

Indice

1	CATENE DI MARKOV	2
1.1	Introduzione	2
1.2	Descrizione probabilistica delle catene di Markov	6
1.3	Matrice di transizione, catene omogenee	11
1.4	Classificazione degli stati	18
1.5	Teoria ergodica per le catene di Markov	27
1.6	Catene di nascita o morte	38
2	PROCESSI STOCASTICI A TEMPI DISCRETI: TRATTAZIONE GENERALE	45
2.1	Premessa e prime definizioni	45
2.2	La descrizione hilbertiana dei processi discreti	53
2.3	La predizione lineare ottimale (caso finito)	71
2.4	Operatori lineari in uno spazio di Hilbert e loro rappresentazione matriciale. Propagazione della covarianza	82
2.5	Operatori causali e processi con lo stesso ordine temporale	86
2.6	Innovazione; processi regolari; decomposizione di Cholesky	96
2.7	Il problema della predizione lineare ottimale; caso infinito	102
2.8	Un'applicazione al filtro di Kalman-Bucy	106
3	I PROCESSI STAZIONARI	118
3.1	Processi stazionari in senso stretto	118
3.2	Processi stazionari in senso debole	122
3.3	Funzione di covarianza: caratteristiche	127
3.4	Lo spettro di potenza ed il calcolo spettrale	134
3.5	La predizione ottimale; calcolo di M ed A dallo spettro	145
3.6	Problemi di stima della funzione di covarianza e dello spettro	151

1 CATENE DI MARKOV

Premessa

Le catene di Markov costituiscono una particolare classe di processi stocastici, essendo costituite da una famiglia di variabili casuali definite in funzione di un parametro mono-dimensionale o n -dimensionale discreto che interpreteremo come un tempo, $n \in N^+ \equiv \{0, 1, 2, 3 \dots\}$. Per esse quindi si può definire un “prima” e un “dopo” ed una corrispondenza con il principio di causa ed effetto. La possibilità di individuare un senso obbligato (il “prima” può causare il “dopo” ma non può essere il viceversa) determina anche la trattazione matematica del problema.

Poiché il parametro varia per ipotesi in maniera discreta, il processo stocastico può essere visto come la generalizzazione di una variabile tempo casuale n -dimensionale: se $X^{(n)} = [X_0, X_1, X_2, \dots, X_n]^+$ è una variabile le cui estrazioni sono in \mathbb{R}^{n+1} , $X^{(\infty)} = [X_0, X_1, X_2, \dots, X_n, \dots]^+$ è una variabile casuale in \mathbb{R}^∞ (\mathbb{R}^∞ è il suo insieme di definizione) e ogni estrazione è rappresentata da una successione di numeri reali.

1.1 Introduzione

Consideriamo ora tra le serie temporali quella classe il cui insieme di definizione Ω è un reticolo in \mathbb{R}^∞ , ottenuto dal prodotto cartesiano di sottoinsiemi di \mathbb{R} discreti e finiti. I punti di Ω sono quindi successioni di numeri, ognuno dei quali può avere un valore scelto in un insieme discreto e finito di \mathbb{R} ; ogni punto, ovvero ogni successione di questo tipo costituisce una realizzazione del processo.

Le variabili marginali X_k di tali processi stocastici $X = \{X_0, X_1, X_2, \dots, X_k, \dots\}$ non sono delle variabili casuali continue definite in \mathbb{R} , ma delle variabili casuali discrete con un numero finito di valori argomentali.

Un esempio molto semplice è rappresentato dal lancio di una moneta infinite volte.

In tal caso, al variare di k , X_k può assumere solo due valori, testa o croce, ovvero 0,1, ovvero -1,1, con probabilità $\frac{1}{2}$, $\frac{1}{2}$.

I valori che X_k può assumere verranno di seguito indicati come *stati del sistema*.

Fig.1.1.1

In Fig. 1.1.1 sono stati indicati i valori argomentali rispettivamente della marginale X_0 , della marginale (X_0, X_1) , della marginale (X_0, X_1, X_2) del processo in esame. Essi sono ottenuti dal prodotto cartesiano dell'insieme discreto e finito $(0,1)$ con se stesso 1, 2, volte rispettivamente per le marginali a 2 e 3 dimensioni.

Un caso più complesso è rappresentato dal RANDOM WALK ¹.

Gli stati del processo sono, al variare di k :

$$\begin{aligned} s_0 &= x_0 \\ s_1 &= x_0 + x_1 \\ \dots & \\ s_k &= x_0 + x_1 + \dots + x_k \\ \dots & \end{aligned} \tag{1.1.1}$$

in cui x_k è il risultato del “lancio di una moneta”.

¹Non rientra, a rigore, nella classe delle catene di MARKOV, in quanto l'insieme dei possibili stati del sistema varia da istante a istante e non è limitato nel tempo

Una tipica realizzazione del processo, può essere rappresentata come in Fig. 1.1.2 quando x_k può assumere i valori (0,1)

Fig. 1.1.2

o come in Fig. 1.1.3

Fig. 1.1.3

quando x_k può assumere i valori (-1,1).

Caratteristica fondamentale di tali processi è che lo stato del sistema al tempo $k + 1$ dipende da quello al tempo k , ma è indipendente da come si è arrivati a tale istante. In termini probabilistici ciò si può esprimere nel seguente modo:

$$\begin{aligned} & P\{S_{k+1} = s_{k+1} | S_k = s_k, S_{k-1} = s_{k-1}, \dots, S_0 = s_0\} = \\ & = P\{S_{k+1} = s_{k+1} | S_k = s_k\} \end{aligned} \tag{1.1.2}$$

Ciò esprime la proprietà di Markov in senso debole.

Definizione 1.1.1: una CATENA DI MARKOV è un processo stocastico il cui insieme di definizione è un reticolo in \mathbb{R}^∞ , in cui una qualunque distribuzione condizionata soddisfa alla proprietà di Markov.

Osservazione 1.1.1: da un punto di vista fisico, la proprietà di Markov significa che, definito lo stato del sistema al tempo t , l'evoluzione successiva del sistema stesso raccoglie come unica informazione sul suo passato proprio il valore di x_t , mentre le successive variazioni sono legate ad una qualche causa stocastica, una "innovazione", indipendente dal passato.

Un notevole modello generale per leggi di questo tipo potrebbe essere

$$x_{k+1} = x_k + g(x_k; \varepsilon_{k+1}) \quad (1.1.3)$$

che ovviamente traduce a tempi discreti una equazione differenziale (o un sistema di equazioni differenziali se \underline{x}_k è una variabile pluridimensionale) del tipo

$$\dot{x} = g(x, \varepsilon) \quad (1.1.4)$$

Perché una equazione come la (1.1.4) dia luogo ad un processo di tipo markoviano basta assumere che ε_{k+1} sia stocasticamente indipendente da $\varepsilon_{k-1}, \varepsilon_{k-2}, \dots$; in tal caso infatti, è facile vedere che è soddisfatta la (1.1.4). Per avere una vera e propria catena di Markov, occorre anche supporre di discretizzare i valori assunti da x_k e che questi siano in numero limitato.

Riassumendo, nello studiare un particolare tipo di processi di Markov che chiameremo Catene di Markov incontreremo sempre la seguente situazione: sia dato un insieme discreto e finito dato l'insieme degli stati

$$\mathcal{S} = \{s_1, s_2, \dots, s_m\} \in \mathbb{R};$$

l'insieme Ω degli eventi elementari del nostro processo è formato da un reticolo regolare finito

$$r = \mathcal{S} \otimes \mathcal{S} \dots \otimes \mathcal{S} \dots = \mathcal{S}^\infty .$$

1.2 Descrizione probabilistica delle catene di Markov

Si consideri una catena associata ad una sequenza finita di tempi $k = 0, 1, 2, \dots, N$; i possibili risultati si rappresentano in \mathbb{R}^{N+1} come una successione di punti ognuno di coordinate $(x_0, x_1, x_2, \dots, x_N)$ e l'insieme degli stati S è costituito da m valori reali qualsiasi o da m punti; è sempre possibile immaginare di trasformare la variabile x in modo tale che i nuovi valori argomentali siano esattamente $x_k = i$; $i = 1, 2, \dots, m$. La descrizione probabilistica di questa catena sarà poi data dall'insieme delle probabilità discrete

$$P\{X_0 = x_0, X_1 = x_1, \dots, X_N = x_N\} = p^{(N)}(x_0, x_1, \dots, x_N). \quad (1.2.1)$$

Si noti che, come sempre nelle variabili a più dimensioni, è la distribuzione congiunta (1.2.1) che conta e non le singole distribuzioni

$$p_i(x_i) = P\{X_i = x_i\} \quad (1.2.2)$$

che sono invece le marginali della (1.2.1).

Il primo dubbio che si potrebbe avere è che, data la natura essenzialmente discreta del problema, sia possibile, anche per una catena associata ad una catena infinita di tempi, definire una regola di probabilità assegnando le probabilità punto a punto in Ω , cioè per ogni singola realizzazione.

