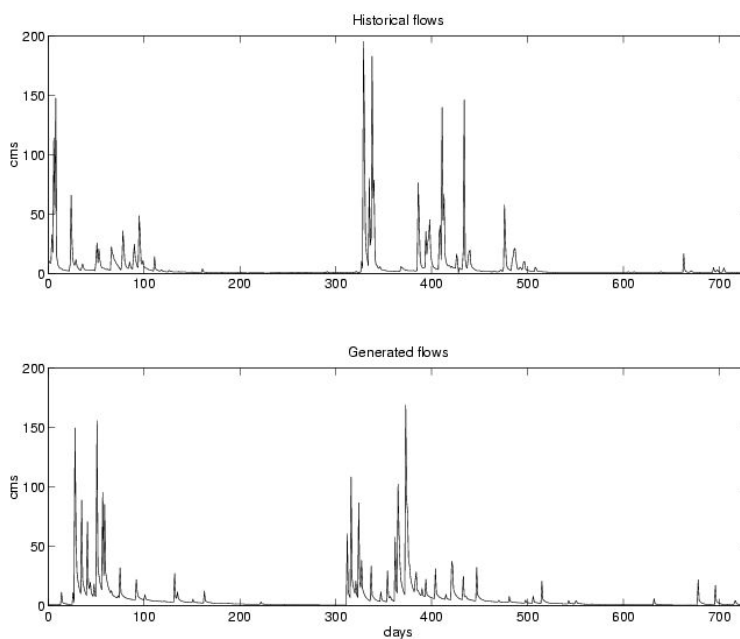


SERIE STORICHE, PROCESSI E MODELLI STOCASTICI PER L'IDROLOGIA E LA GESTIONE DELLE RISORSE IDRICHE

Pierluigi Claps

DITIC – POLITECNICO DI TORINO
[claps@polito.it]



1. INTRODUZIONE

Si parla di serie storiche (o serie temporali) quando si considera un fenomeno in relazione alla sua evoluzione nel tempo. L'osservazione quotidiana dei fenomeni temporali ne offre infiniti esempi: si pensi ad esempio all'andamento dei prezzi di certi beni, specie in periodo di economia fortemente dinamica, o all'altezza del mare in un dato luogo.

La registrazione del fenomeno, posta su un grafico per evidenziarne la dinamica temporale, potrebbe, in linea di principio, essere considerata come una successione di dati costituenti un insieme campionario per la cui investigazione sono disponibili strumenti statistici adeguati. Tuttavia, la statistica matematica, e l'inferenza in particolare, si sviluppa per lo più su dati non connessi temporalmente; ciò spiega il motivo per cui sono stati ricercati ulteriori strumenti, i processi stocastici, per effettuare l'analisi delle serie storiche.

E' bene chiarire il significato di serie storica, processo e modello:

- la serie storica è una collezione di numeri reali, ordinati secondo la variabile tempo, la quale costituisce una parte finita di una realizzazione di un processo stocastico.
- per processo stocastico, a parametro discreto, si intende una successione di variabili casuali la cui completa conoscenza è assicurata solo dalla conoscenza della famiglia delle ripartizioni finite.
- un modello stocastico costituisce una parametrizzazione di un processo in termini di una funzione esplicita di parametri noti. Mentre un processo stocastico è noto oppure no, un modello può essere stimato a partire dai dati, ovvero dalla serie storica osservata

Finalità dello studio di una serie storica è quello di descrivere ed interpretare il fenomeno fisico da essa rappresentato in modo da poter effettuare delle previsioni sulla dinamica temporale del fenomeno stesso, o delle generazioni di serie sintetiche con estensione molto maggiore del periodo di osservazione

Dal punto di vista statistico questa indagine è di tipo inferenziale in quanto bisogna risalire da un campione (nel caso in esame temporale) ad un modello teorico il quale sia in grado di generare una serie sintetica o di effettuare operazioni di previsione dell'evoluzione del fenomeno temporale.

L'indagine statistica sulle serie temporali si articola essenzialmente in tre fasi:

- 1) identificazione del modello teorico;
- 2) stima dei parametri del modello;
- 3) controllo della bontà di adattamento del modello ai dati.

Molto importante è la fase di identificazione del modello che descrive ed interpreta la serie temporale. Questa operazione richiede un'approfondita conoscenza dei processi stocastici e molta esperienza. Infatti, l'individuazione univoca di un modello univoco avviene solamente in casi particolari (es. quando si dispone di informazioni a priori sul fenomeno). Nell'usuale approccio statistico, la tecnica usata per l'identificazione dei modelli consente di ridurre l'indeterminazione ad un numero limitato (due o tre) fra i quali si procede alla selezione del modello definitivo.

Per usuale approccio statistico intendiamo, ormai, quello proposto da Box e Jenkins (1970), i quali hanno proposto un metodo di analisi nel quale sia la serie storica ad orientare verso il modello e non viceversa (*series speaking for themselves*), evidenziando, così, che eventuali conoscenze a priori sulla serie temporale che si intende esaminare potrebbero portare all'identificazione di modelli non ottimali sotto il profilo della bontà di adattamento. Si vedrà, più avanti, che in ambito idrologico esistono invece molti punti a favore di modelli selezionati sulla

base di informazioni a priori.

2. SERIE TEMPORALI E PROCESSI STOCASTICI

2.1. Alcuni esempi di serie temporali osservate

Qualunque fenomeno venga considerato in relazione al tempo t dà luogo ad una serie temporale. Gli esempi che si possono fare sono, pertanto, innumerevoli. Importanti sono i fenomeni appartenenti al mondo economico, alcuni dei quali sono seguiti con particolare interesse allo scopo di individuare tendenze che possano permettere di capire l'evoluzione del fenomeno stesso. Si pensi ad esempio alle serie temporali relative ai prezzi di diverse merci e servizi, alle importazioni ed alle esportazioni, alla bilancia dei pagamenti etc. Possiamo ritrovare esempi di serie temporali anche in altri campi quali l'economia aziendale, la produzione industriale, per la quale l'analisi delle serie storiche ha costituito il punto di partenza per lo studio del "controllo statistico di qualità", la meteorologia.

Si possono schematizzare diverse categorie di serie temporali, riferibili ad altrettante categorie di processi stocastici:

- processi a fenomeno discreto ed a parametro discreto (n. di giorni piovosi in un mese)
- processi a fenomeno discreto ed a parametro continuo (particelle radioattive registrate da un contatore geiger)
- processi a fenomeno continuo ed a parametro discreto (caso più diffuso)
- processi a fenomeno continuo ed a parametro continuo (elettroencefalogramma)

Le serie temporali di interesse idrologico riguardano essenzialmente la successione dei deflussi misurati alle stazioni idrometriche, ed, eventualmente, quella degli afflussi meteorici, misurata ai pluviometri. Si riferiscono quindi al terzo dei casi sopra enunciati. Nelle figure 2.1 e 2.2 sono riportati andamenti cronologici tipici dei due fenomeni.

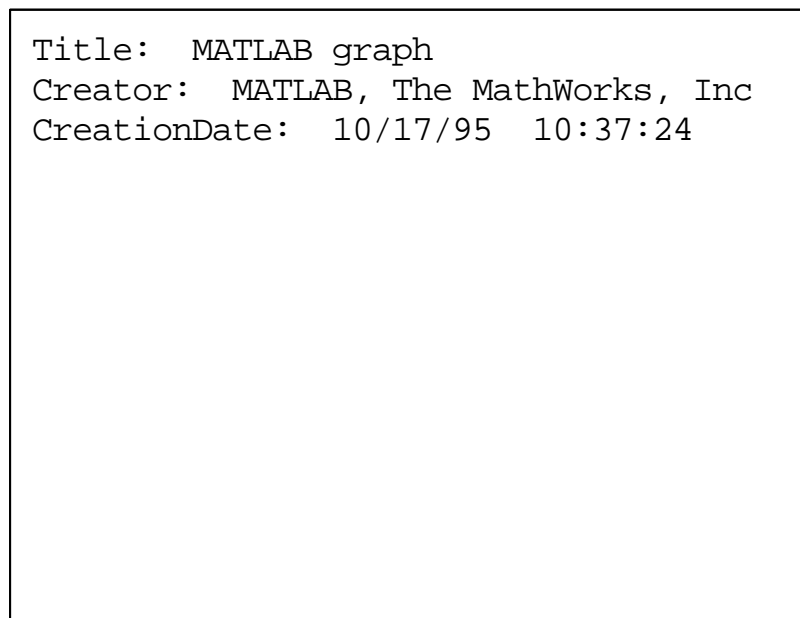


Fig. 2.1. Serie di deflussi mensili.

```

Title:  MATLAB graph
Creator:  MATLAB, The MathWorks, Inc
CreationDate:  10/17/95  10:54:28

```

Fig. 2.2. Serie di precipitazioni mensili.

2.2. Le serie temporali continue e discrete

Consideriamo una grandezza z la quale vari nel tempo t , quale, ad esempio, la temperatura in un dato luogo. Possiamo trovarci in presenza di due diverse situazioni:

- 1) la grandezza z è funzione continua del tempo t ed in tal caso la indicheremo con $z(t)$. In questa ipotesi il diagramma rappresentativo dell'andamento temporale del fenomeno si presenta come indicato nella figura 2.3.a e la serie temporale è detta *continua*;
- 2) la grandezza z viene rilevata solamente in certi istanti $t=0, 1, 2, 3, \dots$. In questa ipotesi il diagramma rappresentativo dell'andamento temporale del fenomeno si presenta come indicato nella figura 2.3.b e la serie temporale è detta *discreta*.

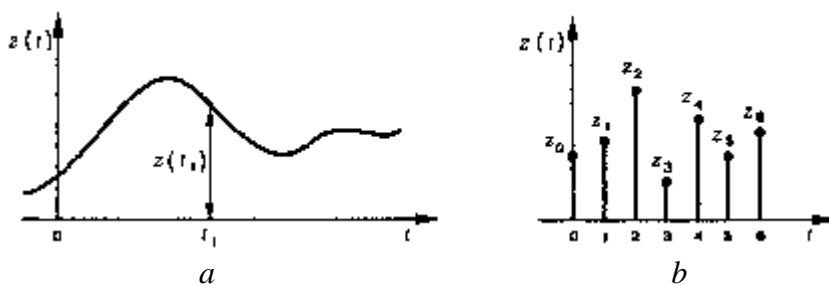


Fig. 2.3. Serie temporale continua (a) e discreta (b).

2.3. Serie deterministiche e serie probabilistiche

Quando si studia una serie temporale ci si può trovare in presenza di due comportamenti, concettualmente ben distinti, indicati, rispettivamente, come *andamento deterministico* ed *andamento stocastico* o probabilistico.

