geq 1 \end{cases} \quad (4.14)$$

e

$$\begin{cases} v_0 = \gamma_X(1, 1), \\ \theta_{n,n-k} = v_k^{-1} \left(\gamma_X(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right), & k = 0, 1, \dots, n-1, \\ v_n = \gamma_X(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \end{cases} \quad (4.15)$$

Risulta semplice risolvere il precedente sistema, calcolando $v_0; \theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3$ e così via.

Dimostrazione. Grazie alla definizione di \widehat{X}_j , abbiamo che $(X_i - \widehat{X}_i)$ appartiene a \mathcal{H}_{j-1} , per $i < j$, e che $(X_j - \widehat{X}_j) \perp \mathcal{H}_{j-1}$. Ne segue che, l'insieme $\{X_1 - \widehat{X}_1, X_2 - \widehat{X}_2, \dots, X_n - \widehat{X}_n\}$ è ortogonale. Prendendo il prodotto scalare con $(X_{k+1} - \widehat{X}_{k+1})$, per $0 \leq k < n$, in entrambi i membri dell'equazione (4.14), otteniamo

$$\langle \widehat{X}_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle = \theta_{n,n-k} v_k.$$

Poiché $(X_{n+1} - \widehat{X}_{n+1}) \perp (X_{k+1} - \widehat{X}_{k+1})$, ricaviamo

$$\theta_{n,n-k} = v_k^{-1} \langle X_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle \quad \text{per } k = 0, \dots, n-1. \quad (4.16)$$

Usando la (4.14) otteniamo

$$\theta_{n,n-k} = v_k^{-1} \left(\gamma_X(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \widehat{X}_{j+1} \rangle \right).$$

Per la (4.16), abbiamo $\langle X_{n+1}, X_{j+1} - \widehat{X}_{j+1} \rangle = v_j \theta_{n,n-j}$, per $0 \leq j < n$, e quindi possiamo riscrivere la precedente uguaglianza come

$$\theta_{n,n-k} = v_k^{-1} \left(\gamma_X(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right).$$

Dal Teorema della proiezione ortogonale e dalle sue proprietà (vedi Appendice A), ricaviamo

$$v_n = \|X_{n+1} - \widehat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\widehat{X}_{n+1}\|^2 = \gamma_X(n+1, n+1) - \sum_{k=0}^{n+1} \theta_{n,n-k}^2 v_k,$$

cioè l'equazione (4.15). □

Osservazione 4.2. Osserviamo che, mentre l'algoritmo di Durbin-Levinson fornisce i coefficienti di X_1, \dots, X_n nella rappresentazione

$$\widehat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j},$$

4.2 Alcuni metodi ricorsivi per calcolare il migliore predittore lineare

l'algoritmo della proposizione precedente fornisce i coefficienti dell'innovazione $(X_j - \widehat{X}_j)$, per $j = 1, \dots, n$, nella rappresentazione ortogonale

$$\widehat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}).$$

Questa seconda eguaglianza è molto facile da usare e, nel caso di un processo $ARMA(p, q)$, può essere semplificata ulteriormente come vedremo nel prossimo paragrafo.

Il secondo algoritmo ci dà inoltre una rappresentazione per il processo X_{n+1} , ponendo $\theta_{n0} = 1$, e

$$X_{n+1} = \sum_{j=0}^n \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}), \quad n \geq 0.$$

Esempio 17 (Previsione di un processo MA(1)). Sia $(X_t)_{t \in \mathbb{Z}}$ il processo

$$X_t = Z_t + \theta Z_{t-1}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

allora $\gamma_X(i, j) = 0$ per $|i - j| > 1$, $\gamma_X(i, i) = \sigma^2(1 + \theta^2)$ e $\gamma_X(i, i + 1) = \theta\sigma^2$. Utilizziamo dapprima l'algoritmo delle innovazioni: dall'equazione (4.15) ricaviamo che

$$\begin{aligned} \theta_{nj} &= 0, \quad 2 \leq j \leq n, \\ \theta_{n1} &= v_{n-1}^{-1} \theta \sigma^2, \\ v_0 &= (1 + \theta^2) \sigma^2, \end{aligned}$$

e

$$v_n = [1 + \theta^2 - v_{n-1}^{-1} \theta^2 \sigma^2] \sigma^2.$$

Ponendo $r_n = v_n / \sigma^2$, possiamo scrivere il predittore come

$$\widehat{X}_{n+1} = \frac{\theta(X_n - \widehat{X}_n)}{r_{n-1}},$$

dove $r_0 = 1 + \theta^2$ ed $r_{n+1} = 1 + \theta^2 - \theta^2 / r_n$. Usiamo le precedenti relazioni per calcolare \widehat{X}_6 basandoci sulle osservazioni X_1, \dots, X_5 , con $\theta = -0.9$.

Notiamo inoltre che v_n è non decrescente rispetto a n e, dal fatto che $\|X_n - \widehat{X}_n - Z_n\| \rightarrow 0$ quando $n \rightarrow \infty$, abbiamo che $v_n \rightarrow \sigma^2$. Nella seguente tabella riportiamo i risultati ottenuti.

t	X_{t+1}	\widehat{X}_{t+1}	v_t
0	-2.58	0	1.810
1	1.62	1.28	1.362
2	-0.96	-0.22	1.215
3	2.62	0.55	1.144
4	-1.36	-1.63	1.102
5		-0.22	1.075

Adesso utilizziamo l'algoritmo di Durbin-Levinson:

$$v_0 = 1.810$$

$$\phi_{11} = -0.4972, v_1 = 1.362$$

$$\phi_{22} = -0.3285, \phi_{21} = -0.6605, v_2 = 1.215$$

$$\phi_{33} = -0.2433, \phi_{32} = -0.4892, \phi_{31} = -0.7404, v_3 = 1.144$$

$$\phi_{44} = -0.1914, \phi_{43} = -0.3850, \phi_{42} = -0.5828, \phi_{41} = -0.7870, v_4 = 1.102$$

$$\phi_{55} = -0.1563, \phi_{54} = -0.3144, \phi_{53} = -0.4761, \phi_{52} = -0.6430,$$

$$\phi_{51} = -0.8169, v_5 = 1.075,$$

ed otteniamo $\widehat{X}_6 = \phi_{55}X_1 + \dots + \phi_{51}X_5 = -0.22$ come nel caso dell'algoritmo delle innovazioni. Osserviamo che le costanti ϕ_{nm} , $n = 1, 2, \dots, 5$, sono rispettivamente le autocorrelazioni parziali ai passi $1, 2, \dots, 5$.

4.2.3 Il calcolo della previsione a più passi

Adottiamo la notazione P_n per l'operatore proiezione $P_{\mathcal{H}_n}$. Allora la previsione ad h passi $P_n X_{n+h}$ può essere calcolata utilizzando l'algoritmo delle innovazioni e le proprietà della proiezione ortogonale (vedi Appendice A), per $h \geq 1$,

$$\begin{aligned} P_n X_{n+h} &= P_n P_{n+h-1} X_{n+h} = P_n \widehat{X}_{n+h} \\ &= P_n \left(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}) \right). \end{aligned}$$

Poiché $(X_{n+h-j} - \widehat{X}_{n+h-j}) \perp \mathcal{H}_n$, per $j < h$, ricaviamo che

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), \quad (4.17)$$

dove i coefficienti θ_{nj} sono determinati dalle equazioni (4.15). Inoltre l'errore quadratico medio può essere espresso come

$$\begin{aligned} E[(X_{n+h} - P_n X_{n+h})^2] &= \|X_{n+h}\|^2 - \|P_n X_{n+h}\|^2 \\ &= \gamma_X(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-j-1}. \end{aligned} \quad (4.18)$$

4.3 La previsione di un processo ARMA

Possiamo applicare l'algoritmo delle innovazioni direttamente ad un processo ARMA causale

$$\phi(B)X_t = \theta(B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

dove ϕ e θ sono i polinomi di grado p e q rispettivamente dati da $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ e $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$. È possibile però ottenere una notevole semplificazione dei calcoli, applicando l'algoritmo al processo trasformato

$$\begin{cases} W_t = \sigma^{-1} X_t & t = 1, \dots, m, \\ W_t = \sigma^{-1} \phi(B) X_t, & t > m, \end{cases} \quad (4.19)$$

dove $m = \max(p, q)$.

Senza ledere la generalità, assumiamo $\theta_0 = 1$, $p \geq 1$ e $q \geq 1$. Per $n \geq 1$ consideriamo il sottospazio

$$\mathcal{H}_n = \overline{\text{sp}}\{X_1, \dots, X_n\} = \overline{\text{sp}}\{W_1, \dots, W_n\}.$$

Per $n \geq 1$, denotiamo la proiezione su \mathcal{H}_n di X_{n+1} e W_{n+1} , rispettivamente con \widehat{X}_{n+1} e \widehat{W}_{n+1} e poniamo $\widehat{X}_1 = \widehat{W}_1 = 0$.

Per calcolare la funzione di autocovarianza $\gamma_X(\cdot)$ di $(X_t)_{t \in \mathbb{Z}}$ utilizziamo uno dei metodi descritti nel paragrafo 2.4. Le funzioni di autocovarianza

$\gamma_W(i, j) = E[W_i W_j]$ risultano definite da

$$\gamma_W(i, j) = \begin{cases} \sigma^{-2} \gamma_X(i - j) & 1 \leq i, j \leq m, \\ \sigma^{-2} [\gamma_X(i - j) - \sum_{r=1}^p \phi_r \gamma_X(r - |i - j|)] & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|} & \min(i, j) > m, \\ 0 & \text{altrimenti,} \end{cases} \quad (4.20)$$

dove abbiamo usato la convenzione $\theta_j = 0$ per $j > q$.

Applicando l'algoritmo delle innovazioni al processo $(W_t)_{t \in \mathbb{Z}}$, otteniamo

$$\begin{cases} \widehat{W}_{n+1} = \sum_{j=1}^n \theta_{nj} (W_{n+1-j} - \widehat{W}_{n+1-j}), & 1 \leq n < m, \\ \widehat{W}_{n+1} = \sum_{j=1}^q \theta_{nj} (W_{n+1-j} - \widehat{W}_{n+1-j}), & n \geq m, \end{cases} \quad (4.21)$$

dove i coefficienti θ_{nj} e l'errore quadratico medio $r_n = E[(W_{n+1} - \widehat{W}_{n+1})^2]$ sono calcolati ricorsivamente usando l'equazione (4.15) con la funzione k definita come in (4.20). Osserviamo che quando $n \geq m$ e $j > q$ il termine θ_{nj} non compare in (4.21). Questo è una conseguenza dell'equazione (4.15) e del fatto che $\gamma_W(n, j) = 0$ se $n > m$ e $|n - j| > q$.

Dalla proiezione del sistema (4.19) sullo spazio \mathcal{H}_{t-1} , otteniamo

$$\begin{cases} \widehat{W}_t = \sigma^{-1} \widehat{X}_t & t = 1, \dots, m, \\ \widehat{W}_t = \sigma^{-1} [\widehat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}], & t > m, \end{cases} \quad (4.22)$$

che, insieme con la (4.19), ci dà

$$X_t - \widehat{X}_t = \sigma(W_t - \widehat{W}_t) \quad \text{per ogni } t \geq 1.$$

Sostituendo $\sigma^{-1}(X_j - \widehat{X}_j)$ al posto di $(W_j - \widehat{W}_j)$ nei sistemi (4.21) e (4.22), otteniamo

$$\begin{cases} \widehat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}), & 1 \leq n < m, \\ \widehat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}) & n \geq m, \end{cases} \quad (4.23)$$

e

$$E[(X_{n+1} - \widehat{X}_{n+1})^2] = \sigma^2 E[(W_{n+1} - \widehat{W}_{n+1})^2] = \sigma^2 r_n,$$

dove θ_{nj} e r_n sono calcolati grazie alla (4.15) con la funzione di autocovarianza k definita nella (4.20).

Osservazione 4.3. La funzione di autocovarianza $\gamma_W(i, j)$ del processo trasformato $(W_t)_{t \in \mathbb{Z}}$ dipende solo da $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ e non da σ^2 . Questo vale anche per θ_{nj} e r_n .

Osservazione 4.4. La rappresentazione della previsione \widehat{X}_{n+1} del sistema (4.23) è particolarmente vantaggiosa da un punto di vista pratico, perché per $n \geq m$ richiede la memorizzazione di al più p osservazioni passate, X_n, \dots, X_{n+1-p} , e al più q innovazioni passate, $(X_{n+1-j} - \widehat{X}_{n+1-j})$, $j = 1, \dots, q$, per determinare \widehat{X}_{n+1} . Applicando direttamente l'algoritmo delle innovazioni al processo $(X_t)_{t \in \mathbb{Z}}$ avremmo ottenuto una rappresentazione di \widehat{X}_{n+1} in funzione di tutte le n precedenti innovazioni $(X_j - \widehat{X}_j)$, $j = 1, \dots, n$.

Esempio 18 (Previsione di un processo AR(p)). Applicando la (4.23) ad un processo ARMA(p,1) con $\theta_1 = 0$, troviamo che

$$\widehat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p}, \quad n \geq p.$$

Esempio 19 (Previsione di un processo MA(q)). Applicando la (4.23) ad un processo ARMA(1,q) con $\phi_1 = 0$, otteniamo

$$\widehat{X}_{n+1} = \sum_{j=1}^{\min(n,q)} \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}), \quad n \geq 1,$$

dove i coefficienti θ_{nj} sono calcolati utilizzando l'algoritmo delle innovazioni e la funzione di autocovarianza $\gamma_W(i, j)$ definita nella (4.20). Poiché il processo $(W_t)_{t \in \mathbb{Z}}$ e il processo $(\sigma^{-1} X_t)_{t \in \mathbb{Z}}$ sono identici, la funzione di autocovarianza risulta

$$\gamma_W(i, j) = \sigma^{-2} \gamma_X(i - j) = \sum_{r=0}^{q-|i-j|} \theta_r \theta_{r+|i-j|}.$$

Esempio 20 (Previsione di un processo ARMA(1,1)). Sia dato il processo ARMA(1,1)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

con $|\phi| < 1$, allora il sistema (4.23) si riduce ad un'unica equazione

$$\widehat{X}_{n+1} = \phi X_n + \theta_{n1} (X_n - \widehat{X}_n), \quad n \geq 1.$$

Per calcolare θ_{n1} , utilizziamo l'equazione vista nel paragrafo 2.4

$$\gamma_X(k) - \phi_1 \gamma_X(k-1) - \dots - \phi_p \gamma_X(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k}, \quad \text{per } 0 \leq k < 2,$$

e troviamo che $\gamma_X(0) = \frac{\sigma^2(1+2\theta\phi+\theta^2)}{(1-\phi^2)}$. Sostituendo questa formula nella (4.20), per $i, j \geq 1$, otteniamo

$$\gamma_W(i, j) = \begin{cases} (1 + 2\theta\phi + \theta^2)/(1 - \phi^2), & i = j = 1, \\ 1 + \theta^2, & i = j \geq 2, \\ \theta, & |i - j| = 1, i \geq 1, \\ 0, & \text{altrimenti.} \end{cases}$$

Con questi valori di $\gamma_W(i, j)$, utilizzando l'algoritmo delle innovazioni, troviamo

$$\begin{cases} r_0 = (1 + 2\theta\phi + \theta^2)/(1 - \phi^2), \\ \theta_{n1} = \theta/r_{n-1}, \\ r_n = 1 + \theta^2 - \theta^2/r_{n-1}. \end{cases}$$

Nella seguente tabella riportiamo i valori simulati di X_1, \dots, X_{10} del processo ARMA(1,1) con $(Z_t)_{t \in \mathbb{Z}} \sim WN(0, 1)$, $\phi_1 = \phi = 0.2$ e $\theta_1 = \theta = 0.4$. Riportiamo anche i valori di r_n e θ_{n1} , per $n = 1, \dots, 10$ e i corrispondenti valori \hat{X}_{n+1} , per $n = 1, \dots, 10$. Siccome $\sigma^2 = 1$, gli errori quadratici medi risultano essere $E[(X_{n+1} - \hat{X}_{n+1})^2] = \sigma^2 r_n = r_n$.

n	X_{n+1}	r_n	θ_{n1}	\hat{X}_{t+1}
0	-1.100	1.3750		0
1	0.514	1.0436	0.2909	-0.5340
2	0.116	1.0067	0.3833	0.5068
3	-0.845	1.0011	0.3973	-0.1321
4	0.872	1.0002	0.3996	-0.4539
5	-0.467	1.0000	0.3999	0.7046
6	-0.977	1.0000	0.4000	-0.5620
7	-1.699	1.0000	0.4000	0.3614
8	-1.228	1.0000	0.4000	0.8748
9	-1.093	1.0000	0.4000	0.3869
10		1.0000	0.4000	-0.5010

Esempio 21 (Previsione di un processo ARMA(2,3)). Sia dato il processo causale ARMA(2,3)

$$X_t - X_{t-1} + 0.24X_{t-2} = Z_t + 0.4Z_{t-1} + 0.2Z_{t-2} + 0.1Z_{t-3}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, 1).$$

I valori simulati di X_1, \dots, X_{10} sono riportati nella tabella alla fine dell'esempio.

Per calcolare i valori di \hat{X}_n , per $n = 2, \dots, 11$, abbiamo bisogno dei valori della funzione di autocovarianza $\gamma_X(h)$, per $h = 0, 1, 2$, che, calcolate come nell'esempio precedente, risultano essere

$$\gamma_X(0) = 7.17133, \quad \gamma_X(1) = 6.44139, \quad \text{e} \quad \gamma_X(2) = 5.06027.$$

Sostituendo questi valori nella (4.20), troviamo la matrice simmetrica $\Gamma_W = [\gamma_W(i, j)]_{i, j=1, 2, \dots}$:

$$\Gamma_W = \begin{pmatrix} 7.17133 & & & & & & & & & & \\ 6.44139 & 7.17133 & & & & & & & & & \\ 5.06027 & 6.44139 & 7.17133 & & & & & & & & \\ 0.10 & 0.34 & 0.816 & 1.21 & & & & & & & \\ 0 & 0.10 & 0.34 & 0.50 & 1.21 & & & & & & \\ 0 & 0 & 0.10 & 0.24 & 0.50 & 1.21 & & & & & \\ \cdot & 0 & 0 & 0.10 & 0.24 & 0.50 & 1.21 & & & & \\ \cdot & \cdot & 0 & 0 & \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \end{pmatrix}.$$

Adesso che abbiamo le funzioni di autocovarianza $\gamma_W(i, j)$, possiamo usare l'algoritmo delle innovazioni per calcolare θ_{nj} e r_{n-1} , con $j = 1, \dots, n$ e $n = 1, \dots, 10$:

$$\hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), \quad n = 1, 2,$$

$$\hat{X}_{n+1} = X_n - 0.24X_{n-1} + \sum_{j=1}^3 \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), \quad n = 3, 4, \dots$$

e

$$E[(X_{n+1} - \widehat{X}_{n+1})^2] = \sigma^2 r_n = r_n.$$

I valori trovati sono riportati nella tabella seguente.

n	X_{n+1}	r_n	θ_{n1}	θ_{n2}	θ_{n3}	\widehat{X}_{t+1}
0	1.704	7.1713				0
1	0.527	1.3856	0.8982			1.5305
2	1.041	1.0057	1.3685	0.7056		-0.1710
3	0.942	1.0019	0.4008	0.1806	0.0139	1.2428
4	0.555	1.0016	0.3998	0.2020	0.0722	0.7443
5	-1.002	1.0005	0.3992	0.1995	0.0994	0.3138
6	-0.585	1.0000	0.4000	0.1997	0.0998	-1.7293
7	0.010	1.0000	0.4000	0.2000	0.0998	-0.1688
8	-0.638	1.0000	0.4000	0.2000	0.0999	0.3193
9	0.525	1.0000	0.4000	0.2000	0.1000	-0.8731
10		1.0000	0.4000	0.2000	0.1000	1.0638
11		1.0000	0.4000	0.2000	0.1000	
12		1.0000	0.4000	0.2000	0.1000	

4.3.1 La previsione a più passi di un processo ARMA

Come nel paragrafo precedente, adatteremo la notazione P_n per indicare la proiezione $P_{\mathcal{H}_n}$. Dall'equazione (4.17) abbiamo dunque

$$\begin{aligned} P_n W_{n+h} &= \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (W_{n+h-j} - \widehat{W}_{n+h-j}) \\ &= \sigma^{-1} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}). \end{aligned}$$

Usando questa relazione e applicando l'operatore P_n ad entrambi i membri dell'equazione (4.19), possiamo concludere che i predittori ad h passi $P_n X_{n+h}$

soddisfano

$$P_n X_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), & 1 \leq h \leq m-n, \\ \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{h \leq j \leq q} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), & h > m-n. \end{cases} \quad (4.24)$$

Dopo aver calcolato i predittori $\widehat{X}_1, \dots, \widehat{X}_n$, utilizzando la (4.24), per un n fissato, possiamo quindi determinare i predittori $P_n X_{n+1}, P_n X_{n+2}, P_n X_{n+3}, \dots$, ricorsivamente.