Dimostreremo tra breve che ciò non è possibile perché Ω possiede la potenza del continuo. Tuttavia cominciamo a cercare di comprendere tale fatto con un controesempio elementare.

Esempio 1.2.1: sia $X = \{X_0, X_1, \dots\}$ la catena che rappresenta ∞ lanci indipendenti di una moneta, così che ciascuna marginale X_i possa assumere i valori 0 e 1 rispettivamente con probabilità $1/2, 1/2$.

Sia

$$C_N = \{X_0 = x_0, X_1 = x_1, \dots, X_N = x_N; \forall X_{N+1}, \forall X_{N+2}, \dots\} \quad (1.2.3)$$

un insieme in cui siano fissati i primi $N + 1$ valori. Le probabilità di tali insiemi devono coincidere con la (1.2.1) qualsiasi sia il valore di N e di $x_0, x_1, x_2, \dots, x_N$.

Nel nostro caso C_N avrà probabilità $P(C_N) = \frac{1}{2^N}$, così che, ovviamente, quando si cerchi la probabilità di una singola realizzazione

$$\begin{aligned} \underline{x} &= \{\bar{x}_0, \bar{x}_1, \dots, \} = \bigcap_{N=0}^{\infty} C_N = \\ &= \bigcap_{N=0}^{\infty} \{X_0 = \bar{x}_0, X_1 = \bar{x}_1, \dots, X_N = \bar{x}_N\} \end{aligned} \quad (1.2.4)$$

risulterà

$$P(\underline{x}) = \lim_{N \rightarrow \infty} P(C_N) = \lim_{N \rightarrow \infty} \frac{1}{2^N} = 0 \quad (1.2.5)$$

Come si vede ciò avviene per ogni realizzazione \underline{x} , così che appare ovvio che la distribuzione su Ω non può essere ricostruita dalla probabilità delle singole realizzazioni.

Come già detto, ciò che l'esempio precedente mette in luce è il fatto che Ω ha la potenza del continuo e quindi, come per ogni insieme con tale cardinalità, una distribuzione di probabilità non può essere definita sommando probabilità di punti individuali.

Ne segue che, se vogliamo dare una descrizione probabilistica complessiva della catena in \mathbb{R}^∞ occorrerà arrivare a definire un'opportuna famiglia di eventi (che deve anche essere una σ -algebra se si vuole che essa contenga tutte le unioni ed intersezioni numerabili di eventi) e le rispettive probabilità. È logico poi richiedere che tale famiglia contenga come caso particolare insiemi del tipo (1.2.3) che chiameremo cilindri. Si tratta cioè di costruire una distribuzione di probabilità, sull'insieme Ω di \mathbb{R}^∞ , che ammetta, per ogni N , le (1.2.2) come distribuzioni marginali.

La risposta ovvia, seppur alquanto astratta, al problema può venire dal seguente ragionamento:

- supponiamo di saper assegnare ad ogni C_N una probabilità P , che soddisfi le seguenti condizioni di compatibilità (Kolmogorov):

$$\sum_y P\{X_0 = x_0, X_1 = x_1, \dots, X_N = x_N, X_{N+1} = y\} =$$

$$= \sum_y p^{(N+1)} \{C_{N+1}(x_0, x_1, \dots, x_N, y)\} = \quad (1.2.6)$$

$$= p^{(N)} \{C_N(x_0, x_1, \dots, x_N)\} = P\{X_0 = x_0, X_1 = x_1, \dots, X_N = x_N\}$$

che esprimono il fatto che ogni $p^{(N)}$ può essere vista come distribuzione marginale di $p^{(N+1)}$,

- definiamo una famiglia \mathcal{A} di eventi che sia la minima σ -algebra di eventi che contiene la sottofamiglia dei cilindri

$$\{C_N; \quad \forall N, x_0, x_1, \dots, x_N\},$$

- supponiamo di poter estendere per continuità P da C_N ad un qualsiasi $A \in \mathcal{A}$ (vi sono appositi teoremi che dimostrano che tale estensione è possibile e noi ne vedremo tra breve una versione particolarmente semplice),

in tal modo avremo definito una distribuzione di probabilità che soddisfa ai requisiti minimali richiesti.

Naturalmente l'astrattezza del procedimento ci impedisce di comprendere chiaramente la natura della distribuzione di probabilità così costruita.

Lemma 1.2.1: *Ω ha la potenza del continuo.*

Dimostriamo il lemma nel caso semplice che il sistema abbia solo due stati, cui associamo i valori numerici $(0,1)$; la dimostrazione può essere generalizzata in modo semplice al caso in cui il sistema abbia m stati distinti.

Sia ϑ l'applicazione che fa corrispondere alla realizzazione $\underline{x} = \{x_0, x_1, \dots\}$ il numero reale, compreso tra 0 e 1 (estremi inclusi)

$$\vartheta(\underline{x}) = x_0 \frac{1}{2} + x_1 \frac{1}{2^2} + x_2 \frac{1}{2^3} + \dots \quad (1.2.7)$$

È ovvio che, potendo x_k assumere al più i valori 0 o 1,

$$\vartheta(\underline{x}) \in [0, 1] \quad (1.2.8)$$

cioè

$$\vartheta(\underline{x}) : \Omega \rightarrow [0, 1] \quad (1.2.9)$$

D'altro canto è facile anche vedere che per ogni $r \in [0, 1]$ esiste \underline{x} tale che

$$\vartheta(\underline{x}) = r \quad (1.2.10)$$

In effetti dalla (1.2.7), con ragionamento dicotomico, si vede che se $0 \leq r < \frac{1}{2}$ si può scegliere $x_0 = 0$ e se $\frac{1}{2} \leq r \leq 1$, $x_0 = 1$; se $0 \leq r < \frac{1}{4}$ allora $x_1 = 0$ e se invece $\frac{1}{4} \leq r < \frac{1}{2}$ $x_1 = 1$ e così via. Procedendo con tale criterio si determina una successione $\{x_k\}$ tale che

$$\left| r - \sum_{k=0}^N x_k \frac{1}{2^{k+1}} \right| < \frac{1}{2^{N+1}} \quad (1.2.11)$$

che, per $N \rightarrow \infty$, dimostra la (1.2.10).

È bene notare che ϑ non è biunivoca; in effetti ϑ^{-1} è non univoca per tutti i punti con ascissa del tipo

$$r = \frac{k}{2^N} (k \text{ intero}, 0 < k < 2^N) \quad (1.2.12)$$

L'esempio

$$\frac{1}{2} = 1 \frac{1}{2} + 0 \frac{1}{2^2} + 0 \frac{1}{2^3} + \dots = 0 \frac{1}{2} + 1 \frac{1}{2^2} + 1 \frac{1}{2^3} + \dots \quad (1.2.13)$$

dovrebbe essere sufficiente a chiarire il problema.

Tuttavia i numeri del tipo (1.2.12) costituiscono al più un insieme numerabile, cioè possono essere ordinati in una successione, così che per tutti i casi in cui i singoli punti hanno sempre probabilità zero come nell'Esempio 1.2.1, tale insieme ha probabilità nulla e non muta i nostri calcoli probabilistici.

Comunque poiché ogni $r \in [0, 1]$ ha almeno una controimmagine in Ω , resta dimostrato che Ω ha la potenza del continuo.

Osservazione 1.2.1: sia \overline{C}_N l'insieme

$$\overline{C}_N = \{X_0 = \overline{x}_0, X_1 = \overline{x}_1, \dots, X_N = \overline{x}_N, \forall X_{N+1} \dots\} \quad (1.2.14)$$

ci chiediamo quale sia l'immagine di \overline{C}_N secondo ϑ .

Sarà

$$\begin{aligned}
\{r\} &= \vartheta(\overline{C}_N) = \\
&= \left\{ \overline{x}_0 \frac{1}{2} + \dots + \overline{x}_N, \frac{1}{2^{N+1}} + \frac{1}{2^{N+1}} \left(x_{N+1} \frac{1}{2} + x_{N+2} \frac{1}{2^2} + \dots \right) \right\} = \\
&= \left\{ r_{0N} + \frac{1}{2^{N+1}} t \right\} \tag{1.2.15}
\end{aligned}$$

con t reale arbitrario in $[0,1]$.

Si vede che

$$\vartheta(\overline{C}_N) = \left[r_{0N}, r_{0N} + \frac{1}{2^{N+1}} \right] \tag{1.2.16}$$

Inoltre, per quanto detto,

$$\vartheta^{-1} \left[r_{0N}, r_{0N} + \frac{1}{2^{N+1}} \right] = (\overline{C}_N) \tag{1.2.17}$$

a meno possibilmente di una particolare successione (x_0, x_1, \dots) , cioè di un punto di Ω che non appartiene a \overline{C}_N .