Si dice che un andamento temporale è di tipo deterministico quando si può prevedere il suo sviluppo futuro senza errore. Esempi di andamenti temporali di tipo deterministico sono riportati

nella figura 2.4. Tali andamenti temporali sono rappresentati da funzioni matematiche che consentono di determinare, in modo certo, il loro valore in qualunque istante t .

Molti comportamenti temporali sono invece caratterizzati da andamenti erratici, con oscillazioni irregolari di segno positivo e negativo. Sono queste le caratteristiche degli andamenti di tipo stocastico il cui nome deriva dal fatto che essi si possono spiegare come manifestazioni di eventi casuali. Rispetto al problema della previsione, questi fenomeni si presentano in modo esattamente opposto a quelli deterministici, in quanto la previsione delle loro future manifestazioni è sempre affetta da errori.

E' opportuno ricordare, però, che gli eventi casuali non sono prevedibili nelle loro singole manifestazioni, ma si possono fare previsioni (in termini probabilistici) quando si considerano "masse di casi". L'esame statistico delle serie temporali di tipo non deterministico può fornire utili informazioni tali da consentire di scoprire il "meccanismo" che governa la serie stessa consentendo, così, una previsione abbastanza precisa del futuro andamento del fenomeno temporale.

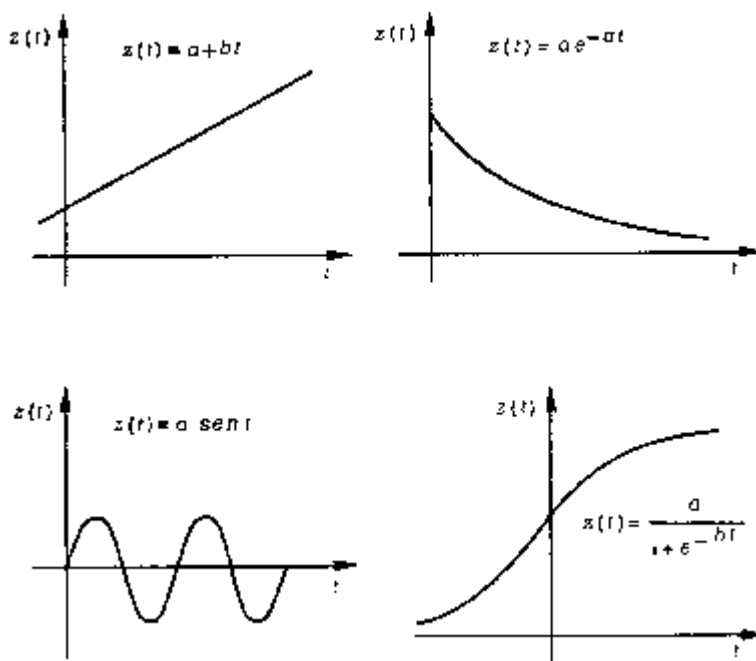


Fig. 2.4. Esempi di serie temporali deterministiche

La gran parte degli andamenti temporali che è possibile osservare sembrano ottenuti dalla sovrapposizione di un andamento deterministico con uno probabilistico (vedi figura 2.5). In questo caso si parla di andamento misto che si suppone sia ottenuto dalla somma di due andamenti elementari.

Il modello che rappresenta un andamento temporale misto può essere scritto, in generale, come segue:

$$z(t) = f(t) + a(t)$$

dove:

- $f(t)$ è la componente deterministica del modello;
- $a(t)$ è la componente probabilistica del modello;
- $z(t)$ è l'andamento risultante, cioè la serie temporale generata dal modello.

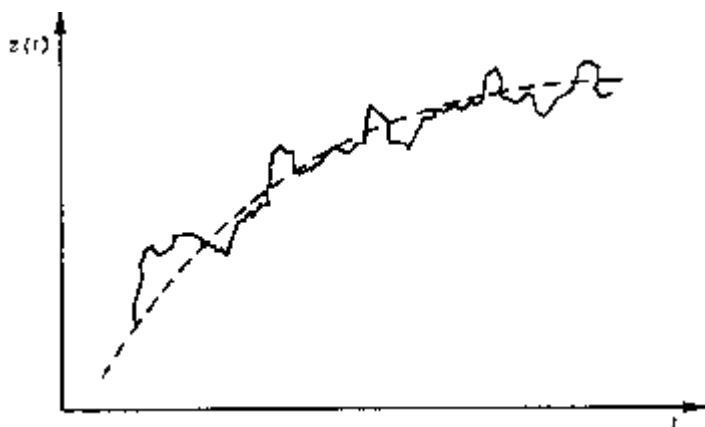


Fig. 2.5. Esempio di serie temporale con andamento misto

2.4 Le componenti di una serie temporale

Il concetto di composizione di più andamenti di tipo diverso costituisce un importante aspetto della teoria delle serie temporali. Quando si osservano sequenze temporali si riscontrano evoluzioni che sembrano non avere caratteristiche di similitudine. Ad un'analisi più attenta, però, ci si accorge che ci sono dei comportamenti elementari che si ripetono frequentemente sia da soli, sia accompagnati ad altri. In una serie temporale empirica z_t si possono osservare i seguenti andamenti elementari, tutti funzione del tempo t :

- 1) un *trend* (tendenza) che indicheremo con T_t
- 2) una *componente ciclica* che indicheremo con C_t
- 3) una *componente stagionale* che indicheremo con S_t
- 4) un *residuo* o componente casuale a_t

Con il nome di trend si intende un andamento non stazionario privo di irregolarità accidentali, che si evidenzia quando si considerano serie lunghe. Il trend viene generalmente rappresentato con una funzione matematica di tipo semplice, come, ad esempio, un polinomio o una funzione esponenziale nel tempo. L'andamento temporale riportato nella figura 2.6 è un chiaro esempio di trend.

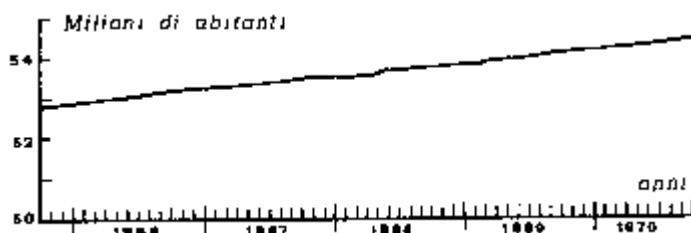


Fig. 2.6. Esempio di trend in una serie temporale

La gran parte delle serie temporali relative a fenomeni economici, sociali e soprattutto meteorologici, contengono una componente ciclica, più o meno regolare, con un periodo prossimo ai 12 mesi. Questa componente, che causa effetti che tendono a ripetersi in tempi corrispondenti, si chiama componente stagionale e si evidenzia quando si esaminano serie con intervallo di rilevazione inferiori all'anno (serie mensili, settimanali, etc.). Un esempio evidente della componente ciclica in una serie temporale è riportata nella figura 2.7. La componente

stagionale è, in generale, la più agevole da riconoscere in quanto la periodicità di tipo regolare che la caratterizza è evidenziata dal semplice esame del grafico.

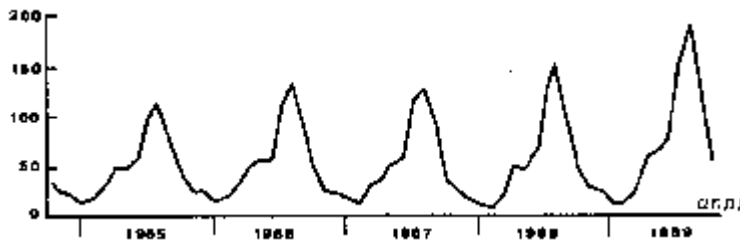


Fig. 2.7. Serie temporale con presenza di componente stagionale

In tutte le serie temporali che non siano di tipo deterministico sono sempre presenti delle irregolarità, di segno positivo e negativo rispetto alla media del fenomeno, che sembrano prodotte da un comportamento di tipo aleatorio. Tale componente si chiama casuale (o accidentale) ed assomma in sè gli elementi di incertezza (o aleatorietà) che caratterizzano fenomeni naturali, sociali, economici etc. L'andamento temporale mostrato nella figura 2.8 evidenzia la presenza della componente accidentale nella serie storica rappresentata.

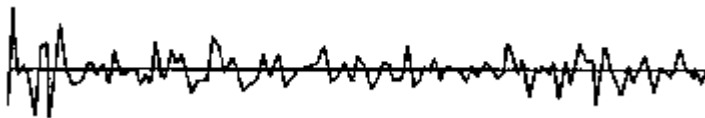


Fig. 2.8. Componente casuale di una serie temporale

Le serie temporali osservate o reali sono date dalle successioni delle misurazioni dei valori assunti da un certo fenomeno. E' però possibile costruire artificialmente delle serie temporali mediante l'uso di appositi metodi matematici. La simulazione di serie sintetiche consente di valutare le prestazioni del modello stocastico utilizzato, mediante confronto tra serie generate ed osservate. Consente inoltre di effettuare previsioni, sia dirette, a breve termine, che probabilistiche, intese come valutazioni del livello di criticità di una certa configurazione dei dati idrologici o di un sistema idrico sottoposto a lunghe sequenze di input simulati.

3. PROCESSI STOCASTICI E LORO CARATTERIZZAZIONE

3.1 Definizioni

Un processo stocastico Z_t è una famiglia di variabili casuali descritte da un parametro t appartenente ad un insieme parametrico T .

Nei fenomeni di tipo idrologico le rilevazioni avvengono ad intervalli equispaziati. I processi stocastici che prenderemo in considerazione sono quindi definiti da successioni di v.c. continue $Z_t, t \in T$ e $T = \{1, 2, 3, \dots\}$. Tali processi sono detti processi stocastici continui a parametro discreto.

Consideriamo, ad esempio, la successione dei deflussi annui misurati in una stazione idrometrica:

$$S = \{dt\} \quad t=1, 2, 3, \dots$$

Per t fissato, dt è una variabile casuale continua; il parametro t del processo stocastico è un parametro discreto ed il processo è, quindi, continuo a parametro discreto.

Per la conoscenza di un processo stocastico continuo a parametro discreto del tipo $\{Z_1, Z_2, \dots\}$ occorre specificare le funzioni di densità di probabilità di ciascuna combinazione di esse. Più precisamente, un processo continuo a parametro discreto è noto se, contemporaneamente:

- 1) sono note le funzioni di probabilità univariate $f(Z_1), f(Z_2), \dots$ per ciascuna v.c. $\{Z_1, Z_2, \dots\}$ componente il processo;
- 2) sono note le funzioni di densità bivariate $f_{1,2}(Z_1, Z_2), f_{1,3}(Z_1, Z_3), \dots, f_{n-1,n}(Z_{n-1}, Z_n)$ per ciascuna coppia ordinata di v.c. $(Z_1, Z_2), (Z_1, Z_3), \dots, (Z_{n-1}, Z_n)$;
- 3) sono note le funzioni di densità trivariate $f_{1,2,3}(Z_1, Z_2, Z_3), f_{1,2,4}(Z_1, Z_2, Z_4), \dots$ per ciascuna terna ordinata di v.c. $(Z_1, Z_2, Z_3), (Z_1, Z_2, Z_4), \dots$; etc...