Assumendo $n > m$ otteniamo, per $h \geq 1$,

$$P_n X_{n+h} = \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), \quad (4.25)$$

dove, per $h > q$, la seconda somma è zero. Usando la relazione

$$X_{n+h} = \widehat{X}_{n+h} + (X_{n+h} - \widehat{X}_{n+h}),$$

possiamo riscrivere l'equazione precedente come

$$X_{n+h} = \sum_{i=1}^p \phi_i X_{n+h-i} + \sum_{j=0}^q \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), \quad (4.26)$$

dove $\theta_{n0} := 1$ per ogni n . Sottraendo la (4.25) dalla (4.26) otteniamo

$$\begin{aligned} X_{n+h} - P_n X_{n+h} - \sum_{i=1}^p \phi_i (X_{n+h-i} - P_n X_{n+h-i}) \\ = \sum_{j=0}^{h-1} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), \end{aligned}$$

e dunque

$$\Phi \begin{bmatrix} X_{n+1} - P_n X_{n+1} \\ \vdots \\ X_{n+h} - P_n X_{n+h} \end{bmatrix} = \Theta \begin{bmatrix} X_{n+1} - \widehat{X}_{n+1} \\ \vdots \\ X_{n+h} - \widehat{X}_{n+h} \end{bmatrix}, \quad (4.27)$$

dove Φ e Θ sono le matrici triangolari inferiori

$$\Phi = -[\phi_{i-j}]_{i,j=1}^h, \text{ con } \phi_0 := -1, \phi_j := 0, \text{ se } j > p \text{ o } j < 0,$$

e

$$\Theta = [\theta_{n+i-1, i-j}]_{i,j=1}^h, \text{ con } \theta_{n0} := 1, \theta_{nj} := 0, \text{ se } j > q \text{ o } j < 0.$$

Dall'equazione (4.27) ricaviamo che la matrice di covarianza del vettore $(X_{n+1} - P_n X_{n+1}, \dots, X_{n+h} - P_n X_{n+h})'$ è

$$C = \Phi^{-1} \Theta V \Theta' \Phi'^{-1},$$

dove $V = \text{diag}(v_n, v_{n+1}, \dots, v_{n+h-1})$ e $\Phi^{-1} = [\chi_{i-j}]_{i,j=1}^h$, con $\chi_0 := 1, \chi_j := 0$ se $j < 0$. Questa relazione ci permette di calcolare l'errore quadratico medio del predittore ad h passi:

$$\sigma_n^2(h) := E[(X_{n+h} - P_n X_{n+h})^2] = \sum_{j=0}^{h-1} \left(\sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 v_{n+h-j-1}, \quad (4.28)$$

dove $\chi_r = \sum_{k=1}^{\min(p,r)} \phi_k \chi_{r-k}$, per $r \geq 1$.

Se assumiamo che il processo ARMA sia invertibile, possiamo passare al limite per $n \rightarrow \infty$ nella (4.25) ed ottenere le seguenti approssimazioni

$$P_n X_{n+h} \cong \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_j (X_{n+h-j} - \widehat{X}_{n+h-j})$$

e

$$\sigma^2(h) \cong \sigma^2 \sum_{j=0}^{h-1} \left(\sum_{r=0}^j \chi_r \theta_{j-r} \right)^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2,$$

dove

$$\sum_{j=0}^{h-1} \psi_j z^j = \left(\sum_{j=0}^{\infty} \chi_j z^j \right) \left(\sum_{j=0}^q \theta_j z^j \right) = \theta(z)/\phi(z), \quad |z| \leq 1.$$

Esempio 22 (Previsione a due e tre passi di un processo ARMA(2,3)). Possiamo applicare la (4.25) e la (4.28) all'esempio 21 per calcolare la previsione a due e tre passi di un processo ARMA(2,3). Dall'equazione (4.25), otteniamo

$$P_{10} X_{12} = \sum_{i=1}^2 \phi_i P_{10} X_{12-i} + \sum_{j=2}^3 \theta_{11,j} (X_{12-j} - \widehat{X}_{12-j})$$

$$= \phi_1 \widehat{X}_{11} + \phi_2 \widehat{X}_{10} + 0.2(X_{10} - \widehat{X}_{10}) + 0.1(X_9 - \widehat{X}_9) = 1.1217$$

e

$$\begin{aligned} P_{10}X_{13} &= \sum_{i=1}^2 \phi_i P_{10}X_{13-i} + \sum_{j=3}^3 \theta_{12,j}(X_{13-j} - \widehat{X}_{13-j}) \\ &= \phi_1 P_{10}X_{12} + \phi_2 \widehat{X}_{11} + 0.1(X_{10} - \widehat{X}_{10}) = 1.0062. \end{aligned}$$

Per $k > 13$, possiamo calcolare ricorsivamente $P_{10}X_k$ usando la relazione

$$P_{10}X_k = \phi_1 P_{10}X_{k-1} + \phi_2 P_{10}X_{k-2}.$$

Per trovare l'errore quadratico medio di $P_n X_{n+h}$ applichiamo la (4.28) con $\chi_0 = 1, \chi_1 = \phi_1 = 1$ e $\chi_2 = \phi_1 \chi_1 + \phi_2 = 0.76$. Usando i valori di θ_{nj} e $v_j = r_j$, contenuti nella tabella dell'esempio 21, otteniamo

$$\sigma_{10}^2(2) = E[(X_{12} - P_{10}X_{12})^2] = 2.960,$$

e

$$\sigma_{10}^2(3) = E[(X_{13} - P_{10}X_{13})^2] = 4.810.$$

Osservazione 4.5. Per un processo stazionario gaussiano, l'errore

$$\Delta_n(h) = X_{n+h} - P_n X_{n+h}$$

ha legge normale di media nulla e varianza

$$\sigma_n^2(h) = E[\Delta_n(h)^2],$$

la quale può essere calcolata usando la (4.18) nel caso generale oppure la (4.28) nel caso di un processo ARMA. Se quindi $q_{1-\frac{\alpha}{2}}$ è il quantile di ordine $(1 - \frac{\alpha}{2})$ della legge normale standard, allora X_{n+h} è compreso tra i valori $P_n X_{n+h} \pm q_{1-\frac{\alpha}{2}} \sigma_n(h)$ con probabilità $(1 - \alpha)$. Questi valori vengono perciò detti “ $(1 - \alpha)$ -prediction bounds” per X_{n+h} .

4.3.2 La previsione di un processo ARMA causale invertibile

Per approssimare $P_n X_{n+h}$ quando n è grande, risulta utile determinare la proiezione di X_{n+h} sul sottospazio $\mathcal{M}_n = \overline{\text{sp}}\{X_j, -\infty < j \leq n\}$. Consideriamo questo problema nel caso di un processo ARMA(p,q) causale invertibile:

$$\phi(B)X_t = \theta(B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2). \quad (4.29)$$

Fissiamo un intero positivo n e definiamo

$$\tilde{X}_t := P_{\mathcal{M}_n} X_t. \quad (4.30)$$

Notiamo che $\tilde{X}_t = X_t$, per $t \leq n$. Vale il seguente risultato.

Proposizione 4.3.1. *Per ogni n e h , risulta*

$$\tilde{X}_{n+h} = - \sum_{j=1}^{\infty} \pi_j \tilde{X}_{n+h-j} \quad (4.31)$$

e

$$\tilde{X}_{n+h} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j}, \quad (4.32)$$

dove $\sum_{j=0}^{\infty} \pi_j z^j = \phi(z)/\theta(z)$ e $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$, $|z| \leq 1$. Inoltre vale

$$E[(X_{n+h} - \tilde{X}_{n+h})^2] = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \quad (4.33)$$

Dimostrazione. Grazie ai Teoremi 2.3.3 e 2.3.4, sappiamo che

$$Z_{n+h} = X_{n+h} + \sum_{j=1}^{\infty} \pi_j X_{n+h}$$

e

$$X_{n+h} = \sum_{j=0}^{\infty} \psi_j Z_{n+h-j}. \quad (4.34)$$

Applicando l'operatore $P_{\mathcal{M}_n}$ ad entrambi i membri delle precedenti eguaglianze ed usando il fatto che Z_{n+k} è ortogonale a \mathcal{M}_n , per ogni $k \geq 1$, otteniamo le

equazioni (4.31) e (4.32). Sottraendo la (4.32) dall'equazione (4.34) troviamo che

$$X_{n+h} - \tilde{X}_{n+h} = \sum_{j=0}^{h-1} \psi_j Z_{n+h-j},$$

da cui segue la (4.33). \square

Osservazione 4.6. L'equazione (4.31) è utile per calcolare ricorsivamente \tilde{X}_{n+h} , usando la condizione $\tilde{X}_t = X_t$, per $t \leq n$:

$$\begin{aligned} \tilde{X}_{n+1} &= - \sum_{j=1}^{\infty} \pi_j X_{n+1-j}, \\ \tilde{X}_{n+2} &= -\pi_1 \tilde{X}_{n+1} - \sum_{j=2}^{\infty} \pi_j X_{n+2-j}, \end{aligned}$$

e così via.

Quando n è grande, un'approssimazione di $P_n X_{n+h}$ può essere ottenuta con la soluzione troncata, la quale è ottenuta ponendo $\sum_{j=n+h}^{\infty} \pi_j X_{n+h-j} = 0$ nell'equazione (4.31) e risolvendo

$$\tilde{X}_{n+h}^T = - \sum_{j=1}^{n+h-1} \pi_j \tilde{X}_{n+h-j}^T \quad \text{con } \tilde{X}_t^T = X_t, \text{ per } t \leq n.$$

Con questo procedimento otteniamo

$$\begin{aligned} \tilde{X}_{n+1}^T &= - \sum_{j=1}^n \pi_j X_{n+1-j}, \\ \tilde{X}_{n+2}^T &= -\pi_1 \tilde{X}_{n+1}^T - \sum_{j=2}^{n+1} \pi_j X_{n+2-j}, \\ &\dots \end{aligned}$$

La quantità $E[(X_{n+h} - P_n X_{n+h})^2]$, quando n è grande, può essere approssimata dall'errore quadratico medio di \tilde{X}_{n+h} , cioè da $E[(X_{n+h} - \tilde{X}_{n+h})^2]$.

Osservazione 4.7. Per un processo $AR(p)$, l'equazione (4.31) diventa

$$\tilde{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p},$$

mentre l'errore quadratico medio risulta

$$E[(X_{n+1} - \tilde{X}_{n+1})^2] = \sigma^2.$$

Per un processo $MA(1)$, dall'equazione (4.31) otteniamo

$$\tilde{X}_{n+1} = - \sum_{j=0}^{\infty} (-\theta_1)^{j+1} X_{n-j},$$

con un errore quadratico medio pari a

$$E[(X_{n+1} - \tilde{X}_{n+1})^2] = \sigma^2.$$

L'approssimazione troncata di $P_n X_{n+1}$ per un processo $MA(1)$ è

$$\tilde{X}_{n+1}^T = - \sum_{j=0}^{n-1} (-\theta_1)^{j+1} X_{n-j},$$

che può non essere una buona approssimazione se $|\theta_1|$ è vicino a 1.

Osservazione 4.8. Per un fissato n , le differenze $X_{n+h} - \tilde{X}_{n+h}$, per $h = 1, 2, \dots$, non sono correlate. Infatti abbiamo

$$E[(X_{n+h} - \tilde{X}_{n+h})(X_{n+k} - \tilde{X}_{n+k})] = \sigma^2 \sum_{j=0}^{h-1} \psi_j \psi_{j+k-h} \quad \text{per } k \geq h.$$

4.4 La previsione nel dominio delle frequenze

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero, con funzione di ripartizione spettrale F e con un processo ad incrementi ortogonali associato $(Z(\lambda))_{-\pi \leq \lambda \leq \pi}$, allora l'applicazione I (vedi Appendice B ed osservazione 3.3 del cap. 3), definita da

$$I(g) = \int_{(-\pi, \pi]} g(\lambda) dZ(\lambda), \quad \text{con } g \in L^2(F), \quad (4.35)$$

è un isomorfismo di $L^2(F)$ su $\overline{\mathcal{H}} = \overline{\text{sp}}\{X_t, t \in \mathbb{Z}\}$, con la proprietà che

$$I(e^{it \cdot}) = X_t, \quad t \in \mathbb{Z}. \quad (4.36)$$

Questo isomorfismo ci permette di calcolare le proiezioni sul sottospazio $\overline{\mathcal{H}}$ (e quindi i predittori), calcolando prima le proiezioni in $L^2(F)$ e poi applicando I . Per esempio il predittore $P_{\mathcal{M}_n} X_{n+h}$ di X_{n+h} in $\mathcal{M}_n = \overline{\text{sp}}\{X_t, -\infty < t \leq \infty\}$ può essere ottenuto come

$$P_{\mathcal{M}_n} X_{n+h} = I(P_{\overline{\text{sp}}\{\exp(it\cdot), -\infty < t \leq n\}} e^{i(n+h)\cdot}). \quad (4.37)$$

Nell'esempio seguente illustriamo come calcolare il membro destro di questa relazione.

Esempio 23. Sia $(X_t)_{t \in \mathbb{Z}}$ un processo stazionario con media zero e densità spettrale f , tale che

$$f(\lambda) = 0, \quad \text{per } 1 < |\lambda| \leq \pi. \quad (4.38)$$

Per $\lambda \in E := [-1, 1]$, vale $|1 - e^{-i\lambda}| < 1$ e dunque la serie

$$\sum_{k=0}^{\infty} (1 - e^{-i\lambda})^k = \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} (-1)^j e^{-i\lambda j}$$

converge uniformemente su E verso $[1 - (1 - e^{-i\lambda})]^{-1} = e^{i\lambda}$. Ne segue

$$e^{i(n+1)\lambda} = \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} (-1)^j e^{i\lambda(n-j)}, \quad \text{per } \lambda \in E, \quad (4.39)$$

dove le serie a destra convergono uniformemente su E . Grazie alla (4.38), queste due serie risultano convergenti anche in $L^2(F)$ verso un elemento di $\overline{\text{sp}}\{\exp(it\cdot), -\infty < t \leq n\}$. Proiettando ogni membro della (4.39) su questo sottospazio, otteniamo

$$P_{\overline{\text{sp}}\{\exp(it\cdot), -\infty < t \leq n\}} e^{i(n+1)\cdot} = e^{i(n+1)\cdot} = \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} (-1)^j e^{i(n-j)\cdot}.$$

Applicando I alla precedente eguaglianza, ed utilizzando la (4.36) e la (4.37), possiamo concludere che

$$P_{\mathcal{M}_n} X_{n+1} = X_{n+1} = \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} (-1)^j X_{n-j}.$$

Esempio 24 (Previsione di un processo ARMA). Consideriamo un processo ARMA causale invertibile:

$$\phi(B)X_t = \theta(B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

con densità spettrale

$$f(\lambda) = a(\lambda)\overline{a(\lambda)},$$

dove

$$a(\lambda) = (2\pi)^{-1/2}\sigma \sum_{k=0}^{\infty} \psi_k e^{-ik\lambda} \quad (4.40)$$

e $\sum_{k=0}^{\infty} \psi_k z^k = \theta(z)/\phi(z)$, $|z| \leq 1$. Dal fatto che il processo è causale segue che $\sum_k |\psi_k| < \infty$ e quindi che la serie in (4.40) converge uniformemente su $[-\pi, \pi]$. La funzione $g(\cdot) = P_{\overline{\text{sp}}\{\exp(it\cdot), -\infty < t \leq n\}} e^{i(n+h)\cdot}$ deve soddisfare

$$\int_{-\pi}^{\pi} (e^{i(n+h)\lambda} - g(\lambda)) e^{-im\lambda} a(\lambda) \overline{a(\lambda)} d\lambda = 0, \quad \text{per } m \leq n.$$

Da questa equazione otteniamo che $(e^{i(n+h)\cdot} - g(\cdot))a(\cdot)\overline{a(\cdot)}$ è un elemento del sottospazio $\mathcal{M}_+ = \overline{\text{sp}}\{\exp(im\cdot), m > n\}$ di $L^2([-\pi, \pi])$. Dalla (4.40) abbiamo che $1/\overline{a(\cdot)} \in \overline{\text{sp}}\{\exp(im\cdot), m \geq 0\}$ e quindi anche $(e^{i(n+h)\cdot} - g(\cdot))a(\cdot)$ risulterà essere un elemento di \mathcal{M}_+ . Osservando che $g(\cdot)a(\cdot)$ è ortogonale in $L^2(d\lambda)$ a \mathcal{M}_+ , possiamo scrivere

$$e^{i(n+h)\lambda} a(\lambda) = g(\lambda)a(\lambda) + (e^{i(n+h)\lambda} - g(\lambda))a(\lambda).$$

D'altra parte, dalla (4.40), ricaviamo

$$e^{i(n+h)\lambda} a(\lambda) = (2\pi)^{-1/2}\sigma e^{in\lambda} \sum_{k=-h}^{\infty} \psi_{k+h} e^{-ik\lambda},$$

e, poiché l'elemento $e^{i(n+h)\cdot} a(\cdot)$ di $L^2(d\lambda)$ ha un'unica rappresentazione come somma di due componenti, una in \mathcal{M}_+ e una ortogonale a \mathcal{M}_+ , deduciamo

$$g(\lambda)a(\lambda) = (2\pi)^{-1/2}\sigma e^{in\lambda} \sum_{k=0}^{\infty} \psi_{k+h} e^{-ik\lambda}.$$

Usando nuovamente la (4.40), otteniamo

$$g(\lambda) = e^{in\lambda} [\phi(e^{-i\lambda})/\theta(e^{-i\lambda})] \sum_{k=0}^{\infty} \psi_{k+h} e^{-ik\lambda},$$

cioè

$$g(\lambda) = \sum_{j=0}^{\infty} \alpha_j e^{i(n-j)\lambda}, \quad (4.41)$$

dove $\sum_{j=0}^{\infty} \alpha_j z^j = [\phi(z)/\theta(z)] \sum_{k=0}^{\infty} \psi_{k+h} z^k$, per $|z| \leq 1$. Applicando I ad ogni membro della (4.41) e usando (4.36) e (4.37), possiamo concludere che

$$P_{\mathcal{M}_n} X_{n+h} = \sum_{j=0}^{\infty} \alpha_j X_{n-j}.$$

Capitolo 5

La costruzione di un modello ARMA

Per costruire un appropriato processo stazionario $X = (X_t)_{t \in \mathbb{Z}}$ come modello per i dati, è di notevole importanza riuscire a dare una stima, a partire dalle osservazioni, del valor medio $E[X_t] = \mu$, della funzione di autocovarianza $\gamma_X(\cdot)$ e della funzione di autocorrelazione $\rho_X(\cdot) = \gamma_X(\cdot)/\gamma_X(0)$.