Ne deriva che, supposto di sapere a priori che nel processo in esame nessun punto di Ω possa avere una probabilità positiva, il problema di costruire una distribuzione di probabilità in Ω (cioè di costruire \mathcal{A} e di estendervi P , definita inizialmente solo su C_N), può essere visto in immagine, tramite ϑ , come il problema di costruire una distribuzione di probabilità su $[0,1]$ a partire dai valori dati di probabilità per intervalli del tipo

$$P = \left[r_{0N}, r_{0N} + \frac{1}{2^{N+1}} \right] = P(\overline{C}_N) \tag{1.2.18}$$

Si noti che le condizioni di Kolmogorov, indicando con

$$r_{0N} = \frac{k}{2^{N+1}} \tag{1.2.19}$$

per k opportuno, diventano

$$\begin{aligned}
&P \left(r \in \left[\frac{k}{2^{N+1}}, \frac{k+1}{2^{N+1}} \right] \right) = \\
&= \sum_{y=0,1} P \left(r \in \left[\frac{k}{2^{N+1}} + \frac{y}{2^{N+2}}, \frac{k}{2^{N+1}} + \frac{y+1}{2^{N+2}} \right] \right) \tag{1.2.20}
\end{aligned}$$

Esse devono essere valide, purché un singolo punto in $[0,1]$ abbia sempre probabilità 0.

Quindi il problema di definire una distribuzione di probabilità in Ω a partire dalle distribuzioni marginali (1.2.2) è sostanzialmente equivalente alla costruzione di una distribuzione di probabilità su $[0,1]$ a partire dalla probabilità nota per gli insiemi,

Fig. 1.2.1

probabilità rese consistenti tra loro dalla condizione (1.2.20).

Poiché tale problema ha soluzione unica in termini di distribuzione su \mathbb{R} , altrettanto si può dire per la questione originalmente posta in Ω .

Come abbiamo detto questa argomentazione ha una ovvia generalizzazione al caso in cui la catena in esame abbia m possibili valori.

Terminate queste considerazioni di carattere generale, passiamo ad esaminare lo studio della evoluzione temporale della catena stessa, ovvero alla determinazione della distribuzione di probabilità tra gli stati al tempo $t = k$, cioè della distribuzione della variabile X_k , singolarmente presa.

1.3 Matrice di transizione, catene omogenee

Si abbia la seguente catena di Markov:

$$\begin{array}{cccccc}
 & X_0 & X_1 & \dots & X_1 & \dots \\
 s_1 & p_1^{(0)} & p_1^{(1)} & \dots & p_1^{(t)} & \dots \\
 s_2 & p_2^{(0)} & p_2^{(1)} & \dots & p_2^{(t)} & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 s_i & \dots & \dots & \dots & \dots & \dots \\
 s_m & p_m^{(0)} & p_m^{(1)} & \dots & p_m^{(t)} & \dots
 \end{array} \tag{1.3.1}$$

in cui con s_i si indicano i possibili stati del sistema ($i = 1, 2, \dots, m$) e con $p_i^{(t)}$ la probabilità che al tempo t il sistema si trovi nello stato s_i .

Si fa osservare che se le variabili casuali X_k fossero indipendenti tra di loro e gli stati equiprobabili, saremmo nel caso lancio di un dado a m facce infinite volte, la distribuzione di probabilità ad ogni istante sarebbe nota; probabilità uguali fra loro e tutte pari a $1/m$.

Nel caso in cui il processo soddisfi alla proprietà di Markov, invece, siamo in grado di determinare tutte le distribuzioni $p^{(t)}$ se, come vedremo, è nota la distribuzione ad un istante t e sono note le matrici di transizione del processo,² che andiamo a definire.

Supponiamo di conoscere la distribuzione del processo al tempo k ; sia $p_j^{(k)} = P\{X_k = j\}$ la probabilità che il sistema al tempo k si trovi nello stato j .

Dimostriamo che la probabilità che il sistema si trovi nello stato i al tempo $k + 1$ è data da:

$$p_i^{(k+1)} = \sum_j P\{X_{k+1} = i | X_k = j\} p_j^{(k)} \quad (1.3.2)$$

è cioè la somma delle probabilità di tutti i percorsi che portano allo stato i al tempo $k + 1$, passando per lo stato j al tempo k . Infatti:

la probabilità che il sistema si trovi nello stato i al tempo $k + 1$ è data da:

$$p_i^{(k+1)} = \sum_{\xi_1, \xi_2, \dots, \xi_k} P\{X_1 = \xi_1, X_2 = \xi_2, \dots, X_k = \xi_k, X_{k+1} = i\} \quad (1.3.3)$$

ma

$$\begin{aligned} & P\{X_1 = \xi_1, X_2 = \xi_2, \dots, X_k = \xi_k, X_{k+1} = i\} = \\ & = P\{X_{k+1} = i | X_1 = \xi_1, X_2 = \xi_2, \dots, X_k = \xi_k\} \cdot \\ & \cdot P\{X_1 = \xi_1, X_2 = \xi_2, \dots, X_k = \xi_k\} \end{aligned} \quad (1.3.4)$$

essendo il processo di Markov

$$\begin{aligned} & P\{X_{k+1} = i | X_1 = \xi_1, X_2 = \xi_2, \dots, X_k = \xi_k\} = \\ & = P\{X_{k+1} = i | X_k = \xi_k\} \end{aligned} \quad (1.3.5)$$

²la matrice di transizione, in generale, varia da istante a istante.

e allora possiamo scrivere che

$$\begin{aligned}
p_i^{(k+1)} &= \sum_{\xi_k} \sum_{\xi_1 \xi_2 \dots \xi_{k-1}} \cdot \\
&\cdot P\{X_{k+1} = i | X_k = \xi_k\} P\{X_1 = \xi_1, X_2 = \xi_2, \dots, X_k = \xi_k\} = \\
&= \sum_{\xi_k} P\{X_{k+1} = i | X_k = \xi_k\} \cdot \\
&\cdot \sum_{\xi_1, \xi_2, \dots, \xi_{k-1}} P\{X_k = \xi_k, X_{k-1} = \xi_{k-1}, \dots, X_1 = \xi_1\} = \\
&= \sum_{\xi_k} P\{X_{k+1} = i | X_k = \xi_k\} P\{X_k = \xi_k\} \tag{1.3.6}
\end{aligned}$$

Definizione 1.3.1: si chiama *matrice di transizione* $P^{(k)}$ la matrice $m \times m$ i cui elementi sono dati da:

$$p_{ji}^{(k)} = P\{X_{k+1} = i | X_k = j\} \tag{1.3.7}$$

Essa realizza la transizione dalla distribuzione all'istante k a quella all'istante $k + 1$ e precisamente

$$\begin{aligned}
p_i^{(k+1)} &= \sum_j p_{ji}^{(k)} p_j^{(k)} \\
\underline{p}^{(k+1)} &= P^{(k)+} \underline{p}^{(k)} \tag{1.3.8}
\end{aligned}$$

Perciò tutta la storia dell'evoluzione della distribuzione può essere ricostruita con la successione

$$\begin{aligned}
\underline{p}^{(1)} &= P^{(0)+} \underline{p}^{(0)} \\
\underline{p}^{(2)} &= P^{(1)+} P^{(0)+} \underline{p}^{(0)} \\
&\dots \\
\underline{p}^{(k+1)} &= P^{(k)+} P^{(k-1)+} \dots P^{(1)+} P^{(0)+} \underline{p}^{(0)} \tag{1.3.9}
\end{aligned}$$

Le matrici $P^{(k)}$ sono stocastiche, cioè soddisfano le due proprietà:

$$\begin{aligned}
p_{ji}^{(k)} &\geq 0 \\
\sum_i p_{ji}^{(k)} &= P\{\forall X_{k+1} | X_k = j\} = 1 \tag{1.3.10}
\end{aligned}$$

Definizione 1.3.2: una catena di Markov con matrice di transizione costante per ogni k è detta omogenea nel tempo.

In tal caso:

$$\begin{aligned} \underline{p}^{(1)} &= P^{(0)+} \underline{p}^{(0)} \\ \underline{p}^{(2)} &= P^{(1)+} P^{(0)+} \underline{p}^{(0)} = (P^+)^2 \underline{p}^{(0)} \\ &\dots \\ \underline{p}^{(k+1)} &= P^{(k)+} P^{(k-1)+} \dots P^{(1)+} P^{(0)+} \underline{p}^{(0)} = (P^+)^{k+1} \underline{p}^{(0)} \end{aligned} \quad (1.3.11)$$

Osservazione 1.3.1: da un punto di vista fisico il concetto di omogeneità è legato alla invarianza per traslazione temporale della dinamica che governa il sistema descritto dalla catena di Markov.

Così ad esempio se il sistema segue una legge del tipo (1.1.3), se gli ξ_k hanno tutti la stessa distribuzione, la probabilità di passare da $X_k = j$ a $X_{k+1} = i$ è pari alla somma delle probabilità di tutti quei valori di ξ_{k+1} per cui

$$i = j + g(j, \xi_{k+1}) ;$$

poiché g è per ipotesi indipendente da k , anche questa probabilità non ne dipende.