In generale, un processo stocastico Z_t è noto se è nota la funzione di densità multivariata della k -pla di v.c. $(Z_{t1}, Z_{t2}, \dots, Z_{tk})$ per ogni k e per ogni k -pla (t_1, t_2, \dots, t_k) di v.c.

La definizione stessa di processo stocastico fa già intuire l'estrema complicazione dello studio di un tale tipo di processi nella loro generalità e, soprattutto, la pratica impossibilità di inferire direttamente su di essi.

Supponiamo che un processo stocastico Z_t rappresenti il risultato di un esperimento in successivi istanti di tempo. Se nel processo Z_t fissiamo la prova da effettuare, otteniamo una successione di risultati campionari $\{Z_1, Z_2, \dots\}$, funzioni della variabile t , chiamata realizzazione o traiettoria del processo. E' evidente che, dato un processo Z_t , esiste un'infinità di possibili realizzazioni che sono, precisamente, tutte quelle che si possono osservare ripetendo indefinitamente l'esperimento: nella figura 3.1 sono illustrate alcune di esse per un ipotetico processo Z_t .

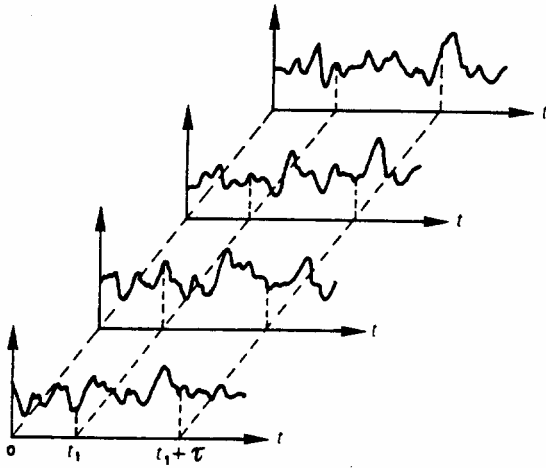


Fig. 3.1. Esempi di possibili realizzazioni di un processo stocastico

Dalle definizioni date in precedenza possiamo trarre una nuova definizione di serie storica $\{z_t\}_{t=1,2,3,\dots,N}$ intesa come una parte finita di una realizzazione di un processo stocastico (finita in quanto le N osservazioni sono solo una parte delle possibili osservazioni del processo).

Una tale definizione di serie storica, però, esplicita anche la limitazione delle informazioni sul processo che sono, in generale, desumibili dalla conoscenza della serie storica. Difatti, la serie storica è una parte finita di una singola realizzazione del processo; quindi, non solo costituisce un campione unico della famiglia delle v.c. che caratterizzano il processo, ma è anche un campione troncato perché si osserva solo per $t=1,2,\dots,N$ istanti di tempo.

Tutto ciò impone, quindi, una limitazione della classe dei processi stocastici su cui è possibile inferire statisticamente, perché solo in una classe più ristretta sarà possibile dedurre informazioni consistenti dalle realizzazioni finite che si conoscono nelle applicazioni reali.

3.2 Momenti di un processo stocastico

I processi stocastici, per loro stessa definizione, sono in grado di generare una serie temporale illimitata, cioè di lunghezza infinita. Poiché, però, non è pensabile di fornire le caratteristiche dei processi stocastici mediante le serie temporali da essi generate, è necessario riassumere le loro principali proprietà mediante poche grandezze caratteristiche: i momenti teorici. Questi sono infiniti in numero ma, in pratica, se ne utilizzano pochi e precisamente:

- il momento del primo ordine, cioè il valor medio (o valore atteso) del processo;
- i momenti del secondo ordine, costituiti dalla covarianza e dalle grandezze ad essa collegate.

Sia, allora, $\{Z_t\}$ un processo stocastico continuo a parametro discreto. Vediamo quali sono le espressioni dei momenti più usati.

Valore medio atteso

Il valore medio teorico, indicato con $\mu(t)$, è espresso dalla relazione:

$$\mathbf{m}(t) = E(Z_t) = \int_{-\infty}^{+\infty} u f_{Z_t}(u) du \quad (3.1)$$

dove il simbolo E indica il valore atteso. In pratica, in base alla (3.1), il valore atteso del processo stocastico al tempo t è pari al valore atteso della v.c. Z_t definita all'istante t . Nel caso del tutto generale espresso dalla (3.1), il valor medio del processo è funzione del tempo t . Ovviamente, il valore atteso $E(Z_t)$ sarebbe noto se fosse nota la pdf della v.c. Z_t .

Varianza (teorica)

In Statistica, data una v.c. Z_t , la varianza viene definita come "il valor medio dei quadrati degli scarti fra la variabile Z e la sua media teorica $\mu(t)$ ". La varianza di un processo stocastico, indicata con $\sigma^2(t)$, è definita dalla relazione seguente:

$$\mathbf{s}^2(t) = E[Z_t - \mathbf{m}(t)]^2 = \text{var}(Z_t) = \int_{-\infty}^{+\infty} [u - \mathbf{m}(t)]^2 f_{Z_t}(u) du \quad (3.2)$$

ed è, in generale, funzione del tempo.

Come si rileva dalla (3.2) la varianza è un momento del secondo ordine e serve ad avere una misura della dispersione media dei dati attorno alla loro media. Spesso, al posto della varianza, se ne usa la radice quadrata che si chiama scarto quadratico medio e si indica con \mathbf{s} .

Autocovarianza (teorica)

Si considerino due variabili casuali X ed Y . In Statistica quando si studiano le relazioni che intercorrono fra due v.c. si usa la covarianza, indicata con \mathbf{g}_{xy} , definita come "valor medio atteso del prodotto degli scarti di X dalla sua media μ_x per gli scarti di Y dalla sua media μ_y ":

$$\mathbf{g}_{xy} = E[(X - \mathbf{m})(Y - \mathbf{m})] = \text{cov}(XY) \quad (3.3)$$

La covarianza, a differenza della varianza, può essere positiva o negativa. E' positiva quando valori crescenti [decrescanti] di X si associano con valori crescenti [decrescanti] di Y . E' negativa quando valori crescenti [decrescanti] di X si associano con valori decrescanti [crescenti] di Y . La covarianza misura, quindi, la tendenza di X ed Y a variare nello stesso senso.

La (3.3), applicata ad una serie temporale, realizzazione di un processo stocastico, ne definisce l'autocovarianza. Infatti, in una serie temporale, la grandezza è sempre la stessa, ma è riferita a due istanti diversi t e $t+k$. Pertanto le v.c. sono Z_t e Z_{t+k} e la relazione che definisce l'autocovarianza del processo stocastico fra gli istanti t e $t+k$ è:

$$\mathbf{g}(t, t+k) = E[(Z_t - \mathbf{m}(t))(Z_{t+k} - \mathbf{m}(t+k))] = \text{cov}(Z_t, Z_{t+k}) \quad (3.4)$$

La varianza è un caso particolare della covarianza: se, infatti, nella (3.4) si pone $k=0$, si ha

$$\mathbf{g}(t, t) = E[(Z_t - \mathbf{m}(t))(Z_t - \mathbf{m}(t))] = E[Z_t - \mathbf{m}(t)]^2 = \text{var}(Z_t) = \mathbf{s}^2(t) \quad (3.5)$$

Autocorrelazione (teorica)

L'autocovarianza, pur essendo una grandezza fondamentale per lo studio delle serie temporali, ha il difetto di non essere compresa fra limiti fissi per cui è difficile giudicare il significato di un dato valore. Per questo motivo è stata introdotta l'autocorrelazione che si ricava immediatamente dall'autocovarianza e che presenta il vantaggio di essere compresa fra i limiti fissi -1 e +1.

L'autocorrelazione si ottiene semplicemente dividendo l'autocovarianza per il prodotto degli scarti quadratici medi di Z_t e Z_{t+k} . La relazione che definisce l'autocorrelazione è, quindi:

$$\mathbf{r}_k(t) = \frac{E[(z_t - \mathbf{m}(t))(z_{t+k} - \mathbf{m}(t+k))]}{\sqrt{E[z_t - \mathbf{m}(t)]^2 E[z_{t+k} - \mathbf{m}(t+k)]^2}} = \frac{\text{cov}(z_t, z_{t+k})}{\mathbf{S}(z_t)\mathbf{S}(z_{t+k})} \quad (3.6)$$

L'autocorrelazione è sempre compresa fra -1 e $+1$; quando $\mathbf{r}=0$, Z_t e Z_{t+k} non sono correlati.

Prima di concludere il paragrafo è opportuno notare che tutte le grandezze finora considerate risultano funzione del tempo t . E' questo un grosso vincolo, sia sul piano teorico che nelle applicazioni concrete. Il problema si aggira introducendo un'ipotesi semplificatrice, molto importante, che va sotto il nome di *stazionarietà*.

3.3 Le ipotesi di stazionarietà e di invertibilità

Nello studio delle serie temporali si incontra molto spesso un'ipotesi semplificatrice, quasi sempre necessaria, che va sotto il nome di *stazionarietà*. Tale ipotesi suppone che certe proprietà statistiche di una serie risultino invarianti rispetto ad una traslazione nel tempo. In altri termini, la stazionarietà suppone che certe proprietà non varino nel tempo. La stazionarietà può riguardare tutti i momenti oppure solamente alcuni. Esistono, pertanto, due tipi di stazionarietà che vengono comunemente indicati come:

- *stazionarietà completa* o totale;
- *stazionarietà ridotta* o debole.

Un processo stocastico $\{Z_t\}$ si dice che è completamente stazionario quando la distribuzione congiunta della n-pla di v.c. $Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}$ è uguale alla distribuzione congiunta della n-pla $Z_{t_1+T}, Z_{t_2+T}, \dots, Z_{t_n+T}$ per tutti i valori di T . Ciò significa che se si fa scorrere il tempo di una quantità arbitraria T , tutte le distribuzioni congiunte non si modificano. E', comunque, opportuno notare che l'ipotesi di stazionarietà completa è una condizione ideale, irraggiungibile nella pratica. Pertanto ci si accontenta, spesso, di stazionarietà più deboli o ridotte. Normalmente ci si limita ad ipotizzare una stazionarietà del secondo ordine, ossia una stazionarietà che investe la media (momento del primo ordine) ed i momenti del secondo ordine.