Nei prossimi due paragrafi di questo capitolo illustreremo degli stimatori di queste grandezze ed alcune loro proprietà. Nel terzo paragrafo affronteremo invece il problema di determinare un appropriato modello ARMA(p,q) a partire dai dati osservati. Ciò consiste nell'individuare in modo opportuno gli ordini p e q e nella stima dei restanti parametri per un fissato valore di p e q .

Nel seguito γ_X e ρ_X saranno denotate semplicemente con i simboli γ e ρ .

5.1 La stima della media

Un naturale stimatore della media μ di un processo stazionario $(X_t)_{t \in \mathbb{Z}}$ è la media campionaria

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

Questo è uno stimatore non distorto: infatti, abbiamo

$$E[\bar{X}_n] = \frac{1}{n} \sum_{t=1}^n E[X_t] = \frac{1}{n} n\mu = \mu.$$

La sua varianza inoltre è

$$\begin{aligned} Var(\bar{X}_n) &= \frac{1}{n^2} Var\left(\sum_{t=1}^n X_t\right) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n Cov(X_j, X_k) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \gamma(j-k) \\ &= \frac{1}{n^2} \sum_{h=-(n-1)}^{n-1} (n-|h|)\gamma(h) = \frac{1}{n} \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h), \end{aligned}$$

dove abbiamo posto $h = (j - k)$ ed usato le proprietà della funzione di autocovarianza, in particolare che è una funzione pari. Ne segue che

$$nVar(\bar{X}_n) = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h) \leq \sum_{|h|<n} |\gamma(h)|$$

e quindi, se $\gamma(n) \rightarrow 0$ per $n \rightarrow \infty$, allora $Var(\bar{X}_n) \rightarrow 0$ in quanto $\lim_{n \rightarrow \infty} n^{-1} \sum_{|h|<n} |\gamma(h)| = 2 \lim_{n \rightarrow \infty} |\gamma(h)| = 0$. In altre parole, se $\gamma(n) \rightarrow 0$ per $n \rightarrow \infty$, il processo stazionario $(X_t)_{t \in \mathbb{Z}}$ è ergodico in media e quindi lo stimatore \bar{X}_n risulta consistente. Se $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, grazie al Teorema della convergenza dominata, ricaviamo

$$\lim_{n \rightarrow \infty} nVar(\bar{X}_n) = \lim_{n \rightarrow \infty} \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \sum_{h=-\infty}^{\infty} \gamma(h).$$

La condizione $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ risulta essere soddisfatta da tutti i processi ARMA(p,q). Questa condizione inoltre implica che il processo $(X_t)_{t \in \mathbb{Z}}$ abbia densità spettrale f e quindi, grazie al Corollario 3.3.4, segue che

$$nVar(\bar{X}_n) \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0).$$

Se poi $(X_t)_{t \in \mathbb{Z}}$ è della forma

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},$$

con $(Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2)$ e $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, allora vale $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ e abbiamo

$$n \text{Var}(\bar{X}_n) \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2.$$

Le considerazioni appena fatte suggeriscono che, aggiungendo opportune condizioni, il processo $(X_t)_{t \in \mathbb{Z}}$ potrebbe essere asintoticamente normale, ossia che, facendo tendere all'infinito il numero n delle osservazioni, $(\bar{X}_n - \mu)$ potrebbe convergere in legge verso la legge normale $N(0, n^{-1} \sum_{h=-\infty}^{\infty} \gamma(h))$. Si dimostra infatti (si veda Teorema 7.1.2 di [7]) che, se $(X_t)_{t \in \mathbb{Z}}$ è un processo stazionario della forma

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (Z_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2),$$

tale che $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ e $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, allora

$$\bar{X}_n \text{ è } AN(\mu, n^{-1}v),$$

ossia $n^{1/2}(\bar{X}_n - \mu) \xrightarrow{d} N(0, v)$, dove $v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 (\sum_{j=-\infty}^{\infty} \psi_j)^2$ e γ è la funzione di autocovarianza di $(X_t)_{t \in \mathbb{Z}}$. Questo risultato è utilizzato per trovare intervalli di confidenza asintotici per μ . Se il processo $(X_t)_{t \in \mathbb{Z}}$ è anche gaussiano, allora abbiamo l'esatta distribuzione di \bar{X}_n , per ogni n finito, ovvero

$$n^{1/2}(\bar{X}_n - \mu) \sim N \left(0, \sum_{|h| < n} \left(1 - \frac{|h|}{n} \right) \gamma(h) \right).$$

Questa relazione fornisce degli intervalli di confidenza per μ esatti se γ è nota; altrimenti degli intervalli di confidenza asintotici nel caso sia necessario stimare γ a partire dalle osservazioni.

L'ipotesi di identità in distribuzione può essere sostituita da condizioni che permettono l'uso di teoremi limite per variabili aleatorie non identicamente distribuite (ad esempio, la limitatezza dei momenti terzi).

5.2 La stima della funzione di autocovarianza e di autocorrelazione

Consideriamo gli stimatori della funzione di autocovarianza e della funzione di autocorrelazione seguenti:

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \bar{X}_n)(X_{i+h} - \bar{X}_n), \quad 0 \leq h \leq n-1, \quad (5.1)$$

e

$$\hat{\rho}_n(h) = \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)}, \quad 0 \leq h \leq n-1. \quad (5.2)$$

Notiamo che nella formula (5.1) i prodotti all'interno della serie sono $n-h$. Se prendessimo valori di i maggiori di $n-h$, nella serie comparirebbero anche le osservazioni successive ad X_n , che non conosciamo. Osserviamo inoltre che $\hat{\gamma}_n$ non è un stimatore non distorto. Ha però la proprietà che la matrice

$$\hat{\Gamma}_n = \begin{pmatrix} \hat{\gamma}_n(0) & \hat{\gamma}_n(1) & \cdots & \hat{\gamma}_n(n-1) \\ \hat{\gamma}_n(1) & \hat{\gamma}_n(0) & \cdots & \hat{\gamma}_n(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_n(n-1) & \hat{\gamma}_n(n-2) & \cdots & \hat{\gamma}_n(0) \end{pmatrix}$$

è semidefinita positiva. Per verificare questo fatto, possiamo scrivere

$$\hat{\Gamma}_n = \frac{1}{n} T_n T_n',$$

dove T_n è la matrice $n \times 2n$

$$T_n = \begin{pmatrix} 0 & \cdots & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n \\ 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n & 0 \\ \vdots & & & \vdots & \vdots & & & \vdots \\ 0 & Y_1 & Y_2 & \cdots & Y_n & 0 & \cdots & 0 \end{pmatrix}$$

con $Y_i = X_i - \bar{X}_n$, per $i = 1, \dots, n$. Ne discende che, per un qualsiasi vettore reale \mathbf{a} , vale

$$\mathbf{a}' \hat{\Gamma}_n \mathbf{a} = \frac{1}{n} (\mathbf{a}' T_n) (\mathbf{a}' T_n)' \geq 0,$$

5.2 La stima della funzione di autocovarianza e di autocorrelazione 407

e, conseguentemente, la matrice $\widehat{\Gamma}_n$ e la matrice $\widehat{R}_n = \widehat{\Gamma}_n / \widehat{\gamma}_n(0)$ sono entrambe semidefinite positive. Queste matrici non risultano semidefinite positive se in (5.1) e (5.2) sostituiamo n^{-1} con $(n-h)^{-1}$. Si dimostra inoltre che $\det \widehat{\Gamma}_n > 0$ se $\widehat{\gamma}_n(0) > 0$.

Sotto opportune ipotesi, è possibile ricavare (si veda Teorema 7.2.1 di [7]) la distribuzione congiunta asintotica di $[\widehat{\rho}_n(1), \dots, \widehat{\rho}_n(h)]$ per $n \rightarrow \infty$: se $(X_t)_{t \in \mathbb{Z}}$ è un processo stazionario della forma

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (Z_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2),$$

tale che $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ e $E[Z_t^4] < \infty$ per ogni t , allora, per ogni $h \geq 1$,

$$\widehat{\boldsymbol{\rho}}_n(h) \text{ è } AN(\boldsymbol{\rho}(h), n^{-1}W),$$

dove

$$\widehat{\boldsymbol{\rho}}_n(h)' = [\widehat{\rho}_n(1), \widehat{\rho}_n(2), \dots, \widehat{\rho}_n(h)], \quad \boldsymbol{\rho}(h)' = [\rho(1), \rho(2), \dots, \rho(h)]$$

e W è la matrice di covarianza, il cui elemento di posto (i, j) è dato dalla formula di Barlett

$$w_{ij} = \sum_{k=-\infty}^{\infty} \{ \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \}.$$

L'ipotesi di identità in distribuzione può essere sostituita da condizioni che permettono l'uso di teoremi limite centrale per variabili aleatorie non identicamente distribuite (ad esempio, la limitatezza dei momenti di ordine 6). È possibile inoltre (si veda Teorema 7.2.2 di [7]) levare la condizione sui momenti quarti, aggiungendo però una condizione più forte sulla successione $(\psi_j)_j$:

$$\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty.$$

Osservazione 5.1. Con dei semplici calcoli, la formula di Barlett può essere riscritta nella forma

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}, \quad (5.3)$$

che risulta più conveniente dal punto di vista computazionale.

Le condizioni $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ e $\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty$ sono soddisfatte da ogni processo ARMA(p,q) associato ad un processo $(Z_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2)$.

I risultati precedenti permettono di determinare degli intervalli di confidenza asintotici per l'autocorrelazione ρ .

Esempio 25. Diamo un esempio di calcolo numerico degli stimatori della funzione di autocovarianza e di autocorrelazione. Consideriamo un insieme di dati generati dal lancio di una moneta e descritti dal processo:

$$X_t = Z_t - 0.7Z_{t-1} + 5,$$

dove $Z_t = 1$ quando otteniamo testa (T) e $Z_t = -1$ quando otteniamo croce (C). Riportiamo i valori ottenuti nella seguente tabella.

t	Evento	Z_t	X_t	$X_t - \bar{X}_{10}$	t	Evento	Z_t	X_t	$X_t - \bar{X}_{10}$
1	T	1	6.7	1.56	6	C	-1	4.7	-0.44
2	T	1	5.3	0.16	7	C	-1	4.7	-0.44
3	C	-1	3.3	-1.84	8	T	1	6.7	1.56
4	T	1	6.7	1.56	9	C	-1	3.3	-1.84
5	C	-1	3.3	-1.84	10	T	1	6.7	1.56

Tabella 5.1: Realizzazioni del processo con $Z_0 = -1$ e $n = 10$.

La stima della media è

$$\bar{X}_{10} = \frac{1}{10} \sum_{t=1}^{10} X_t = \frac{1}{10} (6.7 + 5.3 + \cdots + 6.7) = 5.14.$$

5.2 La stima della funzione di autocovarianza e di autocorrelazione 109

Possiamo calcolare le stime della funzione di autocovarianza e di autocorrelazione utilizzando (5.1) e (5.2). Otteniamo

$$\begin{aligned}\widehat{\gamma}_{10}(0) &= \frac{1}{10} \sum_{t=1}^{10} (X_t - \bar{X}_{10})^2 \\ &= \frac{1}{10} [(1.56)^2 + (0.16)^2 + \dots + (1.56)^2] = 2.0304\end{aligned}$$

e, ad esempio per $h = 4$:

$$\begin{aligned}\widehat{\gamma}_{10}(4) &= \frac{1}{10} \sum_{t=1}^6 (X_{t+4} - \bar{X}_{10})(X_t - \bar{X}_{10}) \\ &= \frac{1}{10} [(-1.84)(1.56) + (-0.44)(0.16) + (-0.44)(-1.84) \\ &\quad + (1.56)(1.56) + (-1.84)(-1.84) + (1.56)(-0.44)] = 0.3002.\end{aligned}$$

La stima della funzione di autocorrelazione è quindi

$$\widehat{\rho}_{10}(4) = \frac{\widehat{\gamma}_{10}(4)}{\widehat{\gamma}_{10}(0)} = \frac{0.3002}{2.0304} = 0.1478.$$

Esempio 26 (Rumore bianco indipendente). Se $(X_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2)$, allora $\rho(l) = 0$ per $|l| > 0$ e quindi, dalla (5.3), ricaviamo

$$w_{ij} = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{altrimenti.} \end{cases}$$

Perciò, quando n è grande, le variabili aleatorie $\widehat{\rho}_n(1), \dots, \widehat{\rho}_n(h)$ sono approssimativamente indipendenti ed identicamente distribuite con distribuzione normale di media zero e varianza n^{-1} . Più precisamente, per ogni $h \geq 1$, abbiamo

$$\frac{\widehat{\rho}_n(h) - \rho(h)}{(\sqrt{n})^{-1} \sqrt{w_{h,h}}} = \widehat{\rho}_n(h) \sqrt{n} \stackrel{d}{\sim} N(0, 1) \quad \text{per } n \rightarrow \infty.$$

Ne segue che, per $\alpha = 0.05$, otteniamo

$$P \{ |\widehat{\rho}_n(h)| \sqrt{n} > q_{1-\frac{\alpha}{2}} \} \cong \alpha \quad \text{per } n \rightarrow \infty,$$

dove $q_{1-\frac{\alpha}{2}} \cong 1.96$ è il quantile di ordine $(1 - \frac{\alpha}{2}) = 0.975$ della legge $N(0, 1)$. Otteniamo quindi l'intervallo

$$-n^{-\frac{1}{2}}q_{1-\frac{\alpha}{2}} < \widehat{\rho}_n(h) < n^{-\frac{1}{2}}q_{1-\frac{\alpha}{2}}.$$

Ciò significa che, se abbiamo dei dati per i quali $\widehat{\rho}_n(h)$ cade in questo intervallo per un buon numero di valori h (circa il 95% dei valori h considerati), allora non può essere respinta l'ipotesi che provengano da un rumore bianco indipendente.

D'altra parte, se i dati non sono generati da un rumore bianco, allora le stime $\widehat{\rho}_n(h)$ cadranno con maggiore frequenza all'esterno del precedente intervallo.

Esempio 27 (Processo a media mobile di ordine q). Se

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (Z_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2),$$

dalla formula di Bartlett (5.3) otteniamo

$$w_{ii} = [1 + 2\rho^2(1) + 2\rho^2(2) + \dots + 2\rho^2(q)] \quad \text{per } i > q,$$

Nel grafico in Figura 5.1 abbiamo riportato i valori della funzione di autocorrelazione campionaria $\widehat{\rho}_n(h)$, con $h = 0, 1, \dots, 40$, per $n = 200$ osservazioni del processo Gaussiano $MA(1)$ della forma

$$X_t = Z_t - 0.8Z_{t-1}, \quad (Z_t)_{t \in \mathbb{Z}} \sim IID(0, 1). \quad (5.4)$$

Osseviamo che $\gamma(0) = 1 + (0.8)^2$, $\gamma(1) = -0.8$ e $\gamma(h) = 0$ per ogni $h \geq 2$, da cui ricaviamo $\rho(h) = 0$ per ogni $h \geq 2$. Nel grafico abbiamo disegnato le rette $\pm 1.96n^{-1/2}[1 + 2\rho^2(1)]^{1/2}$ dove $\rho(1) = -0.8/1.64 = -0.4878$. Notiamo che la maggior parte delle autocorrelazioni campionarie $\widehat{\rho}_n(h)$ per $h \geq 2$ sono comprese tra queste rette. Questo fatto indica la compatibilità dei dati con il modello (5.4). Nella pratica però $\rho(1)$ non è conosciuto in anticipo. Di conseguenza, le autocorrelazioni $\widehat{\rho}_n(2), \dots, \widehat{\rho}_n(40)$ devono essere confrontate con i valori $\pm 1.96n^{-1/2}$ o con i valori $\pm 1.96n^{-1/2}[1 + 2\widehat{\rho}_n^2(1)]^{1/2}$, al fine di verificare l'ipotesi che i dati siano generati da un processo a media mobile di ordine 1.

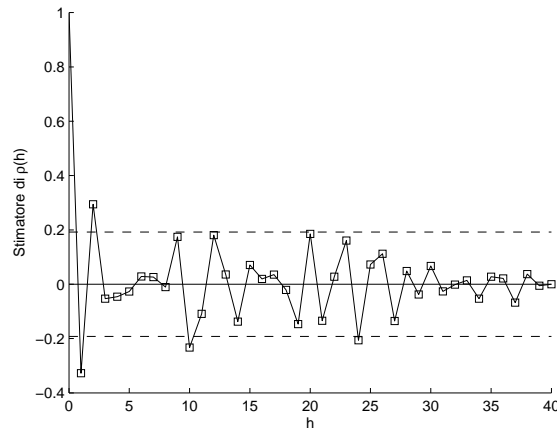


Figura 5.1: Grafico della funzione di autocorrelazione campionaria relativa ad un campione di 200 osservazioni di un processo MA(1) della forma (5.4).

5.3 Alcuni metodi di stima dei parametri di un processo ARMA

Il problema di determinare un appropriato modello ARMA(p, q) per una serie storica osservata consiste nella scelta degli ordini p e q e nella stima dei restanti parametri: i coefficienti $\{\phi_i, \theta_j : i = 1, \dots, p, j = 1, \dots, q\}$ e la varianza σ^2 del rumore bianco. In questo paragrafo illustreremo la procedura di stima, per p e q fissati, dei parametri $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ e σ^2 . Daremo anche delle indicazioni per ricavare delle stime di p e q .

In tutto il seguito assumiamo che ai dati è stata sottratta la media, in modo che il nostro problema diventa determinare un modello ARMA a media nulla. Se quindi un appropriato modello risulta essere

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2),$$

allora il corrispondente modello per la serie stazionaria di partenza $(Y_t)_{t \in \mathbb{Z}}$ è ottenuto sostituendo X_j con $Y_j - \bar{y}_n$ per $j = t, \dots, t-p$, dove $\bar{y}_n = n^{-1} \sum_{j=1}^n y_j$ è la media campionaria dei dati di partenza.

Nel caso in cui $q = 0$, una buona stima di $\boldsymbol{\phi}$ può essere ottenuta eguagliando

le autocovarianze campionarie con quelle teoriche ai passi $0, 1, \dots, p$. Questa è la stima di Yule-Walker. Quando $q > 0$ la procedura di eguagliare le autocovarianze campionarie con quelle teoriche ai passi $0, \dots, p + q$, non risulta né semplice né efficiente. Per ottenere una buona stima dei coefficienti quando $q > 0$, illustreremo quindi un semplice metodo basato sull'algoritmo delle innovazioni. Un'altra procedura di stima è il metodo di massima verosimiglianza, che verrà però descritto nel Paragrafo 6.1.3 quando illustreremo il procedimento di Box e Jenkins.