Esempio 1.3.1: ad esempio se

$$X_{k+1} = [X_k + \xi_{k+1}]_{mod 3} ,$$

se l'insieme degli stati è $I = \{0, 1, 2\}$ e se

$$\xi_k \equiv \begin{cases} -1 & 1 \\ p & q \end{cases} ,$$

allora è chiaro che

$$\{P_{ji}^{(k)}\} = \begin{array}{c|ccc} \text{Stati al} & \text{Stati al tempo } k+1 & & \\ \text{tempo } k & i=0 & 1 & 2 \\ \hline j=0 & 0 & p & q \\ & 1 & p & 0 & q \\ & 2 & q & p & 0 \end{array} ;$$

come si vede la matrice di transizione è appunto indipendente da k .

Esiste una maniera grafica di rappresentare le matrici di transizione che vediamo a partire da un esempio.

Esempio 1.3.2: si abbia la seguente matrice di transizione valida ad ogni istante $t = k$:

$$\begin{array}{cc} & \begin{array}{c} 0 \quad 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{array}{cc} p & q \\ q & p \end{array} \end{array}$$

in cui la colonna a sinistra della matrice indica gli stati al tempo k , la riga sopra la matrice indica gli stati al tempo $k + 1$.

La probabilità che il sistema a partire dallo stato 0 al tempo k passi allo stato 0 al tempo $k + 1$ è p (elemento della riga corrispondente allo stato in considerazione al tempo k e della colonna corrispondente allo stato in considerazione al tempo $k + 1$) e così via.

Il grafo corrispondente è il seguente:

Fig. 1.3.1

in cui si indicano con i nodi gli stati possibili del sistema (in questo caso 2) e con le frecce le transizioni, cui si associa la relativa probabilità.

Osservazione 1.3.2: generalizzando l'Esempio 3.2, possiamo ricavare la seguente regola che permette di istituire una corrispondenza biunivoca tra matrici di transizione e grafi orientati:

- nella matrice di transizione consideriamo tutti e soli quegli elementi per cui

$$p_{ji}^{(k)} = P\{X_{k+1} = i | X_k = j\} > 0$$

lasciando sottinteso che gli altri elementi sono necessariamente nulli; così si crea un inventario delle coppie (j, i) (per cui j è uno stato di entrata al tempo k , i è uno stato di uscita al tempo $(k+1)$ tra cui esiste una probabilità positiva di transizione dal tempo k al tempo $k+1$;

- rappresentiamo gli stati del sistema con m punti nel piano, che saranno i nodi del grafo: ad ogni coppia (j, i) facciamo corrispondere una connessione (lato del grafo) orientata con una freccia da j ad i ; ad ogni lato con freccia associamo un numero (intensità di flusso) corrispondente a $p_{ji}^{(k)}$.

La corrispondenza inversa è del tutto evidente.

La rappresentazione così introdotta è certamente espressione del funzionamento fisico dell'evoluzione degli stati; questa è utile particolarmente per catene omogenee in cui la matrice di transizione e quindi il suo grafo è sempre la stessa al passare del tempo.

Esempio 1.3.3: si consideri ad esempio una catena di Markov omogenea con matrice di transizione

$$P = \begin{pmatrix} 0.5 & 0.25 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix}$$

la sua rappresentazione mediante un grafo orientato sui suoi 7 stati è

Fig. 1.3.2

Fig. 1.3.2: i numeri tra parentesi rappresentano le probabilità di transizione.

1.4 Classificazione degli stati

Definizione 1.4.1: si dice che lo stato j è comunicante con lo stato i se esiste un tempo finito n per cui la probabilità di andare da j a i in n passi è positiva

$$p_{ji}^{(n)} > 0 \quad (1.4.1)$$

Definizione 1.4.2: lo stato i è inessenziale se

$$\exists(j, n) \Rightarrow \begin{cases} p_{ij}^{(n)} > 0 \\ p_{ij}^{(m)} = 0 \end{cases} \quad \forall m \quad (1.4.2)$$

Definizione 1.4.3: se i non è inessenziale allora è essenziale, ovvero

$$\forall j, n \exists m \Rightarrow \begin{cases} p_{ij}^{(n)} > 0 \\ p_{ij}^{(m)} > 0 \end{cases} \quad \text{oppure} \quad \begin{cases} p_{ij}^{(n)} = 0 \\ p_{ij}^{(m)} = 0 \end{cases} \quad (1.4.3)$$

cioè è uno stato a cui, una volta visitato, si ha sempre la possibilità di tornare in un tempo infinito.

Sia \mathcal{S} l'insieme degli stati della catena, esso è l'unione di stati inessenziali ed essenziali

$$\mathcal{S} = I \cup E .$$

Definizione 1.4.4: se i comunica con j e j comunica con i , la coppia (i, j) è detta coppia di stati comunicanti.

Nel sottoinsieme E la probabilità di due stati di essere comunicanti è

riflessiva: (i, i) è comunicante in E , altrimenti i non potrebbe essere essenziale

simmetrica: se i comunica con $j \Rightarrow j$ comunica con i :

Fig. 1.4.1

transitiva: se i comunica con j e j con $k \Rightarrow i$ comunica con k

Fig. 1.4.2

$$\begin{aligned} p_{ik}^{(n+1)} &> p_{ji}^{(n)} p_{jk}^{(1)} > 0 \\ p_{ik}^{(h+m)} &> p_{ki}^{(h)} p_{ji}^{(m)} > 0 \end{aligned} \tag{1.4.4}$$

dunque la relazione binaria di essere stati comunicanti è una relazione di equivalenza.

Pertanto gli stati essenziali si dividono in classi di equivalenza disgiunte

$$E = \bigcup_i E_i \quad E_i \cap E_s = 0 \quad i \neq s$$

Lemma 1.4.1: *non può esistere una catena di Markov con un sistema \mathcal{S} fatto solo di stati inessenziali.*

□ Per assurdo sia \mathcal{S} fatta solo di stati inessenziali.

Fig. 1.4.3

Partiamo da i_1 qualsiasi; deve allora esistere $i_2 \neq i_1$ ed n_1 , tali che

$$\begin{cases} p_{i_1, i_2}^{(n_1)} > 0 \\ p_{i_2, i_1}^{(m)} = 0 \quad \forall m \end{cases} \quad (1.4.5)$$

ora ripartiamo da i_2 : deve esistere $i_3 \neq i_2$ ed n_2 tali che

$$\begin{cases} p_{i_2, i_3}^{(n_2)} > 0 \\ p_{i_3, i_2}^{(m)} = 0 \quad \forall m \end{cases} \quad (1.4.6)$$

Si noti che deve essere anche $i_3 \neq i_1$ perché, per $i_3 = i_1$, la prima delle (1.4.7) risulterebbe in contraddizione con la seconda delle (1.4.6).

Inoltre deve anche essere

$$p_{i_3, i_1}^{(m)} = 0 \quad \forall m \quad (1.4.7)$$

altrimenti

$$p_{i_3, i_2}^{(n_1+m)} > p_{i_3, i_1}^{(m)} p_{i_1, i_2}^{(n_1)} > 0 . \quad (1.4.8)$$

Procedendo in questo modo, si crea una catena che non può mai essere percorsa all'indietro. Pertanto, ripetendo NI passaggi di questo tipo ($NI =$

numero totale di stati), ci si trova in uno stato \bar{i} di I da cui non è possibile uscire

$$P\{X_k = \bar{i} | X_{k-i} = \bar{i}\} = 1 \quad \text{se } k \geq NI \quad (1.4.9)$$

cioè è assorbente e quindi *non inessenziale*. \square

Lemma 1.4.2: *sia data una catena di Markov $\{X_n\}$ omogenea, i cui stati siano divisi nei due sottinsiemi I , degli stati inessenziali, ed E , degli stati essenziali. Ammesso che all'istante iniziale $X_0 \in I$, cioè che $P\{X_0 \in I\} = 1$, allora, quasi certamente, cioè con probabilità $P = 1$, la catena abbandona ad un tempo finito l'insieme degli I , passando definitivamente in E .*

\square L'affermazione significa che

$$P\{X_n \in I, \forall n | X_0 \in I\} = 0 \quad (1.4.10)$$

infatti poiché il ritorno da E in I ha sempre probabilità $P = 0$, si ha che

$$X_n \in I \Rightarrow X_{n-1} \in I \Rightarrow X_{n-2} \in I \dots \quad (1.4.11)$$

cioè

$$\{X_n \in I\} \subset \{X_{n-k} \in I\}, \quad k = 0, \dots, n \quad (1.4.12)$$

mentre

$$X_n \in E \Rightarrow X_{n+1} \in E \Rightarrow X_{n+2} \in E \dots \quad (1.4.13)$$

così che le sole realizzazioni di $\{X_n\}$ che non passano definitivamente in E sono quelle che stanno in I , $\forall n$.