Oltre all'ipotesi di stazionarietà si incontra molto frequentemente l'ipotesi di invertibilità che viene introdotta allo scopo di evitare la "molteplicità dei modelli". Essa indica la possibilità di esprimere un processo X_t tramite le v.c. del passato, ovvero tramite una funzione che collega X_t con le v.c. X_s del tempo $s < t$. Ciò avviene, ovviamente, in senso stocastico, cioè mettendo in conto un errore casuale e_t . In seguito sarà chiarito che esistono dei casi in cui ad uguali strutture statistiche corrispondono due o più modelli stocastici diversi; la condizione di invertibilità permette di individuare in modo univoco il modello utile all'analisi che si sta effettuando.

3.4 Momenti teorici di un processo stocastico a stazionarietà debole

L'ipotesi di stazionarietà debole comporta l'indipendenza del valore medio e della varianza del processo stocastico dal tempo t , ossia:

$$\mathbf{m}(t) = \mathbf{m} \quad (3.7)$$

$$\mathbf{S}^2(t) = \mathbf{S}^2 \quad (3.8)$$

L'ipotesi di stazionarietà debole riduce l'espressione dell'autocovarianza a quella molto più semplice:

$$\mathbf{g}_k = E[(Z_t - \mathbf{m})(Z_{t+k} - \mathbf{m})] = E[(Z_t - \mathbf{m})(Z_{t-k} - \mathbf{m})] \quad (3.9)$$

dalla quale si evince che l'autocovarianza fra due v.c. componenti il processo stocastico dipende solo dal lag k e non dal valore iniziale t del tempo. Si noti che, per $k=0$, la relazione (3.9) fornisce la varianza del processo.

Si mostra facilmente, inoltre, che:

$$\mathbf{g}_k = \mathbf{g}^{-k}$$

In conclusione, la stazionarietà debole richiede che:

- a) media e varianza del processo risultino costanti;
- b) l'autocovarianza fra le componenti il processo stocastico sia funzione solo del lag k .

Trova grande applicazione nello studio delle serie temporali la *matrice delle varianze-covarianze* o *matrice di dispersione* che riassume le varianze e le covarianze di un dato processo stocastico. Essa si indica generalmente con Γ ed è espressa dalla seguente matrice quadrata

$$\Gamma = \begin{bmatrix} \mathbf{g}_0 & \mathbf{g}_1 & \cdots & \mathbf{g}_{t-1} \\ \mathbf{g}_1 & \mathbf{g}_0 & \cdots & \mathbf{g}_{t-2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{g}_{t-1} & \mathbf{g}_{t-2} & \cdots & \mathbf{g}_0 \end{bmatrix} \quad (3.10)$$

Per l'ipotesi di stazionarietà debole, la matrice di dispersione è simmetrica e contiene gli stessi valori su tutta la diagonale principale e sulle diagonali parallele a questa. Si noti, infine, che la matrice delle autocovarianze, in un processo stazionario, è definita positiva e, quindi, i determinanti di tutti i minori principali sono positivi.

La semplificazione indotta sul calcolo dell'autocovarianza fra due istanti diversi del processo dall'ipotesi di stazionarietà debole si ripercuote sull'autocorrelazione la cui espressione diventa:

$$\mathbf{r}_k = \frac{\mathbf{g}_k}{\mathbf{g}_0} \quad (3.11)$$

da cui si evince facilmente che:

$$\mathbf{r}_0 = 1 \quad (3.12)$$

Tale risultato è del tutto logico in quanto l'autocorrelazione di una componente del processo stocastico rispetto a sè stessa non può che essere perfetta.

I coefficienti di autocorrelazione godono delle stesse proprietà dei coefficienti di correlazione ordinari e cioè:

$$\begin{aligned} -1 &\leq \mathbf{r}_k \leq 1 \\ \mathbf{r}_0 &= 1 \\ \mathbf{r}_{-k} &= \mathbf{r}_k \end{aligned} \quad (3.13)$$

Anche i coefficienti di autocorrelazione si possono riunire in una matrice quadrata R che assume la forma seguente:

$$R = \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \dots & \mathbf{r}_{t-1} \\ \mathbf{r}_1 & \mathbf{r}_0 & \dots & \mathbf{r}_{t-2} \\ \dots & \dots & \dots & \dots \\ \mathbf{r}_{t-1} & \mathbf{r}_{t-2} & \dots & \mathbf{r}_0 \end{bmatrix} \quad (3.14)$$

Per la (3.11) le due matrici delle autocovarianze e delle autocorrelazioni sono legate dalla relazione:

$$\mathbf{G} = \mathbf{S}_z^2 R \quad (3.15)$$

In base alla (3.15) si può dedurre che le informazioni fornite dalla funzione di autocorrelazione e dalla varianza del processo sono esattamente uguali a quelle fornite dalla funzione di autocovarianza.

Anche la matrice delle autocorrelazioni di un processo stocastico stazionario è definita positiva. Ciò dà luogo ad alcune conseguenze limitatrici sui valori che possono assumere i coefficienti di autocorrelazione.

Se consideriamo un processo stocastico stazionario e calcoliamo le autocorrelazioni teoriche \mathbf{r}_k per $k=0,1,2,3,\dots$ otteniamo delle coppie di valori del tipo (k, \mathbf{r}_k) che, tradotti in grafico, danno luogo ad un diagramma che prende nome di *correlogramma* di cui vengono forniti due esempi nella figura 3.2.

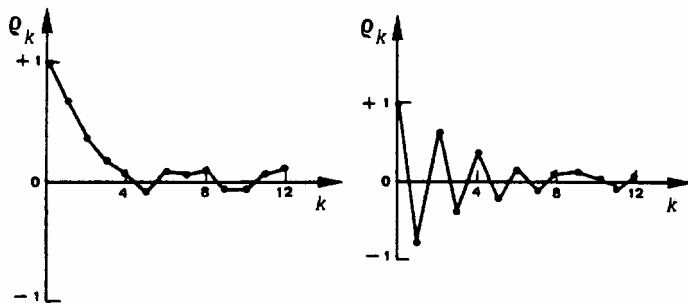


Fig. 3.2. Esempi di autocorrelogrammi di una serie storica.

Il correlogramma è uno strumento di notevole interesse pratico perché serve a discriminare un processo stocastico da un altro (è quindi uno degli strumenti utilizzati nella fase di identificazione del modello).

Oltre ai coefficienti di correlazione totale ρ_k , in un processo stocastico si possono calcolare anche dei coefficienti di correlazione parziali le cui espressioni sono alquanto semplici nel caso di processi stazionari.

Per definire in modo semplice i coefficienti di correlazione parziale consideriamo il caso di una grandezza x_1 che dipenda da più variabili x_2, x_3, \dots in modo lineare attraverso i coefficienti $\mathbf{a}_{12}, \mathbf{a}_{13}, \dots$ secondo la relazione:

$$x_1 = \mathbf{a}_{12}x_2 + \mathbf{a}_{13}x_3 + \dots$$

Si possono pertanto calcolare i coefficienti di correlazione totale fra la variabile dipendente x_1 e le altre singole variabili indipendenti, cioè $\mathbf{r}_{12}, \mathbf{r}_{13}, \mathbf{r}_{14}, \dots$ che, nel caso dei processi stocastici, rappresentano i coefficienti di correlazione totale.

E' però possibile calcolare, ad esempio, anche la correlazione fra le variabili x_1 ed x_3 nell'ipotesi che il valore x_2 sia costante, indicata con r_{13} . Nel caso delle serie temporali questi coefficienti si chiamano di autocorrelazione parziale e si indicano con F_{kk} .

Si può dimostrare che, nel caso di processo stocastico stazionario, questi coefficienti sono espressi dal rapporto:

$$\Phi_{kk} = \frac{R_k^*}{R_k} \quad (3.16)$$

in cui sia il numeratore che il denominatore sono i determinanti di due matrici quadrate definite come segue:

$$R_k = \det \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \dots & \mathbf{r}_{k-1} \\ \mathbf{r}_1 & \mathbf{r}_0 & \dots & \mathbf{r}_{k-2} \\ \dots & \dots & \dots & \dots \\ \mathbf{r}_{k-1} & \mathbf{r}_{k-2} & \dots & \mathbf{r}_0 \end{bmatrix} \quad (3.17)$$

$$R_k^* = \det \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \dots & \mathbf{r}_1 \\ \mathbf{r}_1 & \mathbf{r}_0 & \dots & \mathbf{r}_2 \\ \dots & \dots & \dots & \dots \\ \mathbf{r}_{k-1} & \mathbf{r}_{k-2} & \dots & \mathbf{r}_k \end{bmatrix} \quad (3.18)$$

L'espressione del numeratore R_k^* è uguale a quella di R_k con la sola differenza che l'ultima colonna viene sostituita dai coefficienti r_1, r_2, \dots, r_k :

3.5 Momenti campionari calcolati su una serie osservata

Come abbiamo osservato nel paragrafo precedente, i momenti teorici si riferiscono al processo stocastico che si suppone generi la serie temporale osservata. Gli stessi momenti, quando sono calcolati su quest'ultima, si chiamano *momenti campionari* e questo perchè una serie empirica può sempre essere considerata come un campione ricavato dal modello stocastico. Al solito, secondo le convenzioni della statistica, i momenti campionari vengono indicati con le lettere dell'alfabeto latino per distinguerli dai corrispondenti momenti teorici, indicati con le lettere dell'alfabeto greco.

E' importante precisare che, per il calcolo dei momenti campionari, noi supponiamo sempre di operare su una serie temporale discreta, di tipo stazionario e di ampiezza N. Ciò sarà chiarito alla fine del paragrafo.

Media aritmetica

Come è ben noto, la media aritmetica di n dati z_1, z_2, \dots, z_n è espressa dalla relazione:

$$M(z_i) = \frac{\sum z_i}{n} = \bar{z} \quad (3.19)$$

Va notato l'uso del simbolo M al posto del simbolo E (expected value) utilizzato nel paragrafo precedente.

Varianza campionaria

La varianza campionaria, intesa come stima non distorta (o corretta) della varianza del processo stocastico che ha generato la serie osservata è data dalla relazione:

$$s_*^2(z_t) = \frac{\sum (z_t - \bar{z})^2}{n-1} \quad (3.20)$$

Quando, però, n è grande, come di norma accade nelle serie temporali, la stima della varianza del processo stocastico, data dalla relazione:

$$s^2(z_t) = \frac{\sum (z_t - \bar{z})^2}{n} \quad (3.21)$$

risulta sostanzialmente non distorta.