5.3.1 Lo stimatore di Yule-Walker per i parametri di un processo autoregressivo

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo autoregressivo causale con media nulla,

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2). \quad (5.5)$$

Il nostro obiettivo è trovare degli stimatori per il vettore dei coefficienti $\phi = (\phi_1, \dots, \phi_p)'$ e per la varianza σ^2 del rumore bianco a partire dalle osservazioni X_1, \dots, X_n .

L'ipotesi che il processo sia causale ci permette di scriverlo nella forma

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad (5.6)$$

dove, grazie al Teorema 2.3.3, $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = 1/\phi(z)$, per $|z| \leq 1$. Moltiplicando entrambe le parti dell'equazione (5.5) per X_{t-j} , con $j = 0, \dots, p$, passando alle speranze e usando la (5.6) per valutare la parte destra dell'equazione, otteniamo le *equazioni di Yule-Walker*:

$$\Gamma_p \phi = \gamma_p \quad (5.7)$$

e

$$\sigma^2 = \gamma(0) - \phi' \gamma_p, \quad (5.8)$$

dove Γ_p è la matrice di covarianza $[\gamma(i-j)]_{i,j=1}^p$ e $\gamma_p = (\gamma(1), \gamma(2), \dots, \gamma(p))'$. Queste equazioni possono essere utilizzate per determinare $\gamma(0), \dots, \gamma(p)$ a

partire da σ^2 e ϕ .

D'altra parte, se sostituiamo le autocovarianze $\gamma(j)$, per $j = 0, \dots, p$, nelle (5.7) e (5.8) con le corrispondenti funzioni di autocovarianza campionarie, otteniamo un sistema di equazioni per determinare gli stimatori di Yule-Walker $\hat{\phi} = \hat{\phi}(n)$ e $\hat{\sigma}^2 = \hat{\sigma}^2(n)$ di ϕ e σ^2 , ossia

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \quad (5.9)$$

e

$$\hat{\sigma}^2 = \hat{\gamma}_n(0) - \hat{\phi}' \hat{\gamma}_p, \quad (5.10)$$

dove $\hat{\Gamma}_p = [\hat{\gamma}_n(i-j)]_{i,j=1}^p$ e $\hat{\gamma}_p = (\hat{\gamma}_n(1), \hat{\gamma}_n(2), \dots, \hat{\gamma}_n(p))'$. (Notare che, per semplificare la notazione, nelle formule precedenti abbiamo sottointeso la dipendenza da n .)

Se $\hat{\gamma}_n(0) > 0$, allora $\hat{\Gamma}_p$ è invertibile. Dividendo ogni membro della (5.9) per $\hat{\gamma}_n(0)$, otteniamo

$$\hat{\phi} = \hat{R}_p^{-1} \hat{\rho}_p \quad (5.11)$$

e

$$\hat{\sigma}^2 = \hat{\gamma}_n(0)[1 - \hat{\rho}_p' \hat{R}_p^{-1} \hat{\rho}_p], \quad (5.12)$$

dove $\hat{\rho}_p = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(p))' = \hat{\gamma}_p / \hat{\gamma}_n(0)$. Se $\hat{\phi}$ è definita come nella (5.11), si può dimostrare che $1 - \hat{\phi}_1 z - \dots - \hat{\phi}_p z^p \neq 0$, per $|z| \leq 1$. Il modello

$$X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p} = Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \hat{\sigma}^2),$$

è dunque causale. Le autocovarianze $\gamma_F(h)$, per $h = 0, \dots, p$, associate al modello devono soddisfare le $p + 1$ equazioni lineari ((5.7) e (5.8))

$$\gamma_F(h) - \hat{\phi}_1 \gamma_F(h-1) - \dots - \hat{\phi}_p \gamma_F(h-p) = \begin{cases} 0, & \text{per } h = 1, \dots, p, \\ \sigma^2, & \text{per } h = 0. \end{cases}$$

Utilizzando le equazioni (5.9) e (5.10), otteniamo che la soluzione delle precedenti equazioni è $\gamma_F(h) = \hat{\gamma}_n(h)$, per $h = 0, \dots, p$, ossia la funzione di autocovarianza del modello, ai passi $0, \dots, p$, coincide con la corrispondente funzione di autocovarianza campionaria.

Si può dimostrare (si veda Teorema 8.1.1 di [7]) che, se $(X_t)_{t \in \mathbb{Z}}$ è un processo $AR(p)$ causale della forma (5.5) con $(Z_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2)$ e $\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}(n)$ è lo stimatore di Yule-Walker di $\boldsymbol{\phi}$, allora

$$n^{1/2}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \Gamma_p^{-1}),$$

dove Γ_p è la matrice di covarianza $[\gamma(i-j)]_{i,j=1}^p$. Inoltre,

$$\widehat{\sigma}^2 = \widehat{\sigma}^2(n) \xrightarrow{P} \sigma^2.$$

L'ordine p è di solito non conosciuto. Se l'ordine è p e noi tentiamo di trovare un processo di ordine m , ci aspettiamo che la stima del vettore dei coefficienti $\widehat{\boldsymbol{\phi}}_m = (\widehat{\phi}_{m1}, \dots, \widehat{\phi}_{mm})'$ avrà un valore piccolo per $\widehat{\phi}_{mm}$ per ogni $m > p$. Infatti, se $(X_t)_{t \in \mathbb{Z}}$ è un processo $AR(p)$ causale della forma (5.5) con $(Z_t)_{t \in \mathbb{Z}} \sim IID(0, \sigma^2)$ e se $\widehat{\boldsymbol{\phi}}_m = (\widehat{\phi}_{m1}, \dots, \widehat{\phi}_{mm})' = \widehat{R}_m^{-1} \widehat{\boldsymbol{\rho}}_m$ per $m > p$, allora

$$n^{1/2}(\widehat{\boldsymbol{\phi}}_m - \boldsymbol{\phi}_m) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \Gamma_m^{-1}), \quad (5.13)$$

dove $\boldsymbol{\phi}_m = R_m^{-1} \boldsymbol{\rho}_m$. In particolare, per $m > p$,

$$n^{1/2} \widehat{\phi}_{mm} \xrightarrow{d} N(0, 1). \quad (5.14)$$

(Notare che in $\widehat{\boldsymbol{\phi}}_m$ abbiamo sottointeso la dipendenza da n , per semplificare la notazione.)

5.3.2 L'algoritmo di Durbin-Levinson per la stima dei parametri di un processo autoregressivo

Supponiamo di avere delle osservazioni x_1, \dots, x_n di una serie storica stazionaria con media nulla. Se $\widehat{\gamma}_n(0) > 0$, possiamo trovare un processo autoregressivo di ordine $m < n$, usando le equazioni di Yule-Walker (notare che abbiamo sottointeso la dipendenza da n):

$$X_t - \widehat{\phi}_{m1} X_{t-1} - \dots - \widehat{\phi}_{mm} X_{t-m} = Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \widehat{v}_m), \quad (5.15)$$

dove

$$\widehat{\boldsymbol{\phi}}_m := (\widehat{\phi}_{m1}, \dots, \widehat{\phi}_{mm})' = \widehat{R}_m^{-1} \widehat{\boldsymbol{\rho}}_m \quad (5.16)$$

e

$$\hat{v}_m = \hat{\gamma}(0)[1 - \hat{\rho}'_m \hat{R}_m^{-1} \hat{\rho}_m]. \quad (5.17)$$

Confrontando le equazioni (5.16) e (5.17) con quanto visto nel paragrafo 4.1, vediamo che $\hat{\phi}_m$ e \hat{v}_m sono collegati all'autocovarianza campionaria nello stesso modo in cui ϕ_m e v_m sono collegati all'autocovarianza del processo $(X_t)_{t \in \mathbb{Z}}$. Di conseguenza, se $\hat{\gamma}_n(0) > 0$, le matrici $\hat{R}_1, \hat{R}_2, \dots$ sono invertibili e possiamo utilizzare l'algoritmo di Durbin-Levinson per trovare un modello autoregressivo per i dati di ordine $1, 2, \dots$. Le stime $\hat{\phi}_1, \hat{\phi}_2, \dots$ e $\hat{v}_1, \hat{v}_2, \dots$ vengono calcolate ricorsivamente a partire dalle autocovarianze campionarie nello stesso modo in cui vengono calcolati i coefficienti ϕ_1, ϕ_2, \dots , e v_1, v_2, \dots , a partire dalle autocovarianze (vedi Paragrafo 4.2.1). L'algoritmo risulta quindi il seguente.

Algoritmo di Durbin-Levinson per costruire un modello autoregressivo che si adatti ai dati.

Se $\hat{\gamma}_n(0) > 0$, allora un appropriato modello autoregressivo (5.15), per $m = 1, 2, \dots, n - 1$, può essere determinato ricorsivamente dalle relazioni

$$\begin{aligned} \hat{\phi}_{11} &= \hat{\rho}_n(1), \quad \hat{v}_1 = \hat{\gamma}_n(0)[1 - \hat{\rho}_n^2(1)], \\ \hat{\phi}_{mm} &= \left[\hat{\gamma}_n(m) - \sum_{j=1}^{m-1} \hat{\phi}_{m-1,j} \hat{\gamma}_n(m-j) \right] / \hat{v}_{m-1}, \end{aligned} \quad (5.18)$$

$$\begin{bmatrix} \hat{\phi}_{m1} \\ \vdots \\ \hat{\phi}_{m,m-1} \end{bmatrix} = \hat{\phi}_{m-1} - \hat{\phi}_{mm} \begin{bmatrix} \hat{\phi}_{m-1,m-1} \\ \vdots \\ \hat{\phi}_{m-1,1} \end{bmatrix} \quad (5.19)$$

e

$$\hat{v}_m = \hat{v}_{m-1}(1 - \hat{\phi}_{mm}^2). \quad (5.20)$$

Usando la ricorsività evitiamo di calcolare la matrice inversa richiesta nel calcolo diretto di $\hat{\phi}_m$ e \hat{v}_m con le (5.11) e (5.12). Questo algoritmo ci fornisce, inoltre, anche le stime $\hat{\phi}_{11}, \hat{\phi}_{22}, \dots$, della funzione di autocorrelazione parziale ai passi $1, 2, \dots$. Queste stime sono di estrema importanza, innanzitutto

per decidere se il modello autoregressivo è appropriato o meno, e, successivamente, per la scelta dell'esatto ordine del modello.

Dal paragrafo 1.2.3 sappiamo che per un processo $AR(p)$ le autocorrelazioni parziali $\alpha_X(m) = \phi_{mm}$ sono nulle per $m > p$. Inoltre, per quanto detto nel paragrafo precedente, per un processo $AR(p)$, lo stimatore $\hat{\phi}_{mm} = \hat{\phi}_{mm}(n)$ ha approssimativamente distribuzione normale con media 0 e varianza $1/n$, quando n è grande e $m > p$. Se un modello autoregressivo è appropriato per i dati, di conseguenza, dovrebbe esistere un valore finito oltre il quale i valori di $\hat{\phi}_{mm}$ sono compatibili con la distribuzione $N(0, 1/n)$. In particolare, se l'ordine del processo è p , allora, per $m > p$, $\hat{\phi}_{mm}$ cadrà all'interno dell'intervallo $\pm 1.96n^{-1/2}$ con probabilità circa del 95%. Ciò ci suggerisce di prendere come stima preliminare di p il più piccolo valore r tale che $|\hat{\phi}_{mm}| < 1.96n^{-1/2}$ per $m > r$. Scelto il valore di p , il modello è dato dalle equazioni (5.15), (5.16) e (5.17) con $m = p$.

Gli intervalli di confidenza asintotici per il vettore dei coefficienti ϕ_p e per ogni sua componente ϕ_{pj} possono essere trovati con l'aiuto delle formule (5.13) e (5.14). Se $\chi_{1-\alpha}^2(p)$ denota il quantile di livello $(1 - \alpha)$ della distribuzione chi quadrato con p gradi di libertà, allora, per un grande numero di osservazioni n , la regione

$$\{\phi \in \mathbb{R}^p : (\phi - \hat{\phi}_p)' \hat{\Gamma} (\phi - \hat{\phi}_p) \leq n^{-1} \hat{v}_p \chi_{1-\alpha}^2(p)\} \quad (5.21)$$

contiene ϕ_p con probabilità circa pari a $(1 - \alpha)$.

Allo stesso modo, se $\Phi_{1-\alpha}$ denota il quantile a livello $(1 - \alpha)$ della legge normale standard e \hat{v}_{jj} è l'elemento j -esimo della diagonale principale di $\hat{v}_p \hat{\Gamma}_p^{-1}$, allora, quando n è grande, l'intervallo

$$\{\phi \in \mathbb{R} : |\phi - \hat{\phi}_{pj}| \leq n^{-1/2} \Phi_{1-\alpha/2} \hat{v}_{jj}^{1/2}\} \quad (5.22)$$

conterà ϕ_{pj} con probabilità circa pari a $(1 - \alpha)$.

5.3.3 La stima dei parametri di un processo a media mobile usando l'algoritmo delle innovazioni

In questo paragrafo illustriamo come determinare un modello a media mobile,

$$X_t = Z_t + \hat{\theta}_{m1}Z_{t-1} + \cdots + \hat{\theta}_{mm}Z_{t-m}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \hat{v}_m), \quad (5.23)$$

di ordine $m = 1, 2, \dots$, utilizzando l'algoritmo delle innovazioni. Gli stimatori $\hat{\theta}_m := (\hat{\theta}_{m1}, \dots, \hat{\theta}_{mm})'$ e \hat{v}_m , per $m = 1, 2, \dots$, sono definiti nel seguente modo.

Definizione 5.1. Se $\hat{\gamma}_n(0) > 0$, definiamo gli stimatori $\hat{\theta}_m = \hat{\theta}_m(n)$, $\hat{v}_m = \hat{v}_m(n)$ dell'equazione (5.23) per $m = 1, 2, \dots, n-1$, con le seguenti relazioni ricorsive (notare che, per semplificare la notazione, abbiamo sottinteso la dipendenza da n):

$$v_0 = \hat{\gamma}_n(0),$$

$$\hat{\theta}_{m,m-k} = \hat{v}_k^{-1} \left[\hat{\gamma}_n(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{v}_j \right], \quad k = 0, \dots, m-1, \quad (5.24)$$

e

$$\hat{v}_m = \hat{\gamma}_n(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 \hat{v}_j. \quad (5.25)$$

Osservazione 5.2. C'è una differenza importante tra l'algoritmo di Durbin-Levison per un modello autoregressivo e quello delle innovazioni per un modello a media mobile: per un processo $AR(p)$ lo stimatore di Yule-Walker $\hat{\phi}_p = (\hat{\phi}_{p1}, \dots, \hat{\phi}_{pp})'$ è consistente con ϕ_p (ossia $\hat{\phi}_p \xrightarrow{P} \phi_p$); mentre per un processo $MA(q)$ lo stimatore $\hat{\theta}_q = (\hat{\theta}_{q1}, \dots, \hat{\theta}_{qq})'$ non è consistente.

Osservazione 5.3. La scelta di m , per un certo numero di osservazioni, può essere fatta aumentando il valore di m finché il vettore $(\hat{\theta}_{q1}, \dots, \hat{\theta}_{qq})'$ si stabilizza. Nella pratica c'è un valore di m (m maggiore di $n/4$ per un numero di osservazioni compreso tra 100 e 500) sopra il quale le fluttuazioni di $\hat{\theta}_{mj}$ sono piccole paragonate alla stima asintotica della deviazione standard $n^{-1/2}(\sum_{k=0}^{j-1} \hat{\theta}_{mk}^2)^{1/2}$. Dal paragrafo 2.4 sappiamo che per un processo $MA(q)$

le autocorrelazioni $\rho(m)$ sono nulle per $m > q$. Dalla formula di Barlett sappiamo, inoltre, che la funzione di autocorrelazione campionaria $\hat{\rho}_n(m)$, per $m > q$, è approssimata da una distribuzione normale con media $\rho(m) = 0$ e varianza $n^{-1}[1 + 2\rho^2(1) + \dots + 2\rho^2(q)]$. Questo risultato ci permette di usare il grafico di $\hat{\rho}_n(m)$, per $m = 1, 2, \dots$, per decidere se un dato insieme di osservazioni può essere modellato da un processo a media mobile e anche di ottenere una prima stima dell'ordine q . Questa procedura è stata descritta nell'esempio 27. Se, oltre ad esaminare $\hat{\rho}_n(m)$, per $m = 1, 2, \dots$, esaminiamo anche i vettori dei coefficienti $\hat{\theta}_m$, per $m = 1, 2, \dots$, possiamo anche ottenere una prima stima $\hat{\theta}_{m1}, \dots, \hat{\theta}_{mq}$ dei coefficienti. Per far questo, riportiamo su un grafico i valori $\hat{\theta}_{m1}, \dots, \hat{\theta}_{mm}, 0, 0, \dots$ per $m = 1, 2, \dots$, finché i valori si stabilizzano.

Poiché la varianza asintotica di $\hat{\theta}_{mj}$ è $\sigma_j^2(\theta_1, \dots, \theta_{j-1}) = n^{-1} \sum_{k=0}^{j-1} \theta_k^2$, nel grafico riportiamo anche le rette $\pm 1.96\hat{\sigma}_j$, dove $\hat{\sigma}_j = \sigma_j(\hat{\theta}_{m1}, \dots, \hat{\theta}_{m,j-1})$. Un valore di $\hat{\theta}_{mj}$ al di fuori della regione individuata da queste rette, ci suggerisce che il corrispondente coefficiente θ_j è diverso da zero. La stima di θ_j è $\hat{\theta}_{mj}$, mentre il più grande passo per il quale $\hat{\theta}_{mj}$ cade al di fuori delle rette $\pm 1.96\hat{\sigma}_j$ è la stima dell'ordine q del processo a media mobile.

L'intervallo di confidenza asintotico di livello 95%, per il vettore dei coefficienti θ_q ed ogni sua componente è dato da

$$\left\{ \theta \in \mathbb{R} : |\theta - \hat{\theta}_{mj}| \leq 1.96n^{-1/2} \left(\sum_{k=0}^{j-1} \hat{\theta}_{mk}^2 \right)^{1/2} \right\}.$$

5.3.4 La stima dei parametri di un processo ARMA

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo ARMA(p,q) causale con media nulla:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2). \quad (5.26)$$

Grazie al fatto che il processo è causale, possiamo scrivere

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

dove, per le equazioni (2.15) e (2.16), i coefficienti ψ_j soddisfano

$$\begin{cases} \psi_0 = 1, \\ \psi_j = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i}, \quad \text{per } j = 1, 2, \dots, \end{cases} \quad (5.27)$$

$\theta_j = 0$ per $j > q$ e $\phi_j = 0$ per $j > p$. Per stimare $\psi_1, \dots, \psi_{p+q}$, possiamo usare gli stimatori $\hat{\theta}_{m1}, \dots, \hat{\theta}_{m,p+q}$ del paragrafo precedente. Sostituendo ψ_j con $\hat{\theta}_{mj}$ nelle equazioni (5.27) e risolvendole, otteniamo

$$\hat{\theta}_{mj} = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \hat{\theta}_{m,j-i}, \quad \text{per } j = 1, 2, \dots, p+q. \quad (5.28)$$

Dalle equazioni (5.28), con $j = q+1, \dots, q+p$, vediamo che $\hat{\phi}$ dovrebbe soddisfare il sistema

$$\begin{bmatrix} \hat{\theta}_{m,q+1} \\ \hat{\theta}_{m,q+2} \\ \vdots \\ \hat{\theta}_{m,q+p} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{m,q} & \hat{\theta}_{m,q-1} & \cdots & \hat{\theta}_{m,q+1-p} \\ \hat{\theta}_{m,q+1} & \hat{\theta}_{m,q} & \cdots & \hat{\theta}_{m,q+2-p} \\ \vdots & \vdots & & \vdots \\ \hat{\theta}_{m,q+p-1} & \hat{\theta}_{m,q+p} & \cdots & \hat{\theta}_{m,q} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}. \quad (5.29)$$

Risolvendo (5.29), troviamo lo stimatore di θ :

$$\hat{\theta}_j = \hat{\theta}_{mj} - \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\theta}_{m,j-i}, \quad \text{per } j = 1, 2, \dots, q.$$

La varianza del rumore bianco è stimata infine ponendo

$$\hat{\sigma}^2 = \hat{v}_m.$$

Nel caso di un processo a media mobile, ossia $p = 0$, il metodo si riduce al metodo descritto nel paragrafo 5.3.3.