Poiché la (1.4.12) implica che

$$P\{X_n \in I\} \leq P\{X_{n-k} \in I\} \quad (1.4.14)$$

e poiché

$$\begin{aligned} P\{X_n \in I, \forall n\} &= P\{X_n \in I, \forall n | X_0 \in I\} \leq \\ &\leq P\{X_{\bar{n}} \in I\}, \quad \forall \bar{n} \text{ fisso} \end{aligned} \quad (1.4.15)$$

ci basterà dimostrare che $P\{X_{\bar{n}} \in I\}$ è non solo decrescente, ma anche tendente a zero per $\bar{n} \rightarrow +\infty$; anzi, poiché $P\{X_{\bar{n}} \in I\}$ è decrescente, basterà dimostrare che per una sottosuccessione $n_j (n_j \rightarrow +\infty, j \rightarrow +\infty)$

$$\lim_{j \rightarrow +\infty} P\{X_{n_j} \in I\} = 0 \quad (1.4.16)$$

Sia i_0 un qualunque stato di $I, i_0 \in I$ e supponiamo che $x_0 = i_0$; per definizione di stato inessenziale e seguendo il ragionamento del lemma precedente si vede che per un qualche N finito

$$P\{X_N \in E\} = p_{i_0} > 0 \quad (1.4.17)$$

per la (1.4.13) allora si avrà anche

$$P\{X_n \in E, \forall n \geq N\} \geq p_{i_0} \quad (1.4.18)$$

Poiché questo è vero per tutti gli stati $i_0 \in I$ (che sono in numero finito), è chiaro che esisterà un \bar{N} sufficientemente grande ed un \bar{p} abbastanza piccolo, ma sempre $\bar{p} > 0$, tali che

$$P\{X_{\bar{N}} \in E\} = P\{X_{\bar{N}} \in E | X_0 \in I\} \geq \bar{p} \quad (1.4.19)$$

infatti

$$\begin{aligned} P\{X_{\bar{N}} \in E | X_0 \in I\} &= \sum_{i_0 \in I} P\{X_{\bar{N}} \in E | X_0 = i_0\} P\{X_0 = i_0\} \geq \\ &\geq \bar{p} \sum_{i_0 \in I} P\{X_0 = i_0\} = \bar{p} \end{aligned} \quad (1.4.20)$$

Ma allora

$$P\{X_{\bar{N}} \in I\} = P\{X_{\bar{N}} \in I | X_0 \in I\} \leq 1 - \bar{p} = \bar{q} < 1 ; \quad (1.4.21)$$

inoltre, essendo la catena omogenea, dovrà anche essere, $\forall k$

$$P\{X_{\bar{N}+k} \in I | X_k \in I\} \leq \bar{q} < 1 . \quad (1.4.22)$$

Ora

$$P\{X_{2\bar{N}} \in I | X_0 \in I\} = P\{X_{2\bar{N}} \in I | X_{\bar{N}} \in I\} P\{X_{\bar{N}} \in I\} , \quad (1.4.23)$$

poiché

$$P\{X_{2\bar{N}} \in I | X_{\bar{N}} \in E\} = 0 , \quad (1.4.24)$$

così che per la (1.4.21) e la (1.4.22)

$$P\{X_{2\bar{N}} \in I\} = P\{X_{2\bar{N}} \in I | X_0 \in I\} \leq \bar{q}^2 . \quad (1.4.25)$$

Procedendo analogamente si ha pure

$$P\{X_{j\bar{N}} \in I\} \leq \bar{q}^j \rightarrow 0 \text{ per } j \rightarrow \infty \quad (1.4.26)$$

così che la (1.4.16) risulta dimostrata, con $n_j = j\bar{N}$ ed il teorema è provato. \square

Corollario 1.4.1: vale

$$\lim_{n \rightarrow \infty} P\{X_n \in E\} = 1 \quad (1.4.27)$$

Dal Lemma 1.4.2 discende una prima classificazione qualitativa del processo che stiamo analizzando: divisi gli stati nelle due classi I ed E , se il processo nasce in E ($X_0 \in E$) allora vi rimane: se il sistema nasce in I , dopo un tempo più o meno lungo passa in E senza più ritornare in I .

Il seguente lemma permette di specificare meglio il comportamento di $\{X\}$.

Lemma 1.4.3: *se E si decompone in due classi disgiunte non comunicanti E_1, E_2 .*

$$\begin{aligned} \exists \lim_{n \rightarrow \infty} P\{X_n \in E_1\} &= p_1 \\ \exists \lim_{n \rightarrow \infty} P\{X_n \in E_2\} &= p_2 \\ p_1 + p_2 &= 1 \end{aligned} \quad (1.4.28)$$

\square Infatti

$$P\{X_{n+1} \in E_1\} \supset P\{X_n \in E_1\} \quad (1.4.29)$$

perché

$$X_n \in E_1 \Rightarrow X_{n+1} \in E_1 \quad (1.4.30)$$

e quindi le successioni

$$P\{X_n \in E_1\}, P\{X_n \in E_2\} \quad (1.4.31)$$

sono crescenti, limitate superiormente e perciò hanno limite, inoltre

$$\begin{aligned} p_1 + p_2 &= \lim_{n \rightarrow \infty} P\{X_n \in E_1\} + \lim_{n \rightarrow \infty} P\{X_n \in E_2\} = \\ &= \lim_{n \rightarrow \infty} P\{X_n \in E\} = 1 \end{aligned} \quad (1.4.32)$$

□

Ciò implica che il comportamento di evoluzione asintotica di una distribuzione iniziale di probabilità

$$P\{X_0 = i\} = p_i^{(0)} \quad (1.4.33)$$

abbia le seguenti caratteristiche:

- se $X_0 \in I$ dopo un certo tempo n , X_n va in E , inoltre X_n tende a distribuirsi tra le classi irriducibili di E con una certa distribuzione di probabilità asintotica; si osservi che in generale la distribuzione asintotica tra gli E_k dipende dalla distribuzione iniziale (1.4.33)

Fig. 1.4.4

- se invece $X_0 \in E$ allora X_0 appartiene ad una delle classi disgiunte di equivalenza E_k e vi rimane per sempre.

In conclusione, si può dire che il comportamento asintotico di una catena di Markov si riassume nel fatto che prima o poi lo stato della catena cadrà in una delle classi disgiunte degli stati essenziali e ciò avverrà con probabilità calcolabile. Una volta raggiunta una delle E_i ci chiediamo ancora quale sarà il comportamento della catena in tale classe; questo sarà l'oggetto di studio del prossimo paragrafo almeno per alcuni casi relativamente semplici.

Esempio 1.4.1: riprendiamo l'Esempio 1.3.3; è chiaro che 1 è uno stato inessenziale in quanto il sistema quando lo abbandona non può più ritornarvi. Tolto lo stato 1 il grafo di Fig. 1.3.2 si decompone automaticamente in due classi disgiunte

$$E_1 \equiv \{2, 3, 4\}, \quad E_2 \equiv \{5, 6, 7\} .$$

È chiaro che ciascuna delle due classi è fatta da stati essenziali. Al di là del fatto che appare intuitivo per motivi di simmetria che

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{X_n \in E_1\} &= \frac{1}{2} , \\ \lim_{n \rightarrow \infty} P\{X_n \in E_2\} &= \frac{1}{2} , \end{aligned}$$

proviamo a dimostrare anche direttamente che ciò è vero a partire dalla matrice di transizione del processo. A questo scopo è utile osservare che poiché il fenomeno che vogliamo studiare è la transizione $1 \rightarrow E_1$ e $1 \rightarrow E_2$, mentre per ora non ci interessa sapere cosa succede all'interno di ogni classe, è utile costruire uno schema ridotto della catena in cui le classi E_1, E_2 siano condensate in due unici stati. Si perviene così allo schema ridotto di Fig. 1.4.5

Fig. 1.4.5: $p + q + r = 1$

che nel caso dell'Esempio 1.3.2 corrisponde a

$$p = 1/2 \quad , \quad q = r = 1/4 .$$

La matrice di transizione corrispondente al grafo ridotto è dunque

$$P = \begin{vmatrix} p & q & r \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} .$$

Supposto che al tempo zero il sistema si trovasse in 1, che è l'unico caso significativo per noi, così che

$$\underline{p}^{(0)+} = |1 \ 0 \ 0|$$

è facile verificare che (cfr. (1.3.11))

$$\underline{p}^{(n)+} = P^{(0)+} \underline{p}^n = \left| p^n \quad q \left(\sum_{k=0}^{n-1} p^k \right) \quad r \left(\sum_{k=0}^{n-1} p^k \right) \right| .$$

Quindi passando al limite per $n \rightarrow \infty$ si trova direttamente

$$\lim_{n \rightarrow \infty} P\{X_n = 1\} = 0 ; \lim_{n \rightarrow \infty} P\{X_n = 2\} = \frac{q}{1-p} ; \lim_{n \rightarrow \infty} P\{X_n = 3\} = \frac{r}{1-p}$$

che è appunto quanto ci aspettavamo.