Autocovarianza campionaria

La formula esatta per il calcolo dell'autocovarianza campionaria fra le osservazioni spaziate di k intervalli è data da:

$$c_k^* = \frac{1}{n-k} \sum_{t=1}^{n-k} (z_t - \bar{z}_1)(z_{t+k} - \bar{z}_2) \quad (3.22)$$

in cui

\bar{z}_1 è la media dei primi $n-k$ dati della serie;

\bar{z}_2 è la media degli ultimi $n-k$ dati della serie;

Una formula spesso usata al posto della (3.22) per il calcolo dell'autocovarianza campionaria, avente il vantaggio di non contenere i prodotti degli scarti è riportata di seguito:

$$c_k^* = \frac{1}{n-k} \sum z_t z_{t+k} - \bar{z}_1 \bar{z}_2 \quad (3.23)$$

Autocorrelazione campionaria

Analogamente all'autocorrelazione teorica si calcola l'autocorrelazione campionaria r_k definita dal rapporto:

$$r_k = \frac{c_k}{c_0} \quad (3.24)$$

Al solito, per $k=0$, si ha che $r_0=1$.

Correlogramma campionario

Il correlogramma campionario si costruisce analogamente al correlogramma teorico utilizzando i valori campionari r_0, r_1, r_2, \dots delle autocorrelazioni.

Autocorrelazione parziale campionaria

I coefficienti di autocorrelazione parziale campionaria si calcolano utilizzando la stessa formula valida per il processo teorico avendo cura di sostituire ai valori teorici r_k (ignoti) quelli campionari r_k . Si ha, dunque:

$$\Phi_{kk} = \frac{R_k^*}{R_k}, \quad (3.25)$$

dove (considerando ovviamente $r_0=1$):

$$R_k = \det \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \dots & \mathbf{r}_{k-1} \\ \mathbf{r}_1 & \mathbf{r}_0 & \dots & \mathbf{r}_{k-2} \\ \dots & \dots & \dots & \dots \\ \mathbf{r}_{k-1} & \mathbf{r}_{k-2} & \dots & \mathbf{r}_0 \end{bmatrix} \quad (3.26)$$

$$R_k^* = \det \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \dots & \mathbf{r}_1 \\ \mathbf{r}_1 & \mathbf{r}_0 & \dots & \mathbf{r}_2 \\ \dots & \dots & \dots & \dots \\ \mathbf{r}_{k-1} & \mathbf{r}_{k-2} & \dots & \mathbf{r}_k \end{bmatrix} \quad (3.27)$$

E' opportuno notare che, quando una serie temporale non è stazionaria, non è possibile utilizzare nessuna delle precedenti relazioni per calcolare i momenti campionari della serie. Infatti esse contengono tutte la media aritmetica della serie temporale che si suppone costante. Nelle serie non stazionarie, invece, la media varia con il tempo t per cui è necessario ricorrere ad opportuni metodi per l'individuazione del *trend* e/o della *stagionalità*.

3.6 L'ipotesi di ergodicità

Anticipando quanto sarà meglio specificato in seguito, osserviamo che i momenti campionari di una serie storica definiti nel precedente paragrafo vengono usati come stime corrette dei corrispondenti momenti teorici del processo stocastico stazionario.

Il fatto che si possa ottenere una stima (consistente) delle proprietà statistiche di un processo stocastico stazionario dallo studio di un solo campione temporale di lunghezza finita non è affatto ovvio, contrariamente a quanto potrebbe sembrare.

Per definire in termini quantitativi l'ipotesi di ergodicità, consideriamo, da un lato una singola manifestazione di un processo stocastico (figura 3.3) e, dall'altro, un insieme di manifestazioni campionarie derivate tutte dallo stesso processo (figura 3.4).



Fig. 3.3. Singola realizzazione campionaria derivante da un processo stocastico.

Con riferimento alla figura 3.3, consideriamo n punti temporali equidistanti ottenendo i valori $z_k(t_1), z_k(t_2), \dots$ dai quali possiamo ricavare la quantità:

$$\bar{z}(t) = \frac{\sum z_k(t_i)}{n} \quad (3.28)$$

che prende nome di *media temporale*.

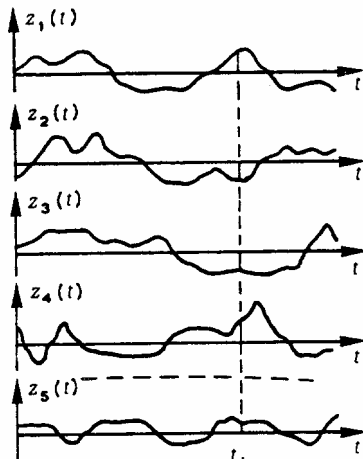


Fig 3.4. Insieme di realizzazioni campionarie provenienti dallo stesso modello stocastico

Con riferimento alla figura 3.4, possiamo considerare un dato istante temporale, per esempio t_1 , ed effettuare la media dei valori temporali di un gran numero di queste manifestazioni $z_1(t_1), z_2(t_1), \dots$ ottenendo, in tal modo, la quantità:

$$\bar{z}(t_1) = \frac{\sum z_i(t_1)}{n} \quad (3.29)$$

che prende nome di *media d'insieme*.

Ciò premesso, un processo temporale si dice *ergodico* se la media d'insieme tende alla media temporale, almeno quando n è molto grande. Si può dimostrare che se un processo temporale è stazionario, è anche ergodico.

Se è valida l'ipotesi di ergodicità, la rilevazione effettuata su una singola manifestazione temporale in un gran numero di punti successivi porta alle stesse distribuzioni statistiche che si otterrebbero considerando un gran numero di valori riferiti allo stesso istante t_i .

3.7 Test di significatività ed intervalli di confidenza

E' nota la profonda differenza concettuale fra i momenti teorici, calcolati su un modello, ed i momenti campionari, calcolati sui dati osservati. Mentre i primi, nell'ipotesi di stazionarietà, sono delle costanti, i momenti campionari, essendo funzione dei dati osservati, variano da campione a campione. Quindi le statistiche campionarie, ossia qualunque grandezza calcolata sui dati campionari, sono esse stesse delle v.c. ed il particolare valore ricavato da un dato campione è una particolare manifestazione di quella variabile casuale.

Nel caso delle serie temporali, le statistiche campionarie che interessano maggiormente sono r_k e F_{kk} . Il problema che si pone è allora quello di valutare se il valore calcolato sul campione dell'autocorrelazione totale e parziale sia rappresentativo o meno del corrispondente valore teorico (si tratta, quindi, di valutare se il campione di osservazioni disponibili possa o meno ritenersi estratto da una popolazione di caratteristiche note). Tale problema si risolve calcolando, per ciascuna quantità stimata, lo scarto quadratico medio della sua distribuzione campionaria (*standard error*) il quale dà un'idea della dispersione dei valori campionari attorno alla media.

La valutazione dello standard error permette di eseguire dei *test di significatività* sui valori calcolati delle statistiche campionarie di interesse.

L'esecuzione di un test di significatività su un valore campionario dell'autocorrelazione deve permettere di giudicare se un dato valore, ad esempio $r_k=0.12$, può ritenersi, con un assegnato livello di significatività, proveniente da un modello con $r_k=0$, che rappresenta la condizione di indipendenza seriale del processo (processo in correlato).

Bartlett ha dimostrato che quando $r_k=0$, la distribuzione campionaria di r_k è all'incirca normale, con media nulla e varianza $s^2(r_k)$ data dalla relazione:

$$s^2(r_k) \cong \frac{1}{n} \left(1 + 2 \sum_{i=1}^q r_i^2 \right) \quad (3.30)$$

nella quale q indica il valore oltre il quale le autocorrelazioni teoriche sono tutte nulle.

Ne consegue che la variabile standardizzata $(r_k - r_k) / s(r_k)$ è anch'essa una v.c., caratterizzata da una distribuzione normale di media nulla e varianza unitaria. Ciò si può esprimere sinteticamente scrivendo:

$$\frac{r_k}{s(r_k)} \rightarrow N(0,1) \quad \text{per } k=q+1, q+2, \dots \quad (3.31)$$

Per le proprietà della distribuzione normale, si ha che:

$$\Pr\{|r_k| \leq 2s(r_k)\} = 0.95 \quad (3.32)$$

Possiamo pertanto definire il seguente test per stabilire se il valore teorico dell'autocorrelazione totale sia nullo:

TEST: se $|r_k| \leq 2s(r_k)$ per $k=q+1, q+2, \dots$ (3.33)
 si accetta l'ipotesi che $r_k=0$.

Poiché si può dimostrare che se X_1, X_2, \dots, X_n sono v.c. identicamente distribuite la stima di r_k (cioè r_k) è distribuita all'incirca in modo normale con parametri:

$$E(r_k) = 0/n$$

$$var(r_k) \approx 1/n,$$

per la precedente proprietà, è anche facile calcolare i limiti di confidenza al 95% di probabilità, che risultano:

$$E(r_k) \pm 2s(r_k) = -\frac{1}{n} \pm \frac{2}{\sqrt{n}} \cong \frac{2}{\sqrt{n}} \quad (3.34)$$

Va notato che circa un coefficiente ogni 20 può cadere al di fuori dei limiti stessi senza che debba essere rifiutata l'ipotesi.

Una procedura assolutamente analoga si segue per la funzione di autocorrelazione parziale campionaria F_{kk} . E' stato infatti dimostrato che quando $F_{kk}=0$ ($k > p$), la stima F_{kk} è distribuita con legge normale con media zero e varianza unitaria. Il test sul valore campionario del coefficiente di autocorrelazione parziale F_{kk} può essere formulato come segue:

TEST: se $|\Phi_{kk}| \leq \frac{2}{\sqrt{n}}$ (3.35)
 si accetta l'ipotesi che $F_{kk}=0$.

4. I MODELLI STOCASTICI DI BOX E JENKINS

Come abbiamo avuto modo di affermare in precedenza, la moderna teoria delle serie temporali ipotizza che una serie osservata sia una manifestazione di un certo processo o modello stocastico. I *modelli di Box e Jenkins*, espressione di processi stocastici stazionari, sono quelli che vengono generalmente usati per le indagini sulle serie temporali.

Prima di esaminare in dettaglio i modelli di Box e Jenkins, consideriamo un particolare tipo di processo stocastico, detto White Noise (rumore bianco) che genera delle serie temporali non correlate.

4.1 Il processo stocastico White Noise

Un processo stocastico White Noise è rappresentato da una serie di prove indipendenti effettuate sulla stessa variabile casuale avente media e varianza costanti. Tale processo dà luogo ad una serie di dati non correlati z_1, z_2, z_3, \dots che costituiscono, quindi, un processo puramente casuale.