Capitolo 6

I processi SARIMA e l'analisi di una serie storica reale

In questo capitolo affronteremo il problema di identificare, a partire da una serie storica osservata $\{x_t, t = 1, \dots, n\}$, un modello SARIMA atto a descrivere le relazioni di dipendenza temporale osservate nei dati e che possa ritenersi il processo generatore della serie stessa.

In questo capitolo illustreremo la *procedura di Box e Jenkins* per la costruzione di modelli SARIMA. Prima di descrivere la procedura di Box e Jenkins, diamo quindi la definizione di processo ARIMA e di processo SARIMA.

Definizione 6.1. Sia d un intero non negativo, allora il processo $(X_t)_{t \in \mathbb{Z}}$ è un processo *autoregressivo-integrato a media mobile* (in inglese, *autoregressive-integrated moving average process*) o, più brevemente, un processo *ARIMA*(p, d, q) se $Y_t := (1 - B)^d X_t$ è un processo ARMA(p, q).

Da questa definizione segue che il processo $(X_t)_{t \in \mathbb{Z}}$ soddisfa l'equazione alle differenze della forma

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2), \quad (6.1)$$

dove $\phi_p(B)$ e $\theta_q(B)$ sono polinomi di grado p e q rispettivamente e B è l'operatore ritardo introdotto nel paragrafo 1.3.

La terminologia *integrato* segue dal fatto che, se $Y_t = (1 - B)^d X_t$, allora il processo $(X_t)_{t \in \mathbb{Z}}$ può essere ottenuto dall'espressione inversa $X_t = (1 - B)^{-d} Y_t$, ossia come somme ripetute di $(Y_t)_{t \in \mathbb{Z}}$, che analogamente, per processi a tempo continuo, corrisponde a un integrale.

Una classe più generale dei processi $ARIMA(p, d, q)$ sono i processi *autoregressivi-integrati a media mobile stagionali moltiplicativi* (in inglese, *seasonal autoregressive-integrated moving average processes*), detti, più brevemente, $SARIMA(p, d, q)(P, D, Q)_s$, i quali sono definiti dalla relazione

$$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d X_t = \Theta_Q(B^s)\theta_q(B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim WN(0, \sigma^2) \quad (6.2)$$

dove p è l'ordine del polinomio autoregressivo regolare, d è il grado delle differenze regolari, q è l'ordine del polinomio media mobile regolare, mentre P, D e Q sono relativi alle componenti stagionali ed s è il periodo di stagionalità. Nel seguito indicheremo questa classe di processi con la notazione più comune $ARIMA(p, d, q)(P, D, Q)_s$.

6.1 La procedura di Box e Jenkins

La procedura iterativa di Box e Jenkins ha come obiettivo l'identificazione di un modello $ARIMA(p, d, q)(P, D, Q)_s$ della forma (6.2) con $(Z_t)_{t \in \mathbb{Z}}$ rumore bianco gaussiano, che possa ritenersi il processo generatore di una serie storica osservata in modo da poterlo usare per fare previsioni. Questa procedura è composta da cinque passi:

1. **Analisi preliminare:** dobbiamo stabilire se la serie è stazionaria e, qualora non lo sia, trasformarla in modo da far sì che lo diventi. In questa fase individuiamo i valori degli ordini d, D e il valore di s ;
2. **Identificazione del modello ARIMA:** individuiamo gli ordini p, q, P e Q in modo da identificare il modello ARIMA che più si adatta alla serie storica osservata resa stazionaria;

3. **Stima del modello:** stimiamo i parametri del modello ARIMA identificato, mediante il metodo della massima verosimiglianza;
4. **Verifica del modello:** verifichiamo la bontà e l'adeguatezza del modello stimato ai dati osservati. Se il modello non supera tale verifica allora la natura iterativa della procedura prevede di rivalutare le scelte effettuate nei primi due passi;
5. **Utilizzo del modello:** utilizziamo il modello stimato e che ha superato la fase di verifica per fare previsioni o proiezioni future.

Nel seguito descriveremo e discuteremo singolarmente i vari passi di questa procedura.

6.1.1 Analisi preliminare

Questa prima fase della procedura di Box e Jenkins consiste nello stabilire se la serie storica è stazionaria, mentre in caso negativo dovremo operare con delle opportune trasformazioni per renderla stazionaria.

Possiamo stabilire se i dati sono o meno la realizzazione di un processo stazionario osservando il grafico della serie storica e il grafico della sua funzione di autocorrelazione. Se la serie storica presenta delle caratteristiche che suggeriscono che è realizzazione di un processo non stazionario, allora dobbiamo applicare una o più trasformazioni ai dati, in modo da ottenere dei dati che risultino compatibili con un processo stazionario.

La non stazionarietà della serie storica può essere dovuta alla presenza del trend e/o della stagionalità e/o di un comportamento irregolare nell'ampiezza delle oscillazioni. Nei primi due casi diremo che la serie storica è *non stazionaria in media*, mentre nell'ultimo caso diremo che la serie storica è *non stazionaria in varianza*.

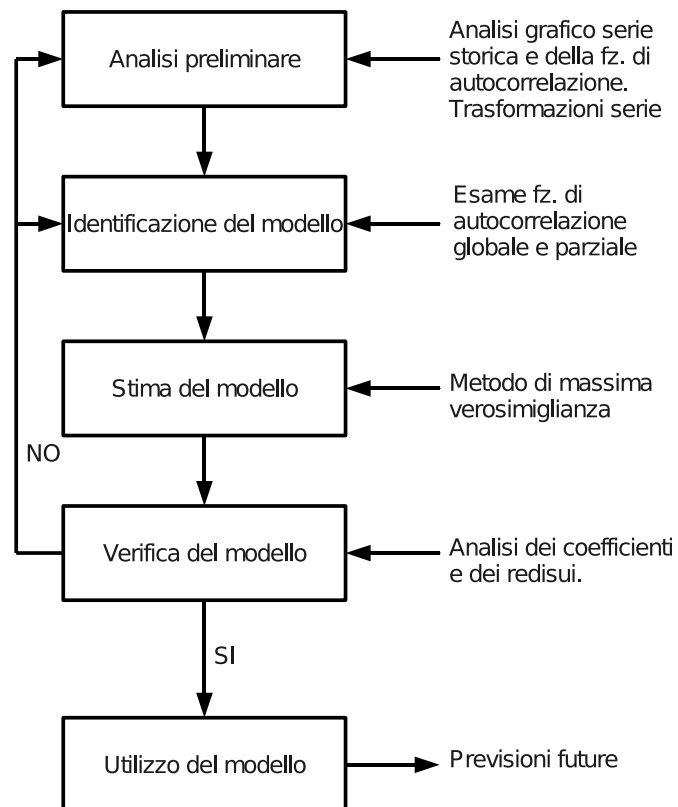


Figura 6.1: Diagramma della procedura di Box e Jenkins.

Se il grafico della serie storica risulta caratterizzato dalla presenza di una tendenza di lungo periodo e, quindi, da un andamento crescente o decrescente allora la serie storica presenta un trend. Quest'ultimo può emergere anche dall'analisi del grafico della funzione di autocorrelazione nel caso in cui tenda a zero molto lentamente. La serie storica può risultare non stazionaria in quanto presenta una componente stagionale, quindi un comportamento periodico che si ripete ad intervalli di tempo successivi di ampiezza pari ad s , con $s > 1$. In questo caso le osservazioni rilevate ogni s istanti temporali hanno lo stesso comportamento. Anche la presenza della stagionalità può essere rilevata dal grafico della serie storica. Per esempio, se abbiamo una serie di dati osservati mensilmente caratterizzati da un picco di massimo o di minimo nello stesso mese di ogni anno, diremo che la serie presenta una stagionalità annuale e quindi $s = 12$. Nel caso in cui la serie evidenzia ampiezze irregolari e variazioni di livello delle oscillazioni lungo l'asse temporale allora risulterà non stazionaria in varianza.

Se la serie storica risulta non stazionaria dobbiamo operare con delle opportune trasformazioni in modo da renderla stazionaria e poter proseguire con la procedura di Box e Jenkins. Per eliminare il trend e la componente stagionale possiamo utilizzare rispettivamente gli operatori differenza regolare $\nabla^d = (1 - B)^d$ e stagionale $\nabla_s^D = (1 - B^s)^D$, definiti in equazioni (1.9) e (1.13). Per eliminare, invece, la non stazionarietà in varianza possiamo utilizzare le seguenti trasformazioni di Box-Cox:

$$f_\lambda(X_t) = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log X_t, & \lambda = 0. \end{cases} \quad (6.3)$$

Il valore di λ può essere individuato attraverso il metodo di massima verosimiglianza o attraverso il metodo basato sulla relazione di Bartlett tra la deviazione standard e la media: $\sigma = k\mu^{1-\lambda}$ con k costante. Con il primo metodo, il valore di λ viene calcolato massimizzando la funzione di log-verosimiglianza

$$l(\lambda) = -\frac{n}{2} \log(\hat{\sigma}^2) + (\lambda - 1) \sum_{t=1}^n \log(X_t) + c,$$

dove n è il numero totale dei dati, c è una costante e $\hat{\sigma}^2$ è la varianza dei dati trasformati, ossia

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \left[f_\lambda(X_t) - \frac{1}{n} \sum_{t=1}^n f_\lambda(X_t) \right]^2.$$

Per quanto riguarda il secondo metodo, invece, supponiamo che la varianza di X_t soddisfi la seguente relazione

$$\text{Var}(X_t) = f(\mu_t),$$

dove $\mu_t = E[X_t]$. Vogliamo trovare una trasformazione $T(\cdot)$ tale che $\text{Var}(T(X_t))$ risulti costante. Grazie allo sviluppo di Taylor al primo ordine di $T(X_t)$ in μ_t , otteniamo

$$T(X_t) \approx T(\mu_t) + T'(\mu_t)(X_t - \mu_t),$$

da cui discende l'approssimazione

$$\text{Var}(T(X_t)) \approx [T'(\mu_t)]^2 \text{Var}(X_t) = [T'(\mu_t)]^2 f(\mu_t).$$

Ricaviamo, dunque, la relazione

$$T'(\mu_t) = \frac{c}{\sqrt{f(\mu_t)}},$$

dove c è una costante, ossia

$$T(x) = \int \frac{c}{\sqrt{f(x)}} dx.$$

Per esempio, se abbiamo una varianza direttamente proporzionale al quadrato della media, ossia una deviazione standard direttamente proporzionale alla media, allora dovremo scegliere una trasformazione del tipo

$$T(x) = \int \frac{c}{\sqrt{x^2}} dx = \int \frac{c}{x} dx = c \log(x) + \tilde{c}.$$

In altre parole, dovremo utilizzare una trasformazione logaritmica sulla serie storica. Per studiare la relazione tra la deviazione standard e la media dei

dati da cui desumere il valore di λ , viene usato il grafico *range mean plot*. Questo grafico viene creato suddividendo le osservazioni della serie in sottoperiodi di uguale ampiezza e calcolando per ognuno di essi il valore medio e la deviazione standard. Realizzando il grafico delle coppie di valori ottenuti, dove sull'asse delle ascisse poniamo la media e su quella delle ordinate la deviazione standard, possiamo determinare il valore di λ interpolando i punti ottenuti e scegliendo:

- $\lambda = 0$, nel caso in cui i punti si distribuiscano lungo una retta del tipo $\sigma = k\mu$;
- $\lambda < 0$, nel caso in cui la deviazione standard cresca più velocemente della media (per esempio $\lambda = -1$, se abbiamo $\sigma = k\mu^2$);
- $0 < \lambda < 1$, nel caso in cui la deviazione standard cresca meno velocemente della media (per esempio $\lambda = 1/2$, se abbiamo $\sigma = k\sqrt{\mu}$);

Nel caso in cui i punti si distribuiscano lungo una retta parallela all'asse delle ascisse, la serie storica non viene trasformata in quanto la deviazione standard è costante e, quindi, la serie risulta già stazionaria in varianza.

Per verificare, infine, che la serie storica trasformata sia stazionaria, si analizza il grafico della funzione di autocorrelazione, che dovrà tendere velocemente verso zero con un andamento coerente con quello della serie originale.

In questa prima fase della procedura di Box e Jenkins si individuano quindi i valori dei parametri s , d e D del modello *ARIMA* che verrà scelto come generatore della serie storica analizzata. In genere, i valori di d e D sono 0, 1 o 2. Se la serie storica non presenta un trend e la componente stagionale prenderemo rispettivamente $d = 0$ e $D = 0$. Nel caso in cui la presenza del trend e della stagionalità abbia un andamento lineare allora sceglieremo $d = 1$ e $D = 1$; nel caso presentino un andamento quadratico prenderemo $d = 2$ e $D = 2$.

6.1.2 Identificazione del modello

L'obiettivo del secondo passo della procedura di Box e Jenkins è quello di identificare il modello $ARIMA(p, d, q)(P, D, Q)_s$ più appropriato a descrivere la serie considerata.

Consideriamo la serie storica stazionaria che abbiamo ottenuto applicando le trasformazioni selezionate nel primo passo della procedura e analizziamo i grafici della funzione di autocorrelazione globale e parziale che permettono di individuare i valori degli ordini dei polinomi regolari, p e q , e stagionali, P e Q , attraverso i seguenti passi:

1. deduciamo l'ordine del polinomio media mobile regolare q dall'analisi dei primi 10 - 12 ritardi della funzione di autocorrelazione globale della serie resa stazionaria: q sarà il numero di ritardi consecutivi in cui la funzione risulterà significativamente diversa da zero;
2. ricaviamo l'ordine del polinomio media mobile stagionale Q dall'analisi dei ritardi stagionali della funzione di autocorrelazione della serie resa stazionaria: cioè analizziamo la funzione in corrispondenza di $k = s, 2s, 3s, \dots$ e Q sarà il numero dei ritardi consecutivi in cui la funzione assume valori significativi;
3. desumiamo l'ordine del polinomio autoregressivo regolare p analizzando i primi 10-12 ritardi della funzione di autocorrelazione parziale della serie resa stazionaria: p sarà il numero di ritardi consecutivi in cui la funzione risulta significativamente diversa da zero;
4. determiniamo l'ordine del polinomio autoregressivo stagionale P analizzando i ritardi stagionali della funzione di autocorrelazione parziale della serie resa stazionaria: cioè analizziamo la funzione in corrispondenza di $k = s, 2s, 3s, \dots$ e l'ordine P sarà il numero dei ritardi consecutivi in cui la funzione di autocorrelazione parziale risulta significativa.

Possiamo verificare se la funzione di autocorrelazione, globale o parziale, è significativamente diverso da zero confrontandola con l'intervallo di confi-

denza (si veda il Paragrafo 5.2), ossia se il suo valore cade al di fuori della regione di confidenza.

6.1.3 Stima del modello

Dopo la fase di identificazione del modello ARIMA più appropriato, la procedura di Box e Jenkins prosegue con la fase di stima dei suoi parametri. Il metodo di stima suggerito da tale procedura è basato sul metodo della massima verosimiglianza e opera sulla serie resa stazionaria. È un metodo iterativo e quindi è importante partire da buoni valori iniziali per avere un minor numero di iterazioni necessarie per individuare la stima ottimale dei parametri.

Metodo della massima verosimiglianza

Sia $(X_t)_{t \in \mathbb{Z}}$ un processo gaussiano $ARMA(p, q)$ invertibile, definito dall'equazione

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (Z_t)_{t \in \mathbb{Z}} \sim GWN(0, \sigma^2).$$

Dobbiamo innanzitutto costruire la funzione di verosimiglianza

$$f_{X_1, X_2, \dots, X_n}(X_1, X_2, \dots, X_n; \boldsymbol{\tau}) = f(\boldsymbol{\tau}),$$

dove $\boldsymbol{\tau} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)'$. Successivamente dobbiamo calcolare il valore del vettore $\boldsymbol{\tau}$ che massimizza la funzione di verosimiglianza $f(\boldsymbol{\tau})$.

Per ottenere un'approssimazione di $f(\boldsymbol{\tau})$, dobbiamo condizionare rispetto ai valori iniziali di $(X_t)_{t \in \mathbb{Z}}$ e a quelli di $(Z_t)_{t \in \mathbb{Z}}$. Supponiamo di conoscere i valori iniziali $\mathbf{X}_0 = (X_0, X_{-1}, \dots, X_{-p+1})'$ e $\mathbf{Z}_0 = (Z_0, Z_{-1}, \dots, Z_{-q+1})'$, allora possiamo calcolare i valori $\{Z_1, Z_2, \dots, Z_n\}$ a partire dai valori $\{X_1, X_2, \dots, X_n\}$ utilizzando una procedura basata sulla relazione

$$Z_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q}, \quad (6.4)$$

per $t = 1, 2, \dots, n$.

La funzione di log-verosimiglianza condizionata corrispondente alla (6.4) è

$$l(\boldsymbol{\tau}) = \log f_{X_1, X_2, \dots, X_n}(X_1, X_2, \dots, X_n | \mathbf{X}_0, \mathbf{Z}_0; \boldsymbol{\tau}) =$$

$$-\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \sum_{t=1}^n \frac{Z_t^2}{2\sigma^2}.$$

Box e Jenkins suggeriscono di partire dall'istante $t = p + 1$ prendendo $\mathbf{Z}_0 = (Z_p, Z_{p-1}, \dots, Z_{p-q+1})' = 0$ e $\mathbf{X}_0 = (X_p, X_{p-1}, \dots, X_1)'$ coincidente con il vettore dei loro reali valori. La funzione di log-verosimiglianza risulta dunque

$$l(\boldsymbol{\tau}) = \log f_{X_{p+1}, X_{p+2}, \dots, X_n}(X_{p+1}, \dots, X_n | \mathbf{X}_0, \mathbf{Z}_0; \boldsymbol{\tau}) =$$

$$-\frac{n-p}{2} \log(2\pi) - \frac{n-p}{2} \log(\sigma^2) - \sum_{t=p+1}^n \frac{Z_t^2}{2\sigma^2}.$$

Per massimizzare la funzione di log-verosimiglianza possiamo utilizzare procedure numeriche o il filtro di Kalman. Osserviamo che l'approssimazione fornita dalla funzione di verosimiglianza condizionata è valida solo se il processo ARMA è invertibile.