1.5 Teoria ergodica per le catene di Markov

Supposto di esserci ristretti ad una catena omogenea costituita da un'unica classe di stati essenziali e comunicanti, ci chiediamo quale sia il comportamento della catena stessa, ovvero come si distribuisca la probabilità tra gli stati essenziali, quando il tempo di evoluzione del sistema tende ad ∞ . Poiché l'evoluzione nel tempo della distribuzione di probabilità tra gli stati di una catena omogenea è legata alla matrice di transizione e alla distribuzione di probabilità iniziale, secondo la (1.3.11), occorrerà studiare come agisce $(P^+)^k$ sui vettori di probabilità quando $k \rightarrow \infty$.

Studiamo questo problema dapprima ponendo la forte ipotesi semplificatrice che la matrice di transizione P (o la sua trasposta che per comodità chiameremo A) sia costituita tutta da elementi positivi, ovvero che tutti gli stati siano comunicanti ad un passo con probabilità positiva; vedremo più oltre come rilassare almeno parzialmente tale ipotesi.

Sia

$$e = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

Sia A la matrice

$$a_{ij} = p_{ji} \quad , \quad 0 < m \leq p_{ij} \leq M < 1 \quad , \quad (1.5.1)$$

allora il fatto che P sia stocastica si traduce nella utile relazione

$$e^+ A = [\dots \sum_i a_{ij} \dots] = [\dots \sum_i p_{ji} = 1 \dots] = e^+ \quad (1.5.2)$$

Sia N il numero degli stati del processo descritto dalla catena di Markov in esame; al tempo k la distribuzione di probabilità tra gli stati sarà un vettore di \mathbb{R}^N . Introduciamo in \mathbb{R}^N una norma (L_1) definita come

$$|r| = \sum_j |r_j| \quad (1.5.3)$$

Come si vede se p è una distribuzione di probabilità, il vettore corrispondente avrà $|p| = 1$ in \mathbb{R}_N secondo tale norma.

Osservazione 1.5.1: se p è una distribuzione di probabilità (cioè $p_i \geq 0$, $e^+p = 1$), allora $q = Ap$ è pure una distribuzione di probabilità.

Infatti

$$e^+q = e^+(Ap) = (e^+A)p = e^+p = 1 \quad (1.5.4)$$

e inoltre

$$q_i = \sum_j a_{ij}p_j = \sum_j p_j p_{ji} \geq 0 \quad (1.5.5)$$

Quindi, poiché la matrice A governa la transizione da un vettore di probabilità ad un altro, avremo che, per un vettore di probabilità p_0 qualsiasi

$$|A^{+\ell}p_0| = |A^{\ell-1}p_0| = \dots = |p_0| = 1. \quad (1.5.6)$$

Lemma 1.5.1: consideriamo il sottospazio $R^\perp = \{v : e^+v = 0\}$

A trasforma R^\perp in se stesso.

□ Infatti

se $e^+v = 0 \rightarrow e^+(Av) = (e^+A)v = e^+v = 0$. □

Ora in maniera particolare ci interessa studiare il comportamento di A quando essa agisce su R^\perp , considerato come sottospazio di \mathbb{R}_N , con la stessa norma.

In generale avremo, sfruttando la proprietà di stocasticità di A ,

$$|Av| = \sum_i \left| \sum_j a_{ij}v_j \right| \leq \sum_i \sum_j |a_{ij}| |v_j| = \sum_j |v_j| \sum_i |a_{ij}| = \sum_j |v_j| \quad (1.5.7)$$

che dimostra che A , come trasformazione lineare di R^1 , è una trasformazione non espansiva, ovvero $|A| \leq 1$. In realtà sfruttando a pieno l'ipotesi (1.5.1) dimostreremo che la A è una contrazione in R^1 .

Lemma 1.5.2: *in R^1 vale*

$$|Av| \leq \rho|v| \quad \rho < 1 \quad (1.5.8)$$

□ Fissato v sono fissati due insiemi di indici I^+, I^- tali che

$$\begin{aligned} v_j &\geq 0 & j \in I^+ \\ v_j &< 0 & j \in I^- \end{aligned} \quad (1.5.9)$$

Si noti che

$$e^+v = 0 \Rightarrow \sum_j v_j = 0 \rightarrow \sum_{I^+} v_j = \sum_{I^-} (-v_j) \quad (1.5.10)$$

e inoltre

$$|v| = \sum_j |v_j| = \sum_{i \in I^+} v_j + \sum_{j \in I^-} (-v_j) = 2 \sum_{j \in I^+} v_j \quad (1.5.11)$$

Fissato v consideriamo

$$u_i = \sum_j a_{ij}v_j = \sum_{j \in I^+} a_{ij}v_j - \sum_{j \in I^-} a_{ij}(-v_j) \quad (1.5.12)$$

sia c_i il segno di u_i

$$\begin{aligned} c_i &= +1 \text{ se } u_i > 0 \\ c_i &= -1 \text{ se } u_i < 0 \end{aligned}$$

allora

$$\sum_i |u_i| = \sum_i c_i u_i = \sum_j v_j \sum_i c_i a_{ij} \quad (1.5.13)$$

I c_i non sono tutti concordi poiché $e^+u = 0$, quindi, ricordando la (1.5.1)

$$\sum_i |c_i a_{ij}| \leq 1 - 2m = \rho < 1 \quad (1.5.14)$$

Perciò

$$\sum_i |u_i| \leq \sum_j |v_j| \sum_i c_i a_{ij} \leq \rho \sum_j |v_j| \quad (1.5.15)$$

ovvero, ogni volta che si applica A ad un vettore di R^\perp , la sua norma si contrae di ρ . \square

Ci serviremo di tale risultato per dimostrare il teorema seguente.

Teorema 1.5.1: *(sullo stato limite). Se A è stocastica, se $m = \min p_{ij} > 0$ e se p_0 è una distribuzione iniziale di probabilità*

($\sum_i p_{0i} = 1, p_{0i} \geq 0$) allora $A^n p_0$ ammette limite

$$\lim_{n \rightarrow \infty} A^n p_0 = \bar{p} \quad (1.5.16)$$

tale che \bar{p} è una distribuzione di probabilità, ovvero ($\bar{p}_i \geq 0, \sum_i \bar{p}_i = 1$); il limite (1.5.16) va inteso nel senso della norma di \mathbb{R}^N , ovvero in senso ordinario per le componenti.

\square 1. La differenza tra due distribuzioni di probabilità $p - q$ è sempre in R^\perp .

Infatti

$$e^+(p - q) = e^+p - e^+q = 0 \quad (1.5.17)$$

2. Consideriamo la differenza

$$A^{n+\ell} p_0 - A^n p_0 = A^n (A^\ell - I) p_0 \quad (1.5.18)$$

e notiamo che

$$v = (I - A^\ell) p_0 = p_0 - A^\ell p_0 \in R^\perp \quad (1.5.19)$$

Perciò, ricordando la (1.5.6) ed osservando che

$$|v| \leq |p_0| + |A^\ell p_0| \leq 2|p_0| = 2 \quad \forall \ell \quad (1.5.20)$$

si ha

$$|A^{n+\ell} p_0 - A^n p_0| = |A^n v| \leq \rho |A^{n-1} v| \leq \dots \leq \rho^n |v| \leq 2\rho^n \xrightarrow[n \rightarrow \infty]{} 0 \quad (1.5.21)$$

\square

Corollario 1.5.1: la distribuzione \bar{p} è temporalmente invariante.

□ Infatti

$$A\bar{p} = A \lim_{n \rightarrow \infty} A^n p_0 = \lim_{n \rightarrow \infty} A^{n+1} p_0 = \bar{p} \quad (1.5.22)$$

□

Ciò significa che se il processo, ad esempio al tempo 0 ha una distribuzione come \bar{p} , allora, in tutti gli istanti successivi, manterrà la stessa distribuzione.

Corollario 1.5.2: la distribuzione invariante \bar{p} è unica.

□ Infatti

se \bar{p}_1, \bar{p}_2 sono invarianti

$$A^n(\bar{p}_1 - \bar{p}_2) = \bar{p}_1 - \bar{p}_2 \quad (1.5.23)$$

inoltre poiché $\bar{p}_1 - \bar{p}_2 \in R^\perp$, per il Lemma 1.5.2

$$A^n(\bar{p}_1 - \bar{p}_2) \xrightarrow[n \rightarrow \infty]{} 0 \quad (1.5.24)$$

quindi

$$\bar{p}_1 = \bar{p}_2 \quad (1.5.25)$$

□

Dunque \bar{p} è un attrattore stabile per il sistema evolutivo descritto da $A^n p_0$.

Si noti che fin qui abbiamo supposto che valesse l'ipotesi

$$a_{ij} = p_{ji} \quad 0 < m \leq p_{ji} \leq M < 1 ; \quad (1.5.26)$$

per ottenere una così forte restrizione dimostriamo il seguente teorema:

Teorema 1.5.2: *condizione sufficiente affinché $A^n p_0$ tenda ad una distribuzione stazionaria è che*

$$(A^k)_{ij} \geq m > 0 \quad (1.5.27)$$

per qualche k fissato.