L'equazione del modello White Noise è data semplicemente da:

$$z_t = a_t \quad (t=1, 2, 3, \dots) \quad (4.1)$$

Se le variabili, come d'altra parte abbiamo supposto, hanno media e varianza costanti, anche la media e la varianza del processo saranno costanti e quindi:

$$E(z_t) = \text{costante} \quad (4.2)$$

$$\text{var}(z_t) = \text{costante} \quad (4.3)$$

Inoltre, essendo le variabili casuali a_t indipendenti, la loro autocovarianza sarà nulla e, quindi:

$$g_k = \text{cov}(z_t, z_{t+k}) = 0 \quad \text{per tutti i valori di } k \quad (4.4)$$

Pertanto sarà nulla anche la serie dei coefficienti di autocorrelazione r_k , salvo $r_0 = 1$.

Per tutto quanto detto, il processo stocastico di tipo White Noise è facilmente riconoscibile in quanto, fra tutte le autocorrelazioni, solo r_0 è diversa da zero.

Poichè i momenti di primo e secondo ordine non dipendono dal tempo t , il processo è sicuramente caratterizzato da una stazionarietà debole. Si può però facilmente dimostrare che il processo è stazionario in modo completo.

4.2 Il modello AR(p)

A. Equazione del modello

La forma generale di un *processo autoregressivo di ordine p*, indicato sinteticamente con $AR(p)$, è espressa dalla relazione:

$$z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \dots + \Phi_p z_{t-p} + a_t \quad (4.5)$$

e risulta, quindi, come una "somma ponderata di valori passati di z_t alla quale si aggiunge un

disturbo a_t calcolato sul valore attuale t del tempo".

La relazione (4.5) si può anche scrivere come:

$$z_t - \Phi_1 z_{t-1} - \Phi_2 z_{t-2} - \dots - \Phi_p z_{t-p} = a_t \quad (4.6)$$

Utilizzando l'operatore all'indietro (backward) B , definito dalla relazione:

$$B z_t = z_{t-1} \quad (4.7)$$

si ha:

$$(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p) z_t = a_t \quad (4.8)$$

e quindi, indicando con $F(B)$ la quantità fra parentesi, si può sintetizzare come segue:

$$F(B) z_t = a_t \quad (4.9)$$

Se il processo non è a media nulla nell'equazione che esprime il modello AR(p) si aggiunge una costante d , che misura il "livello del processo" (approssima la sua media se esso è stazionario):

$$F(B) z_t = a_t + d$$

B. Stazionarietà

La condizione affinché un processo $AR(p)$ sia stazionario è legata alle radici dell'equazione caratteristica che si ottiene uguagliando a zero l'operatore $F(B)$:

$$1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p = 0 \quad (4.10)$$

Considerando B come incognita si dimostra che la stazionarietà di un processo AR(p) si ha quando le radici dell'equazione caratteristica (4.10) sono in modulo maggiori dell'unità, ossia sono esterne al cerchio di raggio unitario:

$$\begin{cases} \Phi B_i = 0 \\ |B_i| > 1 \end{cases} \quad i = 1, \dots, p$$

C. Invertibilità

Come accennato in precedenza, l'utilizzo di un modello di Box-Jenkins richiede, in generale, anche l'esame del problema dell'invertibilità. Nel caso particolare del processo AR(p) questo esame è superfluo in quanto si può dimostrare che non si richiede alcuna condizione ai parametri F_1, F_2, \dots, F_p , affinché il modello sia invertibile.

D. Valor medio

Se le v.c. a_t hanno valor medio nullo, il valore medio atteso del processo $AR(p)$ è dato da:

$$E(z_t) = \frac{d}{1 - \Phi_1 - \Phi_2 - \dots - \Phi_p} \quad (4.11)$$

E. Varianza ed autocovarianza

La varianza g_0 e le autocovarianze g_1, g_2, \dots, g_p si calcolano mediante le relazioni:

$$\mathbf{g}_0 = \Phi_1 \mathbf{g}_1 + \Phi_2 \mathbf{g}_2 + \dots + \Phi_p \mathbf{g}_p + \mathbf{s}_a^2 = \mathbf{s}_z^2 = \text{var}(z) \quad (4.12)$$

$$\mathbf{g}_k = \Phi_1 \mathbf{g}_{k-1} + \Phi_2 \mathbf{g}_{k-2} + \dots + \Phi_p \mathbf{g}_{k-p} \quad k=1,2,\dots,p \quad (4.13)$$

dove con \mathbf{s}_a^2 si indica la varianza della componente casuale del processo.

E' opportuno notare che le autocovarianze teoriche del processo $AR(p)$ sono in numero infinito.

Per i valori di γ_j (con $j > p$) si ricorre ancora ad una relazione del tipo:

$$\mathbf{g}_j = \Phi_1 \mathbf{g}_{j-1} + \Phi_2 \mathbf{g}_{j-2} + \dots + \Phi_p \mathbf{g}_{j-p} \quad (j > p) \quad (4.14)$$

che prende nome di equazione di Yule-Walker.

La relazione (4.14) è molto utile in quanto consente di risolvere due problemi:

- 1) una volta noto un certo modello $AR(p)$ e, quindi, noti i valori dei parametri Φ_1, \dots, Φ_p si possono calcolare le autocovarianze teoriche corrispondenti al modello scelto;
- 2) se di un dato modello sono ignoti i parametri Φ_1, \dots, Φ_p , questi possono essere stimati sostituendo ai valori teorici delle autocovarianze i corrispondenti valori campionari c_1, \dots, c_p che si ricavano dalla serie osservata (metodo dei momenti).

F. Autocorrelazione

Se si divide la relazione (4.13) per \mathbf{g}_0 si ottiene la formula ricorrente:

$$\mathbf{r}_k = \Phi_1 \mathbf{r}_{k-1} + \Phi_2 \mathbf{r}_{k-2} + \dots + \Phi_p \mathbf{r}_{k-p} \quad k=1,2,3,\dots \quad (4.15)$$

anch'essa nota con il nome di equazione di Yule-Walker.

Partendo da $\mathbf{r}_0=1$ si ottengono, con la formula ricorrente (4.15), tutti gli altri coefficienti di autocorrelazione teorica del processo $AR(p)$. Analogamente a quanto notato per le autocovarianze, la relazione (4.15) consente di risolvere due problemi, e precisamente:

- 1) calcolare le autocorrelazioni quando è noto il modello $AR(p)$;
- 2) stimare i parametri Φ_1, \dots, Φ_p sostituendo ai valori teorici delle autocorrelazioni i corrispondenti valori campionari r_1, \dots, r_p desunti dalla serie osservata (metodo dei momenti).

G. Correlogramma

Il correlogramma di un processo $AR(p)$, come si evince dalla (4.15) è costituito da infiniti termini. Si può, inoltre, dimostrare che questi infiniti termini possono tendere a zero in modo monotono oppure con oscillazioni a seconda del valore assunto dai parametri del modello.

H. Autocorrelazione parziale

Nel modello $AR(p)$ l'autocorrelazione parziale è uno strumento molto importante in quanto è costituita da p termini, cioè esattamente da tanti termini quanto è l'ordine del processo. Si dimostra, infatti, che i coefficienti di autocorrelazione parziale F_{kk} sono nulli per $k > p$. Si vede dunque l'utilità dell'autocorrelazione parziale dal momento che quella totale è sempre costituita da infiniti termini in tutti i processi $AR(p)$.

4.3.1 Caso particolare: il processo $AR(1)$

A. Equazione del modello

La forma generale del processo autoregressivo di ordine 1, indicato sinteticamente con $AR(1)$, è espressa dalla relazione:

$$z_t = \Phi_1 z_{t-1} + a_t \quad (4.16)$$

Da questa relazione risulta che il valore della serie temporale al tempo t , ossia z_t , è uguale ad una frazione del valore precedente z_{t-1} aumentata (in senso algebrico) della componente casuale a_t .

B. Stazionarietà

La condizione affinché un processo $AR(1)$ sia stazionario richiede che le radici dell'equazione caratteristica nell'incognita B :

$$1 - \Phi_1 B = 0 \quad (4.17)$$

siano esterne al cerchio di raggio unitario. Ciò equivale a dire che, per la stazionarietà, deve essere verificata la disuguaglianza (questa volta applicata al parametro F):

$$|F_1| < 1 \quad (4.18)$$

C. Valor medio

Se le v.c. a_t hanno come valor medio $E(a_t)=0$, il valore medio atteso del processo $AR(1)$ è anch'esso nullo:

$$E(z_t) = 0 \quad (4.19)$$

D. Varianza

La varianza del processo è data dalla relazione:

$$s^2(z) = \frac{s_a^2}{1 - \Phi_1^2} = g_0 \quad (4.20)$$

Dalla relazione precedente si ha una conferma al fatto che deve risultare $|F_1| < 1$. Infatti, affinché il processo abbia varianza positiva, il parametro F_1 deve essere minore dell'unità, come rilevato in precedenza a proposito della stazionarietà.

E. Autocovarianza

L'autocovarianza di un processo $AR(1)$ si ricava mediante la relazione:

$$g_k = \frac{s_a^2}{1 - \Phi_1^2} \Phi_1^k \quad k=1,2,3,\dots \quad (4.21)$$

Se si tiene conto dell'espressione della varianza del processo, si può scrivere, più semplicemente:

$$g_k = g_0 \Phi_1^k \quad k=1,2,3,\dots \quad (4.22)$$

F. Autocorrelazione

Dalla relazione (4.22) per si ricavano immediatamente le autocorrelazioni del processo che sono espresse da:

$$r_k = \Phi_1^k \quad k=1,2,3,\dots \quad (4.23)$$

G. Correlogramma

Il correlogramma di un processo $AR(1)$, sempre costituito da infiniti termini, ha forma diversa a seconda del segno assunto dal parametro Φ_1 del modello, come si evince dalla figura 4.1.

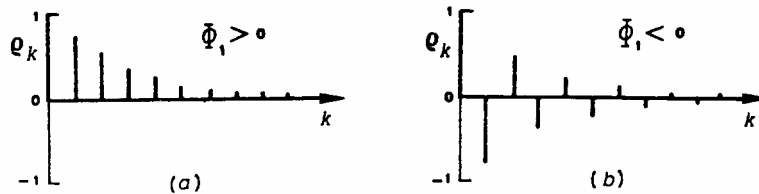


Fig 4.1 Correlogrammi di un modello $AR(1)$

H. Autocorrelazione parziale

Nel modello $AR(1)$, per quanto detto in precedenza, l'autocorrelazione parziale è costituita da un solo termine. L'espressione è:

$$\Phi_{11} = \Phi_1 = \rho_1 \quad (4.24)$$

$$\Phi_{kk} = 0 \quad \text{per } k > 1 \quad (4.25)$$

4.3 Il modello $MA(q)$

A. Equazione del modello

La forma generale di un *processo a media mobile di ordine q*, indicato sinteticamente con $MA(q)$, è espressa dalla relazione:

$$z_t = a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots + \Psi_q a_{t-q} \quad (4.26)$$

e risulta, quindi, come una "somma ponderata dei valori di una successione di variabili casuali indipendenti, aventi media e varianza costanti".