6.1.4 Verifica del modello

La verifica del modello ARIMA identificato e stimato si pone l'obiettivo di provare che il modello suddetto possa essere considerato il generatore della serie analizzata. In caso contrario bisogna rivalutare le scelte fatte nei primi passi della procedura di Box e Jenkins.

Per poter considerare "buono" il modello individuato, dobbiamo analizzare le stime dei parametri, le stime dei residui e dobbiamo verificare che risulti parsimonioso, ossia che il numero di parametri utilizzati sia il minore possibile. Innanzitutto le stime dei parametri devono risultare significativamente diverse da zero, in particolare in valore assoluto il parametro deve risultare almeno due volte superiore alla sua deviazione standard. Inoltre la correlazione fra le stime dei parametri deve risultare bassa, non superiore a 0.8, per evitare problemi di instabilità.

I residui del modello ARIMA ottenuto devono potersi considerare la realizzazione di un processo rumore bianco gaussiano: devono avere media e funzione di autocorrelazione globale non significativamente diversi da zero,

la loro varianza deve mantenersi costante nel corso del tempo e devono essere normalmente distribuiti. Per valutare l'uguaglianza rispetto a zero della funzione di autocorrelazione, possiamo utilizzare la seguente *ipotesi nulla*

$$H_0 : \rho_k(Z) = 0, \quad k > 0,$$

che può essere verificata mediante la seguente statistica test t :

$$t = \frac{r_k(\widehat{Z})}{s[r_k(\widehat{Z})]} \sim t_k,$$

dove $r_k(\widehat{Z})$ è il valore dell'autocorrelazione globale dei residui del modello stimato al ritardo k e $s[r_k(\widehat{Z})]$ è la stima della sua deviazione standard.

Per valutare l'uguaglianza rispetto a zero della funzione di autocorrelazione, possiamo saggiare anche l'*ipotesi nulla congiunta*, al variare di $K = 1, 2, \dots$,

$$H_0 : \rho_1(Z) = \rho_2(Z) = \dots = \rho_K(Z) = 0,$$

attraverso il così detto *test portmanteau*, dal francese “grande valigia”, di Ljung e Box:

$$Q = n(n+2) \sum_{k=1}^K (n-k)^{-1} r_k^2(\widehat{Z}),$$

dove n è il numero totale di osservazioni e $r_k^2(\widehat{Z})$ è il valore dell'autocorrelazione globale dei residui del modello stimato al ritardo k . Se i residui del modello stimato possono ritenersi la realizzazione di un processo rumore bianco gaussiano, la statistica Q si distribuisce come una distribuzione χ^2 con $(K - p - q - P - Q)$ gradi di libertà, dove K è il numero di autocorrelazioni da “mettere in valigia”. La scelta del valore di K deve essere superiore a tutti i ritardi in corrispondenza dei quali c'è autocorrelazione. Ad esempio, per dati mensili che presentano stagionalità dovremo scegliere un valore di K compreso tra 12 e 24.

Nel caso in cui i residui non possono essere considerati realizzazione di un processo rumore bianco gaussiano, allora il modello stimato risulta incapace di cogliere e spiegare la struttura di dipendenza temporale presente nei dati.

In tal caso si possono esaminare nuovamente le funzioni di autocorrelazione globale e parziale della serie originale, oppure si possono considerare i residui stimati come una nuova serie storica su cui ripetere il percorso di identificazione, stima e verifica.

Va menzionato, infine, che nel caso in cui l'analisi delle funzioni di autocorrelazione globale e parziale portino all'identificazione di più di un modello allora è possibile selezionarne uno sulla base di criteri come quelli di informazione asintotica di Akaike, di informazione bayesiano, etc. e sulla base del principio di parsimonia.

6.1.5 Valutazione della capacità previsiva del modello

L'ultimo passo della procedura di Box e Jenkins prevede la valutazione della sua capacità previsiva. Per fare questo utilizziamo come misura l'indice MAPE (*Mean Absolute Percent Error*) che è la media dei valori assoluti degli errori relativi in percentuale

$$MAPE = \frac{100}{L} \sum_{l=1}^L \left| \frac{X_{t+l} - \hat{X}_t(l)}{X_{t+l}} \right|,$$

dove X_{t+l} è il valore osservato al tempo $t+l$, mentre $\hat{X}_t(l)$ è il corrispondente valore previsto ed L è la lunghezza del periodo di previsione. Osserviamo che i valori che andiamo a prevedere $\hat{X}_t(l)$ sono calcolati utilizzando il modello identificato e stimato ottenuto usando non tutte le osservazioni, ma solo le prime $n - L$ osservazioni, dove n è il numero di osservazioni totali e L è il numero dei valori che vogliamo prevedere. Il modello ottenuto risulterà avere una buona capacità previsiva se il MAPE corrispondente risulterà inferiore al 12 - 15%.

6.2 Un esempio di applicazione della procedura di Box e Jenkins

In questo paragrafo illustreremo l'applicazione della procedura di Box e Jenkins sulla serie storica reale dei passeggeri internazionali delle linee aeree osservati mensilmente dal gennaio del 1960 al dicembre del 1971. Per analizzare questa serie utilizziamo il programma R. La serie storica considerata *airpass* è contenuta nel pacchetto *TSA* di R.

Un estratto dalla serie è riportata nella seguente tabella:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1960	112	118	132	129	121	135	148	148	136	119	104	118
1961	115	126	141	135	125	149	170	170	158	133	114	140
...
1971	417	391	419	461	472	535	622	606	508	461	390	432

I PASSO. Disegniamo e analizziamo il grafico della serie storica che riportiamo in Figura 6.2.

Dal grafico della serie storica possiamo notare che la serie risulta non stazionaria in varianza dal momento che l'ampiezza delle oscillazioni non rimane costante lungo l'asse temporale. Questo viene confermato dal *range mean plot*, riportato in Figura 6.3, che suggerisce di applicare una trasformazione logaritmica di Box e Cox ai dati della serie prima di proseguire nell'analisi. Risulta, infatti, evidente la relazione lineare tra i valori della deviazione standard e dei valori medi calcolati per i successivi sottoperiodi della serie. Dopo aver usato una trasformazione logaritmica sui dati della serie storica, generiamo il grafico della funzione di autocorrelazione globale che riportiamo in Figura 6.4, scegliendo di rappresentare le funzioni di autocorrelazione tra le osservazioni fino al quarantottesimo ritardo, cioè fino all'ultima osservazione del quarto anno rilevato.

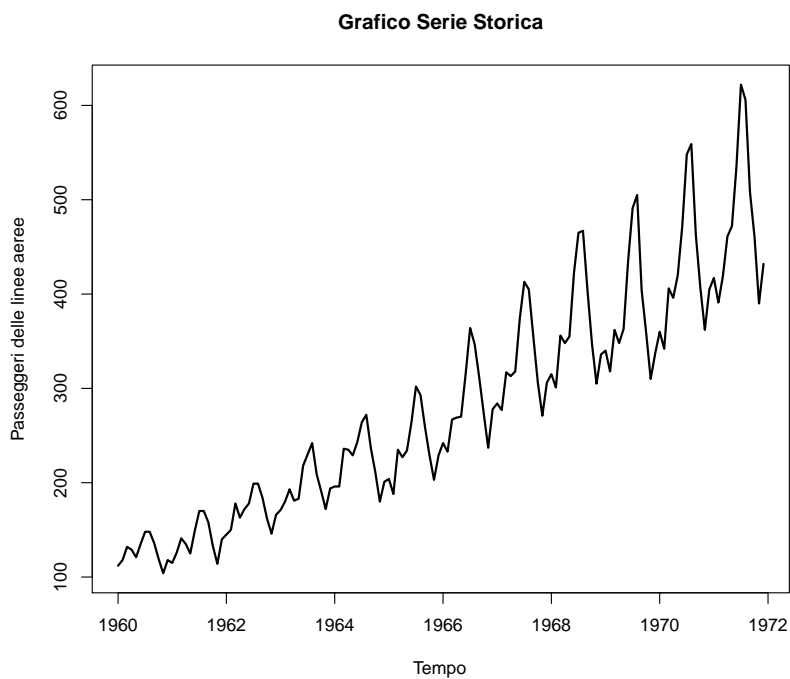


Figura 6.2: Grafico della serie storica dei passeggeri internazionali sulle linee aeree rilevati dal 1960 al 1971.

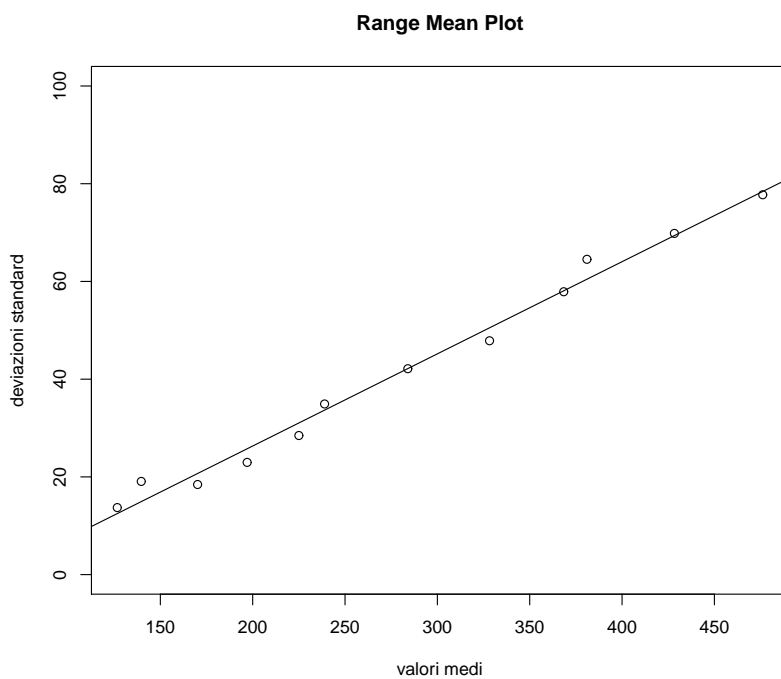


Figura 6.3: Range mean plot della serie storica non stazionaria.

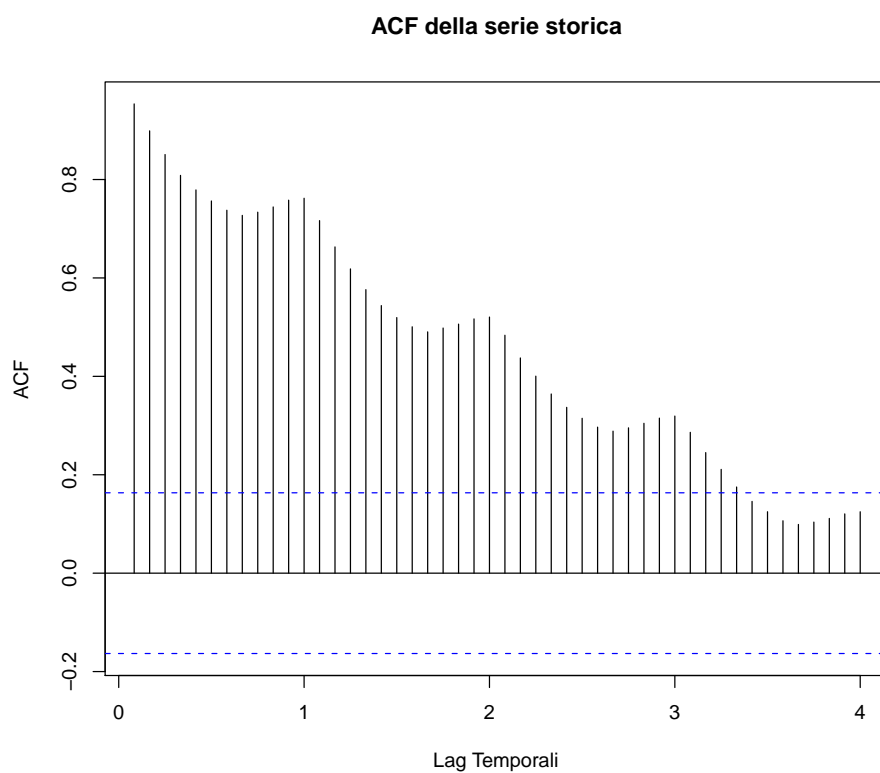


Figura 6.4: Grafico della funzione di autocorrelazione del logaritmo della serie storica oggetto di studio.

Dal grafico della funzione di autocorrelazione possiamo vedere che la funzione di autocorrelazione tende a zero con un andamento decrescente molto lento, che suggerisce la presenza di un trend lineare, e presenta delle correlazioni elevate in corrispondenza dei ritardi stagionali, che indica la presenza della stagionalità. Potevamo osservare che la serie è non stazionaria in media già dal suo grafico riportato in Figura 6.2. La serie, infatti, presenta un trend crescente piuttosto lineare e dei picchi positivi che si ripetono periodicamente. Come accennato nel precedente paragrafo, i picchi che si ripetono ad intervalli di ampiezza regolare pari a $s = 12$ e ai suoi multipli, che vengono chiamati *lag* o *ritardi stagionali*, sono indice della presenza della componente stagionale della serie e contribuiscono alla non stazionarietà in media.

A questo punto, dobbiamo operare con delle trasformazioni per rendere la serie storica stazionaria in media. In particolare, per rimuovere il trend lineare, dobbiamo applicare alla serie l'operatore differenza regolare ∇^d di ordine $d = 1$. La serie così trasformata $\{\nabla \log X_t\}$ risulta ancora non stazionaria dal momento che presenta ancora la componente stagionale. Possiamo, quindi, eliminare la componente stagionale applicando alla serie $\{\log X_t\}$ l'operatore differenza regolare stagionale ∇_s^D di ordine $D = 1$. La serie così trasformata $\{\nabla_s \log X_t\}$ non presenta più la componente stagionale ma presenta di nuovo un trend lineare che la rende ancora non stazionaria in media. Di conseguenza dobbiamo applicare entrambi gli operatori alla nostra serie originale contemporaneamente.

Riassumendo, abbiamo applicato le seguenti trasformazioni ai dati originali: prima una trasformazione logaritmica e successivamente gli operatori differenze prime, ossia i nuovi dati $\{W_t\}$ sono legati ai dati originali $\{X_t\}$ nel modo seguente $W_t = (1 - B)(1 - B^{12})Y_t = (1 - B)(1 - B^{12}) \log X_t$.

6.2 Un esempio di applicazione della procedura di Box e Jenkins 137

II PASSO Dobbiamo analizzare i grafici delle funzioni di autocorrelazione globale e parziale della serie $\{W_t\}$ che riportiamo in Figura 6.5.

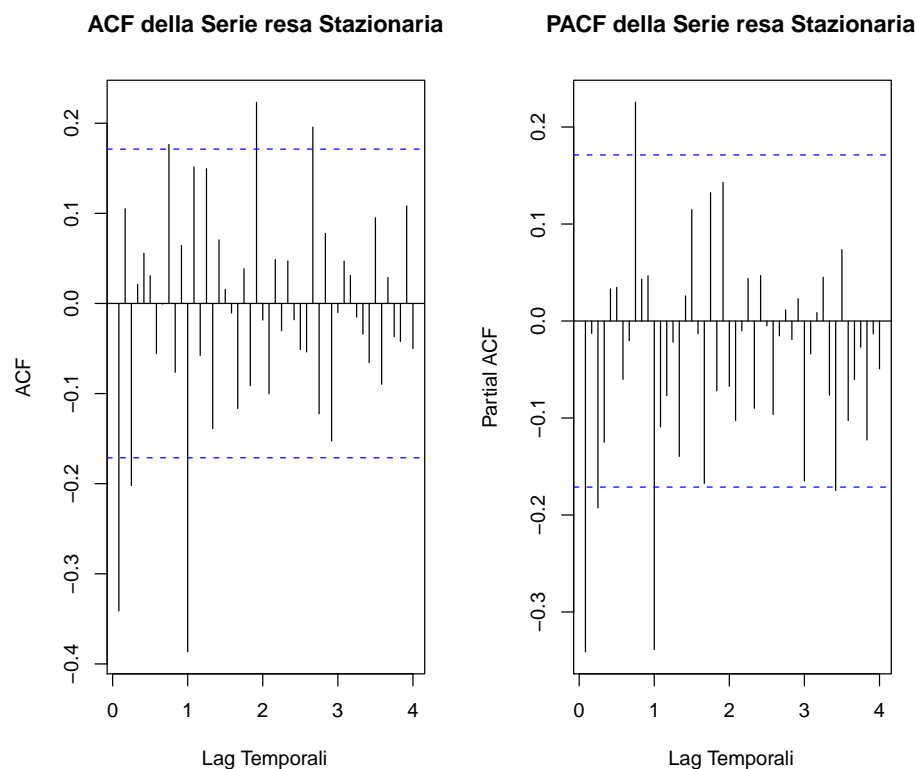


Figura 6.5: Grafico della funzione di autocorrelazione globale e della funzione di autocorrelazione parziale della serie resa stazionaria.

Dall'analisi del grafico della funzione di autocorrelazione globale osserviamo che presenta un picco significativamente diverso da zero in corrispondenza del primo ritardo, per quanto riguarda la parte regolare, e in corrispondenza del primo ritardo stagionale, per quanto concerne la parte stagionale. Possiamo assumere, quindi, che il modello che ha generato la serie oggetto della nostra analisi abbia una componente media mobile regolare di ordine 1 e una componente media mobile stagionale di ordine 1, ossia $q = Q = 1$.

Analizzando il grafico della funzione di autocorrelazione parziale osserviamo che ha un andamento simile a quello della funzione di autocorrelazione globale. Il modello generatore della serie avrà, pertanto, una componente autoregressiva regolare e stagionale di ordine 1, ossia $p = P = 1$.

Riassumendo, il modello stocastico identificato per $\{Y_t\} = \{\log X_t\}$ è un processo $ARIMA(1, 1, 1)(1, 1, 1)_{12}$, che ha la seguente espressione:

$$\Phi_1(B^{12})\phi_1(B)(1-B^{12})(1-B)Y_t = \Theta_1(B^{12})\theta_1(B)Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim GWN(0, \sigma^2).$$

III PASSO Applicando la procedura basata sul metodo iterativo della massima verosimiglianza, proposto da Box e Jenkins, il modello stimato risulta il seguente:

$$(1 - 0.168B)(1 + 0.099B^{12})(1 - B)(1 - B^{12})Y_t = (1 + 0.563B)(1 + 0.497B^{12})Z_t,$$

$$(Z_t)_{t \in \mathbb{Z}} \sim GWN(0, \sigma^2) \text{ dove } \sigma^2 = 0.0013^1.$$

IV PASSO Il primo passo che dobbiamo effettuare riguarda le stime dei suoi coefficienti, attraverso la verifica delle seguenti ipotesi nulle: $H_0 : \phi = 0$, $H_0 : \Phi = 0$, $H_0 : \theta = 0$ e $H_0 : \Theta = 0$. Utilizzando il test t di Student, le prime due ipotesi non vengono rifiutate dal momento che le stime dei parametri autoregressivi sono non significativamente diversi da zero, quindi i parametri stimati sono non rilevanti.

Come previsto dalla procedura di Box e Jenkins dobbiamo rieffettuare l'identificazione del modello. Poiché il test t porta a non rifiutare le ipotesi nulle per i parametri della componente autoregressiva e a rifiutare le ipotesi nulle per i parametri della componente media mobile, identifichiamo come modello generatore un $ARIMA(0, 1, 1)(0, 1, 1)_{12}$, ossia poniamo gli ordini $p = P = 0$.