□ In effetti, sfruttando il Teorema 1.5.1 sullo stato limite, si ha immediatamente

$$\lim_{n \rightarrow \infty} (A^k)^n p_0 = \bar{p} ; \quad (1.5.28)$$

vogliamo però qui dimostrare che è proprio

$$\lim_{n \rightarrow \infty} A^n p_0 = \bar{p} . \quad (1.5.29)$$

Infatti

$$(A^{n+\ell} - A^n)p_0 = A^n(A^\ell - I)p_0 = A^{km} A^\delta v \quad (1.5.30)$$

avendo posto

$$n = km + \delta, \quad (\delta < k) \quad \text{e} \quad v = (A^\ell - I)p_0 \in R^\perp \quad (1.5.31)$$

Essendo $A^\delta v \in R^\perp$

$$|v| \leq 2, \quad |A|^\delta \leq 1 \quad (1.5.32)$$

e per il Lemma 1.5.2

$$|A^{km}|_{R^\perp} = |(A^k)^m|_{R^\perp} \leq \rho^m \quad (1.5.33)$$

si trova:

$$|(A^{n+\ell} - A^n)p_0| \leq \rho^m |A|^\delta |v| \xrightarrow[n \rightarrow \infty]{} 0 \quad (1.5.34)$$

□

Osservazione 1.5.2: la condizione sopra citata è solo sufficiente e comunque vi sono dei casi in cui non è soddisfatta, come dimostra il seguente semplice esempio

Fig. 1.5.1

con matrice di transizione $P = A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$; infatti in tal caso $A^{2n+1} = A$ e $A^{2n} = I$ cosicché tale condizione non è soddisfatta per alcuna potenza di A .

Si noti che in questo caso, partendo da uno stato iniziale $\begin{vmatrix} p \\ q \end{vmatrix}$, ($p + q = 1$) si trova alternativamente $p_{2n} = \begin{vmatrix} p \\ q \end{vmatrix}$, $p_{2n+1} = \begin{vmatrix} q \\ p \end{vmatrix}$, che quindi non dà in generale una successione convergente, tranne che se $p = q = 1/2$. In questo caso lo stato $p = q = 1/2$ è anche lo stato invariante per A .

Lemma 1.5.3: \bar{p} può essere caratterizzato oltre che dalla relazione $A\bar{p} = \bar{p}$ (le componenti \bar{p}_i sono proporzionali ai minori della prima riga di $A - I$), anche dalla seguente notevole proprietà

$$\lim_{n \rightarrow \infty} A^n = \bar{p}e^+ \quad (1.5.35)$$

□ Infatti sia p_0 qualunque (normalizzato con la condizione $\sum_i p_{0i} = 1$)

allora

$$\begin{aligned} A^n p_0 &= A^n(I - \bar{p}e^+)p_0 + A^n\bar{p}e^+p_0 = A^n(p_0 - \bar{p}) + \bar{p}e^+p_0 \rightarrow \\ &\rightarrow \bar{p}e^+p_0 (= \bar{p}) \end{aligned} \quad (1.5.36)$$

perché $e^+p_0 = 1$, $p_0 - \bar{p} \in R^\perp$ e $A^n(p_0 - \bar{p}) \rightarrow 0$.

Quindi

$$A^n \rightarrow \bar{p}e^+ = \begin{bmatrix} \bar{p}_1 & \bar{p}_1 & \cdot & \cdot \\ \bar{p}_2 & \bar{p}_2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \bar{p}_n & \bar{p}_n & \cdot & \cdot \end{bmatrix} \quad (1.5.37)$$

Poiché A è il trasposto della matrice di transizione, saranno le righe di quest'ultima, elevata alla n , a tendere a \bar{p} . □

Teorema 1.5.3: (teorema ergodico o legge dei grandi numeri).

Sia $X = \{X_0, X_1, \dots\}$ una catena di Markov con matrice di transizione $P = A^+$, che ammette un unico stato limite; sia

$$\nu_i^N = \text{numero degli } X_k = i \text{ tra i primi } N ;$$

allora risulta

$$\lim_{N \rightarrow \infty} \frac{\nu_i^N}{N} = \bar{p}_i \quad (1.5.38)$$

che rappresenta la probabilità dello stato i nella distribuzione limite \bar{p} : il limite va intero in media quadratica, cioè

$$\lim_{N \rightarrow \infty} E \left\{ \frac{\nu_i^N}{N} \right\} = \bar{p}_i \quad (1.5.39)$$

$$\lim_{N \rightarrow \infty} \sigma^2 \left\{ \frac{\nu_i^N}{N} \right\} = 0$$

□ 1. Infatti ricordiamo che (teorema di Cesaro)

$$\lim_{n \rightarrow \infty} a_n = a \Rightarrow \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i^n a_n = a \quad (1.5.40)$$

Notiamo poi che si ha

$$\nu_i^N = \sum_{k=0}^N \chi_i(x_k) \quad (1.5.41)$$

dove

$$\chi_i(x_k) = \begin{cases} 1 & \text{se } x_k = i \\ 0 & \text{se } x_k \neq i \end{cases} \quad (1.5.42)$$

Inoltre vale la relazione

$$E\{\chi_i(X_k)\} = 1 \cdot P\{X_k = i\} ; \quad (1.5.43)$$

Ora sia

$$e_i = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 1 \\ \cdot \\ 0 \\ 0 \end{pmatrix} \leftarrow i$$

allora

$$P\{X_k = i\} = e_i^+ A^k p_0 \quad (1.5.44)$$

cosicch 

$$\lim_{k \rightarrow \infty} E\{\chi_i(X_k)\} = \lim_{k \rightarrow \infty} e_i^+ A^k p_0 = e_i^+ \lim_{k \rightarrow \infty} A^k p_0 = e_i^+ \bar{p} = \bar{p}_i \quad (1.5.45)$$

Quindi

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^n E\{\chi_i(X_k)\} = \bar{p}_i \quad (1.5.46)$$

cio 

$$\lim_{N \rightarrow \infty} E\left\{\frac{\nu_i^N}{N}\right\} = \bar{p}_i \quad (1.5.47)$$

2. Dimostriamo ora la seconda della (1.5.30); a tal fine usiamo la relazione

$$\sigma^2 \left\{ \frac{\nu_i^N}{N} \right\} = E \left\{ \frac{(\nu_i^N)^2}{N^2} \right\} - E \left\{ \frac{\nu_i^N}{N} \right\} \quad (1.5.48)$$

Calcoliamo

$$\begin{aligned} E \left\{ \frac{(\nu_i^N)^2}{N^2} \right\} &= \frac{1}{N^2} \sum_{\substack{k=0 \\ \ell=0}}^N E\{\chi_i(X_k)\chi_i(X_\ell)\} = \\ &= \frac{1}{N^2} \sum_{k=0}^N E\{\chi_i(X_k)^2\} + \frac{2}{N^2} \sum_{k=0}^N \sum_{\ell=k+1}^N E\{\chi_i(X_k)\chi_i(X_\ell)\} \end{aligned} \quad (1.5.49)$$

ma per la (1.5.33), essendo $\chi_i^2 = \chi_i$,

$$E\{\chi_i(X_k)^2\} = E\{\chi_i(X_k)\} = P\{X_k = i\} \rightarrow \bar{p}_i \quad (1.5.50)$$

perci 

$$\frac{1}{N} \left(\frac{1}{N} \sum_{k=0}^N E\{\chi_i(x_k)^2\} \right) \approx \frac{\bar{p}_i}{N} (N \rightarrow \infty) \quad (1.5.51)$$

Inoltre, osservando che $\chi_i(j) = \delta_{ij}$ per la (1.5.42),

$$\begin{aligned}
& E\{\chi_i(X_k)\chi_i(X_\ell)\} = E_{X_k}\{E\{\chi_i(X_k)\chi_i(X_\ell)_K\}\} \cdot \\
& \cdot \sum_j E\{\chi_i(X_k)\chi_i(X_\ell)|X_k = j\}P(X_k = j) \cdot \\
& \cdot \sum_j E\{\chi_i(x_\ell)|X_k = j\}\chi_i(j)P(X_k = j) = E\{\chi_i(X_\ell)|X_k = i\} \cdot \\
& \cdot P\{X_\ell = i|X_k = i\}P\{X_k = i\} = e_i^+ A^{1-k} e_i e_i^+ A^k p_0
\end{aligned} \tag{1.5.52}$$

Ma allora

$$\begin{aligned}
& \sum_{\ell=k+1}^N E\{\chi_i(X_k)\chi_i(X_\ell)\} = e_i^+ A^k p_0 \cdot \\
& \cdot \sum_{\ell=k+1}^N e_i^+ A^{\ell-k} e_i = e_i^+ A^k p_0 (N - k) e_i^+ \left(\frac{1}{N - k} \sum_{s=1}^{N-k} A^s \right) e_i
\end{aligned} \tag{1.5.53}$$

Dunque si ha, applicando il teorema di Cesaro e ricordando che $A^n \rightarrow \bar{p}e^+$,

$$C_{N-k} = e_i^+ \left(\frac{1}{N - k} \sum_{s=1}^{N-k} A^s \right) e_i \underset{N \rightarrow \infty}{\approx} e_i^+ \bar{p}e^+ e_i = \bar{p}_i \tag{1.5.54}$$

inoltre

$$B_k = e_i^+ A^k p_0 \rightarrow \bar{p}_i \text{ per } k \rightarrow \infty \tag{1.5.55}$$

ed entrambe le successioni sono limitate.