E' consuetudine, nell'espressione di un modello a media mobile, indicarne i parametri con la lettera Θ per cui, ponendo $Y_i = -Q_i$, si ha:

$$z_t = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - \dots - \Theta_q a_{t-q} \quad (4.27)$$

Utilizzando l'operatore all'indietro backward, la (4.27) può essere scritta sinteticamente come:

$$z_t = \Theta(B) a_t \quad (4.28)$$

avendo indicato con $\Theta(B)$ l'operatore $MA(q)$ espresso dalla relazione

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q \quad (4.29)$$

B. Stazionarietà

Poichè la serie (4.27) è finita, nessuna restrizione è necessaria per i parametri del modello

Media Mobile per assicurarne la stazionarietà.

C. Invertibilità

La valutazione dell'invertibilità del modello stocastico è particolarmente interessante in quanto nei modelli a media mobile esiste il problema della molteplicità dei modelli, ossia, ad una certa funzione di autocorrelazione possono corrispondere due o più modelli diversi. Questo fatto sarà mostrato in seguito con riferimento al caso semplice del modello $MA(1)$.

La condizione affinché un processo $MA(q)$ sia invertibile si desume dalle radici dell'equazione caratteristica:

$$1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q = 0 \quad (4.30)$$

Si dimostra che l'invertibilità di un processo $MA(q)$ si ha quando le radici della (4.30), sono in modulo maggiori dell'unità, ossia sono esterne al cerchio di raggio unitario:

$$\begin{cases} \Theta(B_i) = 0 \\ |B_i| > 1 \end{cases} \quad i = 1, \dots, q$$

D. Valor medio

Se le v.c. a_t hanno come valor medio $E(a_t) = 0$, il valore medio atteso del processo $MA(q)$ è pure nullo e, quindi:

$$E(z_t) = 0 \quad (4.31)$$

E. Autocovarianza e varianza

Le autocovarianze in un processo $MA(q)$ si calcolano mediante le relazioni:

$$\mathbf{g}_k = (-\Theta_k + \Theta_1 \Theta_{k-1} + \Theta_2 \Theta_{k-2} + \dots + \Theta_q \Theta_{k-q}) \mathbf{s}_a^2 \quad k=1, 2, \dots, q \quad (4.32)$$

$$\mathbf{g}_k = 0 \quad k > q \quad (4.33)$$

dove con \mathbf{s}_a^2 si indica la varianza delle componenti casuali del processo.

La varianza teorica del modello $MA(q)$ si ottiene ponendo $k=0$ nella (4.32):

$$\mathbf{g}_0 = (1 + \Theta_1^2 + \Theta_2^2 + \dots + \Theta_q^2) \mathbf{s}_a^2 \quad (4.34)$$

Dalle relazioni precedenti si evince che il processo $MA(q)$ è sempre stazionario in quanto le grandezze ora viste non dipendono dal tempo t .

F. Autocorrelazione

Dalle relazioni (4.32) e (4.34) si ricava subito:

$$\mathbf{r}_k = \frac{-\Theta_k + \Theta_1 \Theta_{k-1} + \dots + \Theta_q \Theta_{q-k}}{1 + \Theta_1^2 + \dots + \Theta_q^2} \quad k=1, 2, \dots, q \quad (4.35)$$

$$\mathbf{r}_k = 0 \quad k > q \quad (4.36)$$

G. Correlogramma

Il correlogramma di un processo $MA(q)$, come si evince dalle (4.35) e (4.36), è costituito da q termini e questa informazione è estremamente preziosa in quanto, se ciò si verifica in un

correlogramma empirico, consente di orientarsi verso tale modello.

H. Autocorrelazione parziale

Le espressioni delle autocorrelazione parziali Φ_{kk} si presentano in forma alquanto complicata. Nel seguito è riportata l'espressione dell'autocorrelazione parziale per il caso semplice del processo $MA(1)$.

4.4.1 Caso particolare: il processo MA(1)

A. Equazione del modello

La forma generale del *processo media mobile di ordine 1*, indicato sinteticamente con $MA(1)$, è espressa dalla relazione:

$$z_t = a_t - \Theta a_{t-1} \quad (4.37)$$

B. Stazionarietà

Per quanto detto in precedenza, il modello $MA(1)$ è sempre stazionario qualunque sia il valore del parametro Θ .

C. Valor medio

Se le v.c. a_t hanno come valor medio $E(a_t)=0$, il valore medio atteso del modello $MA(1)$ è anch'esso nullo:

$$E(z_t) = 0 \quad (4.38)$$

D. Varianza

La varianza del processo è data dalla relazione:

$$g_0 = (1 + \Theta^2) s_a^2 = s_z^2 = \text{var}(z_t) \quad (4.39)$$

E. Autocovarianza

L'autocovarianza di un processo $MA(1)$ si ricava mediante la relazione:

$$g_1 = -\Theta s_a^2 \quad (4.40)$$

$$g_k = 0 \quad k > 1 \quad (4.41)$$

F. Autocorrelazione

Dalle relazioni precedenti si ricavano immediatamente le autocorrelazioni del processo che sono espresse da:

$$r_1 = \frac{g_1}{g_0} = \frac{-\Theta}{1 + \Theta^2} \quad (4.42)$$

$$r_k = 0 \quad k > 1 \quad (4.43)$$

G. Correlogramma

Il correlogramma di un processo $MA(1)$ è costituito da un solo termine, positivo o negativo a seconda del segno del parametro del modello (vedi figura 4.2)

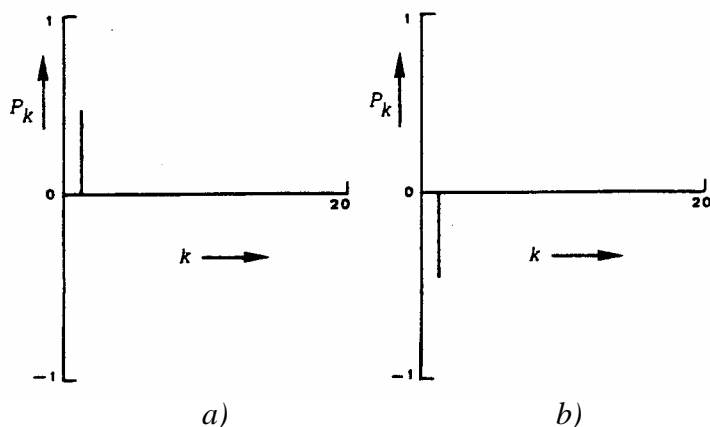


Fig 4.2. Esempio di autocorrelogramma per il modello $MA(1)$: Caso a): $Q < 0$; Caso b:) $Q > 0$.

H. Autocorrelazione parziale

Si può dimostrare che le autocorrelazioni parziali di un modello $MA(1)$ sono date da:

$$\Phi_{kk} = \frac{-\Theta^k (1 - \Theta^2)}{1 - \Theta^{2(k+1)}} \quad k=1, 2, 3, \dots \quad (4.44)$$

Da questa relazione si ricava che i coefficienti di autocorrelazione parziale non si esauriscono di colpo ma si smorzano in modo esponenziale, anche con oscillazioni di segno.

I. Invertibilità

Il problema dell'invertibilità può essere illustrato molto bene nel caso del processo $MA(1)$ considerando la funzione di autocorrelazione.

Supponiamo di avere due processi $MA(1)$ aventi come parametri, rispettivamente, Θ e $1/\Theta$:

$$z_t = a_t - \Theta a_{t-1} \quad (4.45)$$

$$z'_t = a_t - \frac{1}{\Theta} a_{t-1} \quad (4.46)$$

Se calcoliamo i coefficienti di autocorrelazione dei due processi otteniamo:

$$r_1(z_t) = \frac{-\Theta}{1 + \Theta^2} \quad (4.47)$$

$$r_1(z'_t) = \frac{-1/\Theta}{1 + 1/\Theta^2} = \frac{-\Theta}{1 + \Theta^2} \quad (4.48)$$

Come si può notare, i due processi, pur essendo diversi, presentano la stessa autocorrelazione. E' questo un tipico esempio di molteplicità dei modelli.

Il problema può essere affrontato esprimendo a_t in funzione di z_t, z_{t-1}, \dots ottenendo:

$$a_t = z_t - \Theta z_{t-1} + \Theta^2 z_{t-2} - \dots \quad (4.49)$$

$$a'_t = z'_t - 1/\Theta z'_{t-1} + 1/\Theta^2 z'_{t-2} - \dots \quad (4.50)$$

Ricorrendo all'operatore B , si ottiene:

$$a_t = (1 - \Theta B + \Theta^2 B^2 - \dots) z_t \quad (4.51)$$

$$a_t = (1 - 1/\Theta B + 1/\Theta^2 B^2 - \dots) z'_t \quad (4.52)$$

E' utile osservare che, per $|\Theta| < 1$ la prima serie converge mentre la seconda diverge. In altre parole, si dice che, quando $|\Theta| < 1$ il modello $MA(1)$ è invertibile ossia esiste un unico processo che corrisponde ad una certa funzione di autocorrelazione.

Questo risultato coincide con la condizione generale di invertibilità data in precedenza. Infatti, dire che $|\Theta| < 1$ equivale ad affermare che le radici dell'equazione caratteristica del modello $MA(1)$:

$$1 - \Theta(B) = 0 \quad (4.53)$$

cadono al di fuori del cerchio di raggio unitario.

4.5 Il modello ARMA(p,q)

A. Equazione del modello

I due processi che abbiamo considerato separatamente possono venire composti in un unico modello misto in base alla seguente considerazione.

Il modello $AR(p)$ si presenta nella forma:

$$z_t - \Phi_1 z_{t-1} - \Phi_2 z_{t-2} - \dots - \Phi_p z_{t-p} = a_t \quad (4.54)$$

Si suppone, quindi, che il residuo sia costituito dalla variabile casuale a_t le cui manifestazioni siano normali, indipendenti, con media nulla e varianza costante.