¹Le rispettive deviazioni standard dei coefficienti stimati sono 0.246, 0.154, 0.211 e 0.136. Queste ultime saranno utilizzate nel passo successivo per verificare la significatività dei parametri controllando che ogni singola stima in valore assoluto sia almeno due volte superiore alla sua deviazione standard.

6.2 Un esempio di applicazione della procedura di Box e Jenkins 139

A questo punto dobbiamo ripetere la stima dei parametri, come nel terzo passo. Utilizzando nuovamente il metodo di massima verosimiglianza, il modello stimato risulta:

$$(1 - B)(1 - B^{12})Y_t = (1 + 0.402B)(1 + 0.557B^{12})Z_t, \quad (Z_t)_{t \in \mathbb{Z}} \sim GWN(0, \sigma^2)$$

dove $\sigma^2 = 0.0013^2$. Adesso analizziamo la bontà di questo modello verificando le seguenti ipotesi nulle: $H_0 : \theta = 0$ e $H_0 : \Theta = 0$. Utilizzando il test t di Student, entrambe le ipotesi vengono rifiutate dal momento che le stime dei parametri risultano in valore assoluto maggiori di due volte la rispettiva deviazione standard. Il test t usato porta, dunque, a ritenere che le due stime siano significativamente diverse da zero.

Osserviamo che il modello stimato soddisfa le condizioni di invertibilità dal momento che $|\theta| < 1$ e $|\Theta| < 1$. Inoltre, le radici dei due operatori media mobile, regolare e stagionale, sono reali e lontani dall'unità assicurando l'assenza di ridondanza dei parametri.

Dobbiamo, infine, controllare il coefficiente di correlazione tra le stime dei due parametri, θ e Θ , che nel nostro caso è un valore molto basso -0.1107058 , che assicura la stabilità del modello stimato.

Un'ulteriore verifica che dobbiamo effettuare riguarda l'analisi dei residui del modello ARIMA stimato. Dobbiamo, innanzitutto, analizzare il grafico dei residui per valutare che essi risultino stazionari e, successivamente, il loro grafico a dispersione per poter valutare che siano incorrelati. Questi due grafici sono riportati in Figura 6.6.

²Le rispettive deviazioni standard dei coefficienti sono 0.090 e 0.073

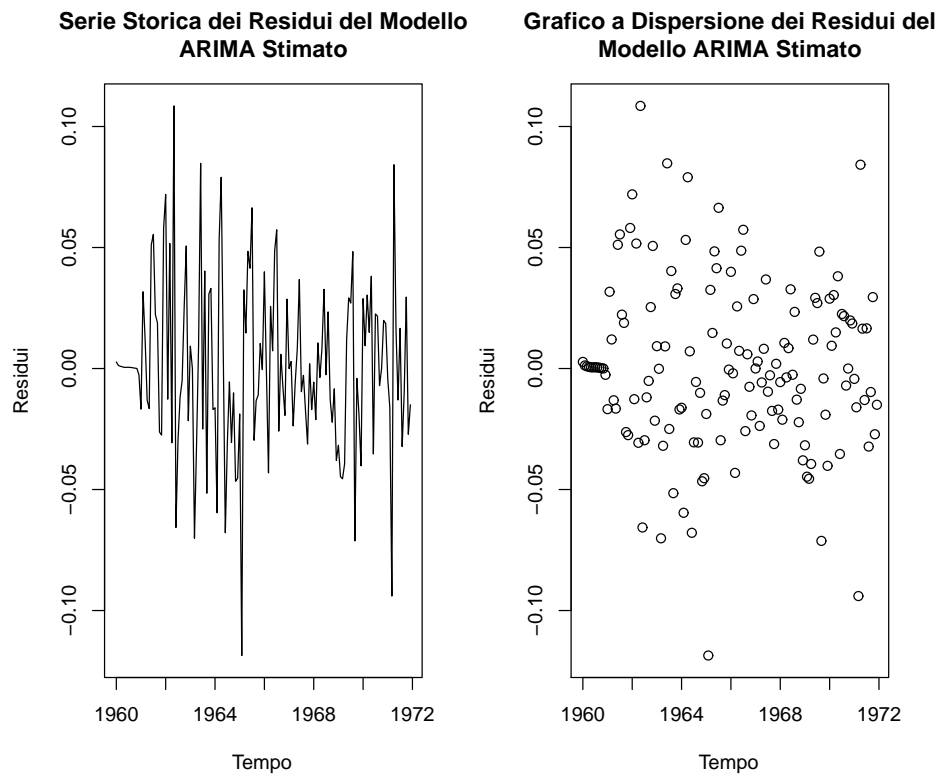


Figura 6.6: Grafici dei residui del modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato sulla serie storica oggetto di studio.

6.2 Un esempio di applicazione della procedura di Box e Jenkins 141

Dall'analisi del grafico della serie storica dei residui osserviamo che questa risulta stazionaria dal momento che fluttua attorno ad un valore costante, lo zero. Mentre dall'analisi del grafico a dispersione possiamo dire che i residui sono incorrelati dato che non presentano una particolare struttura.

Per poter valutare, invece, la capacità del modello stimato di descrivere la dipendenza temporale della serie storica analizzata, dobbiamo studiare l'andamento della funzione di autocorrelazione globale dei residui. Un modello che riesce a cogliere le dipendenze della serie storica presenterà, infatti, una funzione di autocorrelazione globale contenuta nella regione di confidenza, ossia una funzione che sarà non significativamente diversa da zero in corrispondenza di ogni ritardo. Riportiamo in Figura 6.7 il grafico della funzione di autocorrelazione globale dei residui e il grafico dell'andamento del p-value associato al test di Ljung-Box.

In particolare, il primo grafico mostra che le autocorrelazioni dei residui sono non significativamente diverse da zero, portando ad accettare l'ipotesi nulla secondo cui i residui del modello provengono da un processo rumore bianco. Il secondo grafico, invece, rappresenta il valore del p-value associato alla statistica test di Ljung-Box al variare di K , vista nel Paragrafo 6.1.4. Dato che il p-value associato risulta superiore al livello di significatività $\alpha = 0.05$, allora possiamo non rifiutare l'ipotesi nulla congiunta al variare di K .

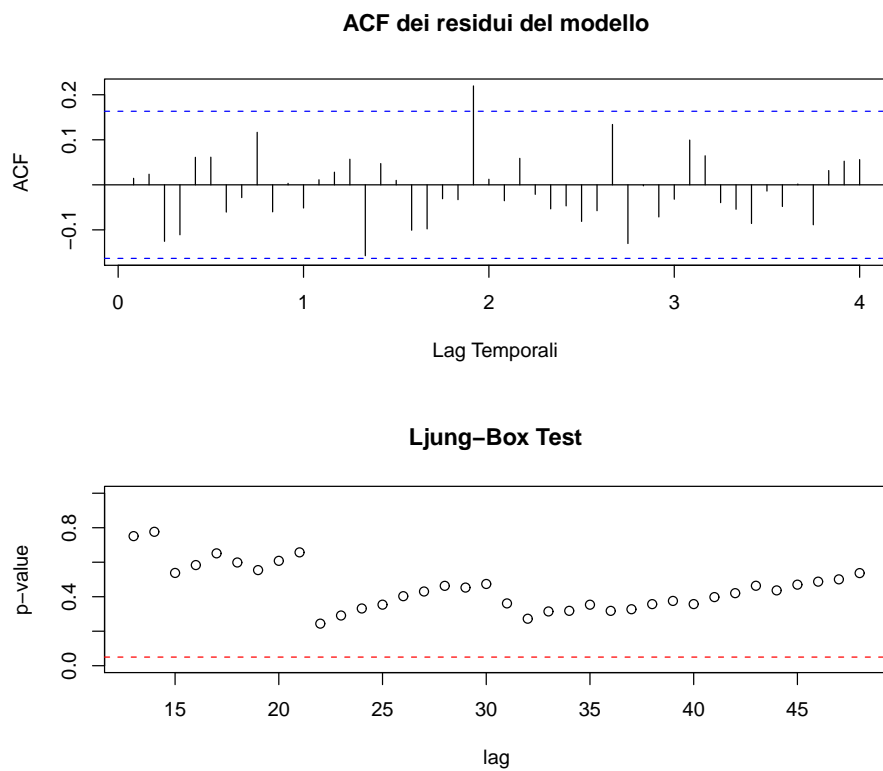


Figura 6.7: Grafici relativi alla funzione di autocorrelazione dei residui del modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato.

6.2 Un esempio di applicazione della procedura di Box e Jenkins 143

Per verificare, infine, che i residui siano normalmente distribuiti usiamo l'istogramma dei residui, che deve risultare distribuito secondo la legge Gaussiana, e il grafico Quantile-Quantile (Q-Q plot), che fornisce una misura della bontà di adattamento della distribuzione dei residui alla distribuzione gaussiana. In quest'ultimo grafico riportiamo in rosso la retta che indica la situazione di perfetta coincidenza tra i quantili della normale standard e quelli della distribuzione dei residui. L'istogramma dei residui e il grafico Quantile-Quantile sono riportati in Figura 6.8.

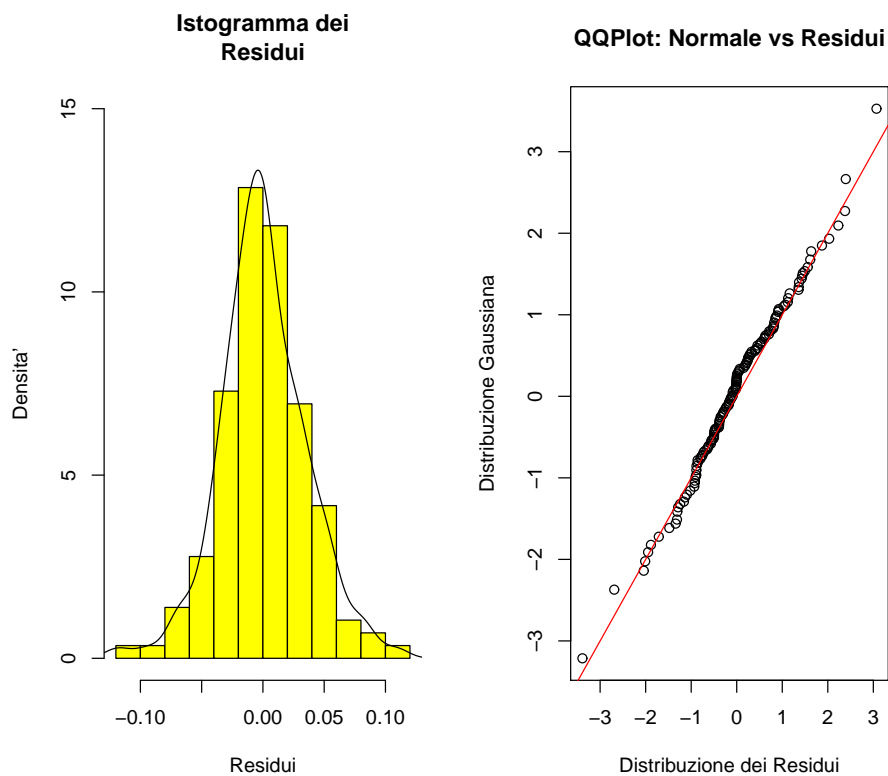


Figura 6.8: Grafici relativi alla valutazione del tipo di distribuzione dei residui del modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato.

Possiamo, inoltre, calcolare il test di Kolmogorov-Smirnov che saggia l'ipotesi di normalità distributiva dei residui standardizzati e nel seguito riportiamo l'output ottenuto:

$D = 0.0822$, $p\text{-value} = 0.3633$

Un altro test più potente del precedente per la verifica della normalità dei residui è il test di Shapiro-Wilk, di cui riportiamo l'output:

$W = 0.9864$, $p\text{-value} = 0.1674$

Analizzando l'istogramma dei residui, il grafico Quantile-Quantile e l'esito dei test appena calcolati possiamo dire che i residui risultano normalmente distribuiti.

V PASSO L'ultimo controllo che dobbiamo effettuare per poter accettare ed utilizzare il modello è valutare la sua capacità previsiva. Scegliamo di prevedere i valori degli ultimi 12 mesi di osservazione ossia del 1971, usando il modello identificato e stimato tralasciando le osservazioni che vogliamo prevedere. Otteniamo i seguenti valori:

6.038649 5.988762 6.145428 6.118993 6.159657 6.304670 6.433296
6.445969 6.266723 6.136196 6.007904 6.114341.

Ricordiamo che abbiamo applicato, inizialmente, la trasformazione logaritmica e quindi la previsione risulta essere per la serie $Y_t = \log X_t$. Per ottenere una previsione dei dati $\{X_t\}$, dobbiamo ritrasformare i valori precedenti mediante una trasformazione esponenziale. Otteniamo:

419.3261 398.9204 466.5793 454.4069 473.2657 547.1210 622.2214
630.1570 526.7484 462.2917 406.6301 452.2979.

Le stime ottenute non si discostano troppo dalle osservazioni iniziali, infatti, possiamo osservare come in corrispondenza della settima e ottava osservazione il numero di passeggeri sia più alto, in accordo con i dati osservati.

Il grafico delle osservazioni $\{Y_t\}$ con gli intervalli di confidenza per le previsioni effettuate, che riportiamo in Figura 6.9, mostra che la serie delle previsioni effettuate è contenuta nella regione di confidenza e presenta, a meno del terzo valore previsto, un andamento coerente con quella delle corrispondenti

osservazioni reali.

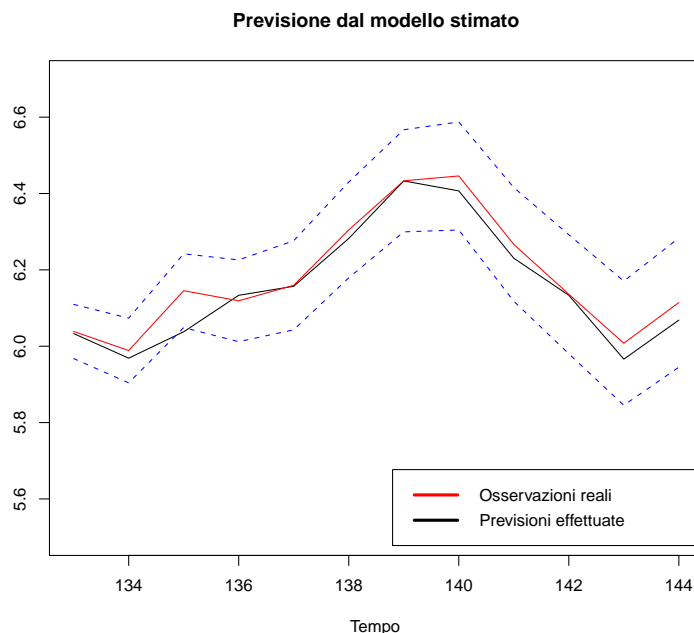


Figura 6.9: Grafico delle ultime 12 osservazioni della serie previste mediante il modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ stimato.

Una misura sintetica della capacità previsiva del modello stimato è il MAPE, che nel nostro caso risulta $MAPE = 0.462$, ossia il modello identificato rivela una buona capacità previsiva.

Il modello $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ identificato e stimato costituisce, dunque, una buona rappresentazione della nostra serie storica *airpass*, pertanto, possiamo utilizzarlo per fare previsioni future.

Utilizziamo, ora, il modello stimato per effettuare la previsione per l'anno 1972, di cui riportiamo i valori ottenuti:

450.4224 425.7172 479.0068 492.4045 509.0550 583.3449 670.0108
667.0776 558.1894 497.2078 429.8720 477.2426.

Dal grafico riportato in Figura 6.10 possiamo notare che la previsione effettuata segue le caratteristiche della serie originale $\{X_t\}$.

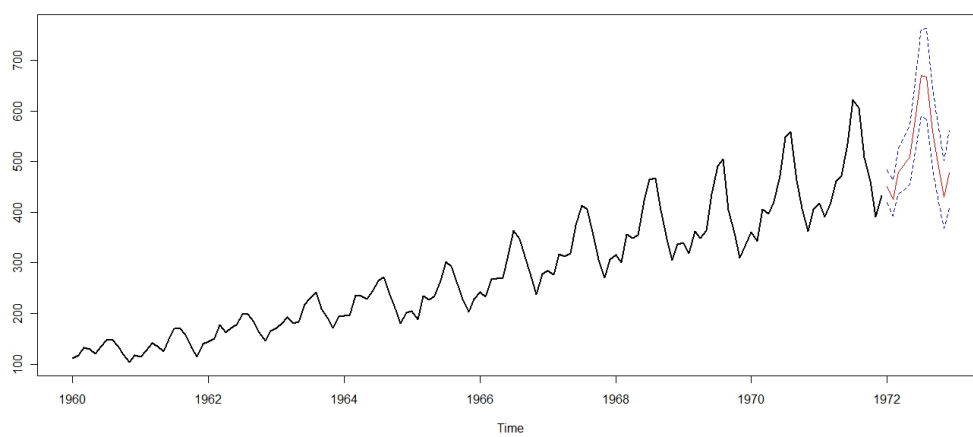


Figura 6.10: Grafico della serie storica airpass con le previsioni per l'anno 1972.

Appendice A

Alcuni richiami sugli spazi di Hilbert

Definizione A.1. Sia X uno spazio vettoriale su \mathbb{C} . Sia data un'applicazione:

$$\begin{aligned}\langle \cdot, \cdot \rangle : X \times X &\rightarrow \mathbb{C} \\ (x, y) &\longmapsto \langle x, y \rangle\end{aligned}$$

tale che

1. $\langle x, y \rangle \rightarrow \langle x, y \rangle$ è lineare;
2. $\overline{\langle x, y \rangle} = \langle y, x \rangle$ in \mathbb{C} ;
3. $\langle x, x \rangle = \overline{\langle x, x \rangle} > 0$ se $x \neq 0$.

Quest'applicazione è detta *prodotto scalare*.

Osservazione A.1. Sia $x \in X$ e sia $\langle \cdot, \cdot \rangle$ un prodotto scalare definito su questo spazio. Allora $\|x\| := \sqrt{\langle x, x \rangle}$ è una norma.

Definizione A.2. Uno *spazio di Hilbert* \mathcal{H} è uno spazio vettoriale su cui è definito un prodotto scalare che induce una norma $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ che rende lo spazio completo, ossia che ogni successione di Cauchy converge in norma a un elemento $x \in \mathcal{H}$.

Definizione A.3. Un sottospazio \mathcal{M} di uno spazio di Hilbert \mathcal{H} è detto *sottospazio chiuso* di \mathcal{H} se data una successione in \mathcal{M} questa converge ad un elemento di \mathcal{M} .

Definizione A.4. Definiamo il sottospazio *ortogonale* di un sottoinsieme \mathcal{M} di \mathcal{H} come l'insieme \mathcal{M}^\perp di tutti gli elementi di \mathcal{H} ortogonali rispetto a ogni elemento di \mathcal{M} . Quindi

$$x \in \mathcal{M}^\perp \Leftrightarrow \langle x, y \rangle = 0, \quad \forall y \in \mathcal{M}.$$

Teorema A.1 (Teorema della proiezione ortogonale). *Sia \mathcal{M} un sottospazio chiuso di uno spazio di Hilbert \mathcal{H} e sia $x \in \mathcal{H}$, allora*

- esiste un unico elemento $\hat{x} \in \mathcal{M}$, detto *proiezione ortogonale di x su \mathcal{M}* , tale che

$$\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|,$$

- $\hat{x} \in \mathcal{M}$ e $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$ se e solo se $\hat{x} \in \mathcal{M}$ e $(x - \hat{x}) \in \mathcal{M}^\perp$.

Poniamo $\hat{x} := P_{\mathcal{M}}(x)$.

Di seguito riportiamo alcune proprietà della proiezione ortogonale.

Proposizione A.2. *Sia \mathcal{H} uno spazio di Hilbert e sia $P_{\mathcal{M}}$ la proiezione ortogonale sul sottospazio chiuso \mathcal{M} . Allora valgono:*

1. $P_{\mathcal{M}}(\alpha x + \beta y) = \alpha P_{\mathcal{M}}x + \beta P_{\mathcal{M}}y, \quad x, y \in \mathcal{H}, \quad \alpha, \beta \in \mathbb{C};$
2. $\|x\|^2 = \|P_{\mathcal{M}}x\|^2 + \|(I - P_{\mathcal{M}})x\|^2;$
3. ogni $x \in \mathcal{H}$ ha un'unica rappresentazione come somma di un elemento di \mathcal{M} e un elemento di \mathcal{M}^\perp , ossia

$$x = P_{\mathcal{M}}x + (I - P_{\mathcal{M}})x;$$

4. $P_{\mathcal{M}}x_n \rightarrow P_{\mathcal{M}}x$ se $\|x_n - x\| \rightarrow 0;$
5. $x \in \mathcal{M}$ se e solo se $P_{\mathcal{M}}x = x;$

6. $x \in \mathcal{M}^\perp$ se e solo se $P_{\mathcal{M}}x = 0$;

7. $\mathcal{M}_1 \subseteq \mathcal{M}_2$ se e solo se $P_{\mathcal{M}_1}P_{\mathcal{M}_2}x = P_{\mathcal{M}_1}x$ per ogni $x \in \mathcal{H}$.

Definizione A.5. Definiamo il *sottospazio chiuso* generato da $\{x_t, t \in T\}$, che indicheremo con $\overline{\text{sp}}\{x_t, t \in T\}$, come il più piccolo sottospazio chiuso di \mathcal{H} che contiene tutti gli elementi x_t , con $t \in T$.

Osservazione A.2. Il sottospazio chiuso generato da un insieme finito $\{x_1, \dots, x_n\}$ è l'insieme di tutte le combinazioni lineari, $y = \alpha_1x_1 + \dots + \alpha_nx_n$, con $\alpha_1, \dots, \alpha_n$ in \mathbb{C} .

Osservazione A.3. Sia $\mathcal{M} = \overline{\text{sp}}\{x_1, \dots, x_n\}$. Allora per $x \in \mathcal{H}$, la proiezione $P_{\mathcal{M}}x$ è l'unico elemento della forma

$$P_{\mathcal{M}}x = \alpha_1x_1 + \dots + \alpha_nx_n,$$

tale che

$$\langle x - P_{\mathcal{M}}x, y \rangle = 0, \quad y \in \mathcal{M},$$

o equivalentemente

$$\langle P_{\mathcal{M}}x, x_j \rangle = \langle x, x_j \rangle, \quad j = 1, \dots, n. \quad (\text{A.1})$$

Possiamo riscrivere l'equazione (A.1) come un sistema di equazioni lineari

$$\sum_{i=1}^n \alpha_i \langle x_i, x_j \rangle = \langle x, x_j \rangle, \quad j = 1, \dots, n.$$

Appendice B

I processi ad incrementi ortogonali

Prima di dare la definizione di processo stocastico ad incrementi ortogonali, ricordiamo la seguente definizione.

Definizione B.1. Un processo $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ è detto *continuo a destra* (in L^2) se per tutti i $\lambda \in [-\pi, \pi)$,

$$\|Z(\lambda + \delta) - Z(\lambda)\|^2 = E[Z(\lambda + \delta) - Z(\lambda)]^2 \rightarrow 0 \quad \text{per } \delta \searrow 0.$$

Definizione B.2. Un *processo stocastico ad incrementi ortogonali* su $[-\pi, \pi]$ è un processo stocastico a valori complessi continuo a destra $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ tale che

- $\langle Z(\lambda), Z(\lambda) \rangle < \infty$, per $-\pi \leq \lambda \leq \pi$;
- $\langle Z(\lambda), 1 \rangle = 0$, per $-\pi \leq \lambda \leq \pi$;
- $\langle Z(\lambda_2) - Z(\lambda_1), Z(\lambda_4) - Z(\lambda_3) \rangle = 0$, per $(\lambda_1, \lambda_2] \cap (\lambda_3, \lambda_4] = \emptyset$;

dove il prodotto scalare è definito come $\langle X, Y \rangle = E[X\bar{Y}]$.

Proposizione B.1. Se $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ è un processo ad incrementi ortogonali, allora esiste un'unica funzione di ripartizione F tale che

- $F(\lambda) = 0, \quad \lambda \leq -\pi,$
- $F(\lambda) = F(\pi), \quad \lambda \geq \pi,$
- $F(\mu) - F(\lambda) = \|Z(\mu) - Z(\lambda)\|^2, \quad -\pi \leq \lambda \leq \mu \leq \pi.$

La funzione F è detta *funzione di ripartizione associata al processo* $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$.

Esempio 28. Il moto Browniano $\{B(\lambda), -\pi \leq \lambda \leq \pi\}$ con $E[B(\lambda)] = 0$ e $Var(B(\lambda)) = \sigma^2 \frac{(\lambda+\pi)}{2\pi}$, $-\pi \leq \lambda \leq \pi$, è un processo ad incrementi ortogonali su $[-\pi, \pi]$. La sua funzione di ripartizione vale

$$F(\lambda) = \begin{cases} 0 & \text{per } \lambda \leq -\pi, \\ \sigma^2 \frac{(\lambda+\pi)}{2\pi} & \text{per } -\pi \leq \lambda \leq \pi, \\ \sigma^2 & \text{per } \lambda \geq \pi. \end{cases}$$

Esempio 29. Se $\{N(\lambda), -\pi \leq \lambda \leq \pi\}$ è un processo di Poisson su $[-\pi, \pi]$ con intensità costante c , allora il processo $Z(\lambda) = N(\lambda) - E[N(\lambda)]$, con $-\pi \leq \lambda \leq \pi$, è un processo ad incrementi ortogonali, la cui funzione di ripartizione vale

$$F(\lambda) = \begin{cases} 0 & \text{per } \lambda \leq -\pi, \\ c(\lambda + \pi) & \text{per } -\pi \leq \lambda \leq \pi, \\ 2\pi c & \text{per } \lambda \geq \pi. \end{cases}$$

Se scegliamo $c = \frac{\sigma^2}{2\pi}$, allora il processo di Poisson ha la stessa funzione di ripartizione del moto Browniano dell'esempio precedente.

Vogliamo adesso dare la definizione di integrale stocastico

$$I(f) = \int_{(-\pi, \pi]} f(\nu) dZ(\nu),$$

dove $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ è un processo stocastico ad incrementi ortogonali e f è una funzione su $[-\pi, \pi]$ di quadrato integrabile rispetto alla funzione

di ripartizione F associata al processo $Z(\lambda)$.

Denotiamo con \mathcal{D} l'insieme delle funzioni f della forma

$$f(\lambda) = \sum_{i=0}^n f_i I_{(\lambda_i, \lambda_{i+1}]}(\lambda), \quad -\pi = \lambda_0 < \lambda_1 < \dots < \lambda_{n+1} = \pi, \quad n \in \mathbb{N}. \quad (\text{B.1})$$

Se $f \in \mathcal{D}$, definiamo $I(f)$ come

$$I(f) = \sum_{i=0}^n f_i [Z(\lambda_{i+1}) - Z(\lambda_i)]. \quad (\text{B.2})$$

Si dimostra che l'applicazione I è ben definita su \mathcal{D} . Estendiamo adesso questa applicazione ad un isomorfismo da $L^2(F)$ su un sottospazio di $L^2(P)$. Denotiamo con $\overline{\mathcal{D}}$ la chiusura di \mathcal{D} in $L^2(F)$. Se $f \in \overline{\mathcal{D}}$ allora esiste una sequenza (f_n) di elementi di \mathcal{D} tale che $\|f_n - f\|_{L^2(F)} \rightarrow 0$. Allora possiamo definire

$$I(f) = \lim_{n \rightarrow \infty} I(f_n) \quad \text{in } L^2(F).$$

Si dimostra che questo limite esiste ed è lo stesso per ogni successione (f_n) tale che $\|f_n - f\|_{L^2(F)} \rightarrow 0$. Utilizzando la definizione di $I(f)$ su $\overline{\mathcal{D}}$ e ricordando che $\overline{\mathcal{D}} = L^2(F)$, possiamo dare la seguente definizione di integrale stocastico rispetto ad un processo ad incrementi ortogonali.

Definizione B.3. Sia $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ un processo ad incrementi ortogonali con funzione di ripartizione F . Se $f \in L^2(F)$, allora l'integrale stocastico $\int_{(-\pi, \pi]} f(\lambda) dZ(\lambda)$ è la variabile aleatoria $I(f)$ definita precedentemente, ossia

$$\int_{(-\pi, \pi]} f(\lambda) dZ(\lambda) := I(f).$$

L'integrale stocastico così definito è lineare e tale che

- $E[I(f)\overline{I(g)}] = \int_{(-\pi, \pi]} f(\nu)\overline{g(\nu)}dF(\nu)$ con $f, g \in L^2(F)$;
- $E[I(f)] = 0$ con $f \in L^2(F)$.

Da queste proprietà, discende la proposizione seguente.

Proposizione B.2. *Sia $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$ un processo stocastico ad incrementi ortogonali con funzione di ripartizione F , allora*

$$X_t = I(e^{it\cdot}) = \int_{(-\pi, \pi]} e^{it\nu} dZ(\nu), \quad t \in \mathbb{Z},$$

è un processo stazionario con media nulla e funzione di autocovarianza data da

$$E[X_{t+h}\bar{X}_t] = \int_{(-\pi, \pi]} e^{i\nu h} dF(\nu).$$

Appendice C

Alcuni richiami sulla serie di Fourier

Consideriamo lo spazio di Hilbert $L^2([-\pi, \pi])$ dove il prodotto scalare di $f, g \in L^2([-\pi, \pi])$ è definito come

$$\langle f, g \rangle = E[f\bar{g}] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

Definiamo le funzioni $\{e_n, n \in \mathbb{Z}\}$ come

$$e_n(x) = e^{inx}.$$

Esse sono ortonormali in $L^2([-\pi, \pi])$: infatti

$$\begin{aligned} \langle e_m, e_n \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)x} dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [\cos(m-n)x + i \sin(m-n)x] dx \\ &= \begin{cases} 1 & \text{se } m = n \\ 0 & \text{se } m \neq n. \end{cases} \end{aligned}$$

Definizione C.1. L'approssimazione di Fourier di ordine n di una funzione $f \in L^2([-\pi, \pi])$ è definita come la proiezione di f su $\overline{\text{sp}}\{e_j, |j| \leq n\}$, ossia

$$S_n f = \sum_{j=-n}^n \langle f, e_j \rangle e_j. \tag{C.1}$$

I coefficienti

$$\langle f, e_j \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ijx} dx$$

sono detti coefficienti di Fourier della funzione f .

Possiamo riscrivere la (C.1) in maniera più esplicita:

$$S_n f(x) = \sum_{j=-n}^n \langle f, e_j \rangle e^{ijx} \quad \text{per } x \in [-\pi, \pi].$$

Valgono i seguenti risultati (si vedano i Paragrafi 2.8 e 2.11 di [7]).

Teorema C.1. *La successione $\{S_n f\}$ converge in L^2 verso f per $n \rightarrow \infty$.*

Teorema C.2. *Sia f una funzione continua su $[-\pi, \pi]$ tale che $f(\pi) = f(-\pi)$. Allora*

$$\frac{1}{n} (S_0 f + S_1 f + \cdots + S_{n-1} f) \rightarrow f$$

uniformemente su $[-\pi, \pi]$ per $n \rightarrow \infty$.

Teorema C.3. *Se $f \in L^2([-\pi, \pi])$ e $\langle f, e_j \rangle = 0$ per ogni $j \in \mathbb{Z}$, allora $f = 0$ quasi ovunque.*

Appendice D

Codice in R

Riportiamo di seguito il codice scritto nel programma R utilizzato per analizzare la serie storica descritta nel Paragrafo 6.2.

```
library(TSA)
# Caricamento e grafico della serie storica
data(airpass)
serie<- airpass
tserie <- ts(serie, start=c(1960,1), end=c(1971,12), freq=12)
is.ts(tserie)
plot(tserie,type=1,lwd=2,col=red,xlab=Tempo,ylab=Passeggeri
     delle linee aeree,main=Grafico Serie Storica)
```

I PASSO.

```
#Realizzazione del grafico range mean plot
medieann <- vector()
devst <- vector()
k <- 1
for(i in seq(1, length(tserie), by=12)){
  medieann[k] <- mean(tserie[i:(i+11)])
  devst[k] <- sd(tserie[i:(i+11)])
```

```

    k <- k+1
  }
  o <- order(medieann)
  plot(medieann[o], devst[o], type=p, ylim=c(0,100), main=Range
    Mean Plot,xlab=valori medi, ylab=deviazioni standard)
  lm.tserie <- lm(devst  medieann)
  lm.tserie
  lm.tserie$coef
  abline(a=lm.tserie$coef[1], b=lm.tserie$coef[2], col=red)
  #Trasformazione logartmica
  lserie=log(serie)
  #Grafico della funzione di autocorrelazione globale
  series<- as.ts(lserie)
  acf(series,lag.max=48,main=ACF della serie storica,
    xlab=Lag Temporal)
  #Eliminazione del trend e della stagionalità
  series.diff.fin <- diff(diff(series, lag=12, differences=1),
    differences=1)

```

II PASSO.

```

#Identificazione del modello
par(mfrow=c(1,2))
acf(series.diff.fin, lag.max=48, main=ACF della Serie resa
  Stazionaria, xlab=Lag Temporal)
pacf(series.diff.fin, lag.max=48, main=PACF della Serie resa
  Stazionaria, xlab=Lag Temporal)

```

III PASSO.

```

#Stima del modello ARIMA(1, 1, 1)(1, 1, 1)12
fit <- arima(series, order=c(1,1,1), seasonal=list
  (order=c(1,1,1), period=12), method=ML)

```

IV PASSO.

```
#Valutazione del modello stimato: analisi dei coefficienti
abs(fit$coef[1])>2*sqrt(fit$var.coef[1,1])
abs(fit$coef[2])>2*sqrt(fit$var.coef[2,2])
abs(fit$coef[3])>2*sqrt(fit$var.coef[3,3])
abs(fit$coef[4])>2*sqrt(fit$var.coef[4,4])
#Stima del modello ARIMA(0,1,1)(0,1,1)12
fit <- arima(series, order=c(0,1,1), seasonal=list
  (order=c(0,1,1), period=12), method=ML)
fit
#Valutazione del modello stimato: analisi dei coefficienti
abs(fit$coef[1])>2*sqrt(fit$var.coef[1,1])
abs(fit$coef[2])>2*sqrt(fit$var.coef[2,2])
rho12 <- fit$var.coef[1,2]/sqrt(fit$var.coef[1,1]*fit$
  var.coef[2,2])
rho12
#Valutazione della bontà del modello stimato: analisi dei residui
par(mfrow=c(1,2))
plot(fit$residuals, type=1, main=Serie Storica dei Residui
  del Modello ARIMA Stimato, xlab=Tempo, ylab=Residui)
plot(fit$residuals, type=p, main=Grafico a Dispersione dei
  Residui del Modello ARIMA Stimato, xlab=Tempo, ylab=Residui)
# Test di Ljung-Box
acf(fit$residuals, lag.max=48, main=ACF dei residui del
  modello, xlab=Lag Temporal)
LBQPlot(fit$residuals, lag.max = 48, StartLag = k,
  k = 0, SquaredQ = FALSE)
#Grafico Q-Q plot
par(mfrow=c(1,2))
hist(fit$residuals, col=yellow, freq=F, ylim=c(0,15),
  main=Istogramma dei Residui, xlab=Residui, ylab=Densità )
```

```
lines(density(fit$residuals))
qqplot(scale(fit$residuals,T,T), rnorm(1000), main=QQPlot:
  Normale vs Residui, xlab=Distribuzione dei Residui,
  ylab=Distribuzione Gaussiana)
abline(0,1,col=red)
#Test Kolmogorov-Smirnov
ks<-ks.test(scale(fit$residuals,T,T),rnorm(1000))
ks
#Shapiro-Wilk
sw<-shapiro.test(scale(fit$residuals,T,T))
sw
```

V PASSO.

```
#Valutazione della capacità previsiva del modello
pred.fit <- arima(series[1:132], order=c(0,1,1),
  seasonal=list(order=c(0,1,1), period=12), method=ML)
pred.model <- predict(pred.fit, n.ahead=12)
U <- pred.model$pred+1.96*pred.model$se
L <- pred.model$pred-1.96*pred.model$se
ts.plot(series[133:144], pred.model$pred, col=1:2, ylim=
  c(5.5,6.7), xlab=Tempo, main=Previsione dal modello stimato)
lines(U, col=blue, lty=dashed)
lines(L, col=blue, lty=dashed)
legend(x=bottomright, legend=c(Osservazioni reali,Previsioni
  effettuate), col=c(red,black),ncol=1,lty=1,lwd=3,
  inset=c(0.01,0.02),bg=white)
series.fin <- series[133:144]
ser <- sum(abs((series.fin-pred.model$pred[1:12])/
  series.fin))
MAPE <- ser*100/12
```

Bibliografia

- [1] Ash, R.B., M.F. Gardner, *Topics in stochastic processes*, 1975, Probability and Mathematical Statistics, vol. 27.
- [2] Bartlett, M.S., *The Use of Transformations*, 1947, Biometrics, Vol. 3, pp. 39-52.
- [3] Battaglia, F., *Metodi di previsione statistica*, 2007, Springer-Verlag Italia.
- [4] Bee Dagum, E., *Analisi delle serie storiche. Modellistica, previsione e scomposizione*, 2002, Springer-Verlag Italia.
- [5] Box, G.E.P., D.R. Cox, *An Analysis of Transformations*, 1964, Journal of the Royal Statistical Society, Series B, Vol. 26, pp. 211-252.
- [6] Box, George E. P., Gwilym M. Jenkins, *Time series analysis: forecasting and control*, 1970, Holden-Day.
- [7] Brockwell, Peter J., Richard A. Davis, *Time series: theory and methods*, 1991, seconda edizione, Springer.
- [8] Chung, K.L., *A course in probability theory*, 1974, seconda edizione, Academic Press.
- [9] Cox, D.R., H.D. Miller, *The theory of stochastic processes*, 1977, Science Paperbacks.

- [10] Di Lascio, F. Marta L., Marilena Pillati, *Dispensa di Analisi di Serie Storiche*, A.A. 2009/2010.
- [11] Fuller, W.A., *Introduction to statistical time series*, 1996, Wiley Series in Probability and Statistics.
- [12] Hamilton, James D., *Time Series Analysis*, 1994, Princeton University Press.
- [13] Pintacuda, N., *Probabilità* , 1994, Zanichelli.
- [14] Shumway, Robert H., David S. Stoffer, *Time series analysis and its applications*, 2006, Springer.
- [15] Wei, W.S., *Time Series Analysis*, 1990, Addison-Wesley Longman.