In diversi casi è stato però constatato che questo residuo ha le caratteristiche di una media mobile di ordine q, sempre sulla variabile a_t e, quindi, risulta del tipo:

$$z_t = a_t - \Theta_1 a_{t-1} - \dots - \Theta_q a_{t-q} \quad (4.55)$$

L'equazione del modello misto autoregressivo di ordine p con media mobile di ordine q, indicato con ARMA(pq), è, quindi:

$$z_t - \Phi_1 z_{t-1} - \Phi_2 z_{t-2} - \dots - \Phi_p z_{t-p} = a_t - \Theta_1 a_{t-1} - \dots - \Theta_q a_{t-q} \quad (4.56)$$

Ricorrendo all'operatore B si ottiene l'espressione:

$$\Phi(B) = \Theta(B)a_t \quad (4.57)$$

in cui:

$$\Phi(B) = \Phi_1 z_{t-1} - \Phi_2 z_{t-2} - \dots - \Phi_p z_{t-p}$$

rappresenta l'operatore autoregressivo e

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q$$

rappresenta l'operatore media mobile.

L'importanza pratica del modello misto risiede nel fatto che per molte serie temporali esso richiede un numero di parametri inferiore a quelli necessari per un modello autoregressivo puro.

B. Stazionarietà

La condizione affinché un processo $ARMA(p,q)$ sia stazionario è legata ancora alle radici dell'equazione caratteristica della parte autoregressiva del modello, che si ottiene uguagliando a zero l'operatore $\Phi(B)$ considerando B come incognita:

$$1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p = 0 \quad (4.58)$$

Si dimostra che "la stazionarietà di un processo $ARMA(p,q)$ si ha quando le radici

dell'equazione caratteristica (4.58), sono in modulo maggiori dell'unità, ossia sono esterne al cerchio di raggio unitario".

C. Invertibilità

La condizione affinché un processo $ARMA(p,q)$ sia invertibile si desume dalle radici dell'equazione caratteristica della parte media mobile del modello, che si ottiene, uguagliando a zero l'operatore $\Theta(B)$ considerando B come incognita:

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q = 0 \quad (4.59)$$

Si dimostra che "l'invertibilità di un processo $ARMA(p,q)$ si ha quando le radici dell'equazione caratteristica (4.59) sono in modulo maggiori dell'unità, ossia sono esterne al cerchio di raggio unitario".

D. Valor medio

Se il processo misto è completo, e quindi contiene anche una costante δ , la media teorica del modello risulta:

$$E(z_t) = \mathbf{d} + \Phi_1 E(z_{t-1}) + \dots + \Phi_p E(z_{t-p}) + E(a_t) - \Theta_1 E(a_{t-1}) - \dots - \Theta_q E(a_{t-q}) \quad (4.60)$$

da cui si ricava in modo semplice che:

$$E(z_t) = \frac{\mathbf{d}}{1 - \Phi_1 - \Phi_2 - \dots - \Phi_p} \quad (4.61)$$

E. Autocovarianza

Le espressioni delle autocovarianze in un modello $ARMA(p,q)$ sono fornite dalle relazioni:

$$\mathbf{g}_k = \Phi_1 \mathbf{g}_{k-1} + \dots + \Phi_p \mathbf{g}_{k-p} + \mathbf{g}_{za}(k) - \Theta_1 \mathbf{g}_{za}(k-1) - \dots - \Theta_q \mathbf{g}_{za}(k-q) \quad (k < q) \quad (4.61)$$

$$\mathbf{g}_k = \Phi_1 \mathbf{g}_{k-1} + \Phi_2 \mathbf{g}_{k-2} + \dots + \Phi_p \mathbf{g}_{k-p} \quad (k \geq q+1) \quad (4.62)$$

dove con $\mathbf{g}_{za}(k)$ si è indicata la covarianza incrociata fra le variabili z_t ed a_t :

$$\mathbf{g}_{za}(k) = E[(z_{t-k} - \bar{z})(a_t - \bar{a})] \quad (4.63)$$

F. Varianza

La varianza del processo $ARMA(p,q)$ è espressa dalla relazione:

$$\mathbf{g}_0 = \Phi_1 \mathbf{g}_1 + \dots + \Phi_p \mathbf{g}_p + \mathbf{s}_a^2 - \Theta_1 \mathbf{g}_{za}(-1) - \dots - \Theta_q \mathbf{g}_{za}(-q) \quad (4.64)$$

Si noti che la varianza del modello può essere calcolata solo dopo aver risolto le p relazioni (4.62) per $k=1,2,\dots,p$ in modo da ottenere $\gamma_1, \gamma_2, \dots, \gamma_p$.

G. Autocorrelazione

Dalla relazione (4.62) si ottiene immediatamente:

$$\mathbf{r}_k = \Phi_1 \mathbf{r}_{k-1} + \Phi_2 \mathbf{r}_{k-2} + \dots + \Phi_p \mathbf{r}_{k-p} \quad (k \geq q+1) \quad (4.65)$$

H. Autocorrelazione parziale

Nel modello $ARMA(p,q)$ l'autocorrelazione parziale è costituita da infiniti termini ed ha un andamento simile a quello dell'autocorrelazione parziale di un processo a media mobile.

4.5.1 Caso particolare: il processo $ARMA(1,1)$

A. Equazione del modello

Il caso più semplice ed usato del processo di tipo misto è il modello $ARMA(1,1)$, caratterizzato dalla relazione:

$$z_t - \Phi_1 z_{t-1} = a_t - \Theta_1 a_{t-1} \quad (4.66)$$

B. Stazionarietà

La condizione affinché un processo $ARMA(1,1)$ sia stazionario è analoga a quella di stazionarietà di un processo $AR(1)$, cioè:

$$|\Phi_1| < 1 \quad (4.67)$$

C. Invertibilità

La condizione affinché un processo $ARMA(1,1)$ sia invertibile è analoga a quella di invertibilità di un processo $MA(1)$, cioè:

$$|\Theta_1| < 1 \quad (4.68)$$

D. Valor medio

Se il processo $ARMA(1,1)$ è completo, cioè si presenta sotto la forma:

$$z_t - \Phi_1 z_{t-1} = \mathbf{d} + a_t - \Theta_1 a_{t-1} \quad (4.69)$$

il valor medio atteso si ricava dalla seguente relazione:

$$E(z_t) - \Phi_1 E(z_{t-1}) = \mathbf{d} + E(a_t) - \Theta_1 E(a_{t-1}) \quad (4.70)$$

da cui:

$$E(z_t) = \frac{\mathbf{d}}{1 - \Phi_1} \quad (4.71)$$

E. Autocovarianza

Le espressioni delle autocovarianze di un processo $ARMA(1,1)$ risultano:

$$\mathbf{g}_1 = \frac{(1 - \Phi_1 \Theta_1)(\Phi_1 - \Theta_1)}{1 - \Phi_1^2} \mathbf{s}_a^2 \quad (4.72)$$

$$\mathbf{g}_k = \Phi_1 \mathbf{g}_{k-1} \quad k=2,3,\dots \quad (4.73)$$

F. Varianza

La varianza del processo è data dalla relazione:

$$\text{var}(z_t) = \mathbf{g}_0 = \frac{1 - 2\Phi_1 \mathbf{g}_1 + \Theta_1^2}{1 - \Phi_1^2} \mathbf{s}_a^2 \quad (4.74)$$

G. Autocorrelazione

Dalle relazioni precedenti si ricava immediatamente:

$$\mathbf{r}_1 = \frac{\mathbf{g}_1}{\mathbf{g}_0} \quad (4.75)$$

$$\mathbf{r}_k = \Phi_1 \mathbf{r}_{k-1} \quad k=2,3,4,\dots \quad (4.76)$$

La funzione di autocorrelazione presenta un andamento smorzato di tipo esponenziale con segni tutti uguali o alternati a seconda che risulti $F_1 > 0$ oppure $F_1 < 0$

H. Correlogramma

Il correlogramma di un processo $ARMA(1,1)$ può assumere andamenti diversi in funzione dei valori assunti dai parametri che caratterizzano il processo stesso (v. figura 4.3).

I. Autocorrelazione parziale

Nel modello $ARMA(1,1)$ la funzione di autocorrelazione parziale Φ_{kk} parte dal valore iniziale

$$\Phi_{11} = \rho_1$$

e va smorzandosi in modo monotono o con alternanze di segno (v. figura. 4.3).

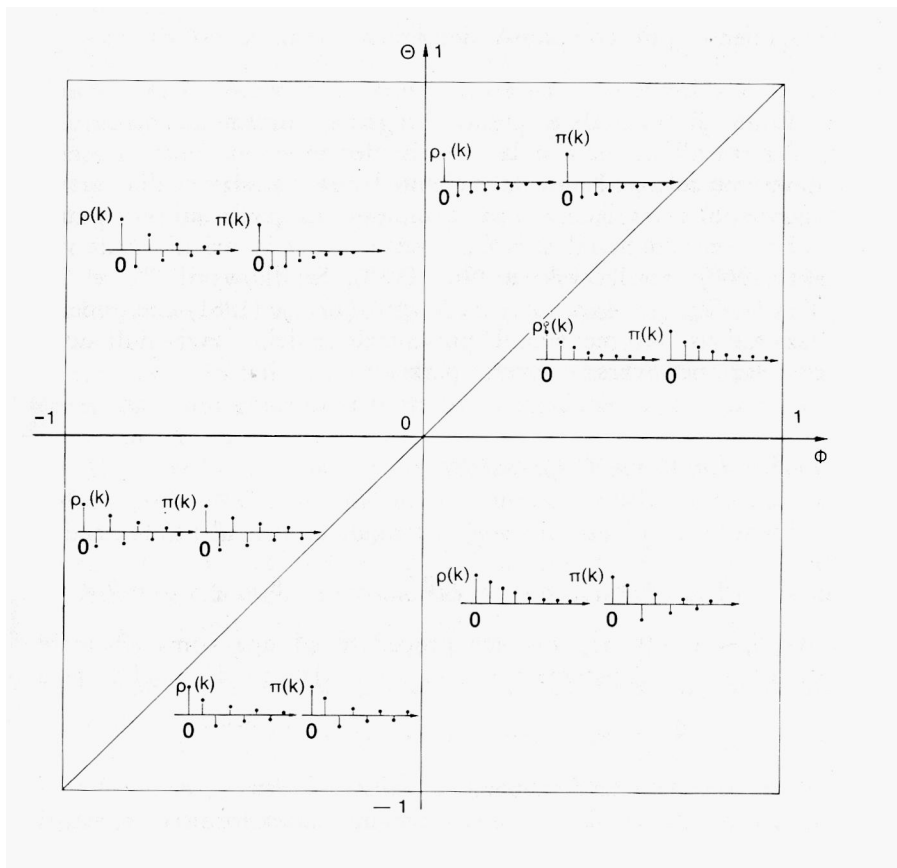


Fig. 4.3. Funzioni di autocorrelazione totale, r e parziale, p per il processo $ARMA(1,1)$.