In questa situazione è facile vedere che

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \frac{2}{N^2} \sum_{k=1}^N B_k C_{N-k} (N - k) = \lim_{N \rightarrow \infty} \frac{2}{N^2} \sum_{k=1}^N \bar{p}_i^2 (N - k) = \\
& \bar{p}_i^2 \lim_{N \rightarrow \infty} \frac{2}{N^2} \frac{N(N + 1)}{2} = \bar{p}_i^2
\end{aligned} \tag{1.5.56}$$

Quindi si ha:

$$\sigma^2 \left\{ \frac{\nu_i^N}{N} \right\} = E \left\{ \frac{(\nu_i^N)^2}{N^2} \right\} - E \left\{ \frac{\nu_i^N}{N} \right\}^2 \approx \bar{p}_i^2 - \bar{p}_i^2 = 0 \tag{1.5.57}$$

□

Essenzialmente questo teorema afferma che, valutando su una realizzazione unica la percentuale di occupazione dello stato i su N tempi e lasciando tendere $N \rightarrow \infty$, si ritrova la probabilità limite \bar{p}_i .

Poiché (1.5.37) e (1.5.38) implicano che $\frac{\nu_i^N}{N} \rightarrow \bar{p}_i$ in probabilità, nel caso che per ipotesi $\bar{p}_i \neq 0$ si avrà anche $\left(\frac{\nu_i^N}{N}\right)^{-1} \rightarrow (\bar{p}_i)^{-1}$; d'altro canto il limite di $\frac{N}{\nu_i^N}$ può essere interpretato come tempo medio di ritorno del sistema allo stato i , perciò, quanto dimostrato, si può anche esprimere dicendo che il tempo di ritorno allo stato i , su tutta una realizzazione, è pari all'inverso della probabilità asintotica che il sistema si trovi in i .

Osservazione 1.5.3: il teorema ergodico che abbiamo qui dimostrato è naturalmente strettamente imparentato con la legge (debole) dei grandi numeri, che afferma che, dato un campione Bernoulliano di numerosità N , la frequenza relativa $f_N(I)$ di estrazioni appartenenti ad un intervallo I , tende in media quadratica alla probabilità teorica $p(I)$. La notevole differenza tra questa legge e quanto dimostrato sopra sta nel fatto che per un campione Bernoulliano le variabili X_1, X_2, \dots sono tutte indipendenti tra loro e tutte uguali in distribuzione, mentre nel caso trattato qui le variabili non sono tra loro indipendenti né ugualmente distribuite.

Esempio 1.5.1: concludiamo il caso iniziato con l'Esempio 1.3.3, cercando la distribuzione asintotica, se esiste, all'interno delle due classi essenziali E_1, E_2 che sono chiaramente tra loro equivalenti. Per semplicità dunque ci riduciamo ad una catena con tre stati con grafico e matrice di transizione P mostrati in Fig. 1.5.2.

Fig. 1.5.2

$$P = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{vmatrix}$$

In primo luogo ci si chiede se è possibile applicare il Teorema 1.5.2. In effetti non è difficile vedere che da P fino a P^4 restano in P^n degli elementi di matrice nulli; ad esempio partendo da 1 è chiaro che in 4 mosse la probabilità di tornare a 1 è nulla.

Tuttavia è

$$P^5 = \begin{vmatrix} 1/4 & 1/4 & 1/8 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 1/4 & 3/8 \end{vmatrix}$$

i cui elementi sono tutti positivi così che vale la condizione (1.5.27). In tal caso esiste la distribuzione asintotica $\underline{\bar{p}}^+ = |\bar{p}_1 \bar{p}_2 \bar{p}_3|$, tale che

$$\underline{\bar{p}}^+ = \underline{\bar{p}}^+ P ;$$

ne segue che la distribuzione asintotica è data da

$$\underline{\bar{p}} = \frac{1}{5} \begin{vmatrix} 1 \\ 2 \\ 2 \end{vmatrix} .$$

1.6 Catene di nascita o morte

Una catena di nascita o morte è una catena di Markov con probabilità di transizione per uno stato intermedio i , rappresentata dal grafico:

Fig. 1.6.1

ovvero tale che ogni stato comunichi solo con se stesso o con quelli adiacenti. Agli estremi di questa catena si possono avere comportamenti diversi, tutti compatibili con tale proprietà di base.

1.

Fig. 1.6.2

la connessione è ancora di tipo locale cioè limitata al più vicino; nel caso che $r_a = 0$ deve essere $p_a = 1$, il che significa che la barriera è “riflettente”

2. caso “circolare”

i punti a e b possono essere disposti in un cerchio su cui si evidenzia che la transizione è sempre al più vicino

Fig. 1.6.3

3.

Fig. 1.6.4

Lo stato del sistema, raggiunto l'estremo a vi si ferma con $P = 1$.

La barriera a è allora detta "assorbente".

Osservazione 1.6.1: se in una catena vi è un solo estremo assorbente e gli stati sono tutti comunicanti, lo stato limite finisce nell'estremo assorbente con $P = 1$; cioè

$$\begin{aligned} & \sum_M P\{X_m = a \text{ se } m \geq M; X_m \neq a \text{ se } m < M\} = \\ = & P\{X \text{ raggiunga } a \text{ in un qualunque tempo finito}\} = 1 \end{aligned}$$

infatti in questo caso tutti gli stati sono inessenziali tranne a .

Osservazione 1.6.2: se vi sono due estremi a, b assorbenti, questi costituiscono classi disgiunte, mentre tutti gli altri stati sono inessenziali.

Si vogliono trovare le probabilità del sistema di essere intrappolato in a o in b (che chiameremo rispettivamente α e β), essendo partito dallo stato x .

Si indichi con $B_m(x)$ e $A_m(x)$ l'insieme delle realizzazioni in cui il sistema arriva rispettivamente allo stato b e a per la prima volta all'istante m essendo partito dallo stato x :

$$\begin{aligned} B_m(x) &= \{\underline{x}; x_m = b; a < x_k < b \quad k < m | x_0 = x\} \\ A_m(x) &= \{\underline{x}; x_m = a; a < x_k < b \quad k < m | x_0 = x\} \end{aligned} \quad (1.6.1)$$

Notiamo che B_m è disgiunto da $B_{m+1}B_{m+2}$ e lo stesso vale per A_m in quanto le singole realizzazioni x sono tutte diverse tra loro; pertanto posto

$$\begin{aligned} B(x) &= \bigcup_{m=1}^{+\infty} B_m(x) = \lim_{M \rightarrow \infty} \bigcup_{m=1}^M B_m(x) = \lim_{M \rightarrow \infty} B^M(x) \\ A(x) &= \bigcup_{m=1}^{+\infty} A_m(x) = \lim_{M \rightarrow \infty} \bigcup_{m=1}^M A_m(x) = \lim_{M \rightarrow \infty} A^M(x) \end{aligned} \quad (1.6.2)$$

le probabilità cercate

$$\begin{aligned}\beta(x) &= P \left[\bigcup_{m=1}^{\infty} \{ \underline{x}; x_m = b; a < x_k < b \text{ se } k < m | x_0 = x \} \right] \\ \alpha(x) &= P \left[\bigcup_{m=1}^{\infty} \{ \underline{x}; x_m = a; a < x_k < b \text{ se } k < m | x_0 = x \} \right]\end{aligned}\quad (1.6.3)$$

saranno date da

$$\begin{aligned}\beta(x) &= P\{B(x)\} = \lim_{M \rightarrow \infty} P\{B^M(x)\} = \lim_{M \rightarrow \infty} \beta_M(x) \\ \alpha(x) &= P\{A(x)\} = \lim_{M \rightarrow \infty} P\{A^M(x)\} = \lim_{M \rightarrow \infty} \alpha_M(x)\end{aligned}\quad (1.6.4)$$

Ma

$$\begin{aligned}\beta_M(x) &= P(B^M) = \\ &= P\{a < x_j < b, 1 \leq j < m; \\ &\quad x_m = b, m = 1, 2, \dots, M | x_0 = x\} = \sum_{m=1}^M P(B_m) = \\ &= \sum_y P\{a < x_j < b; 1 \leq j < m; \\ &\quad x_m = b, 1 \leq m \leq M | x_0 = x, x_1 = y\} P_{xy}\end{aligned}\quad (1.6.5)$$

dove P_{xy} è la probabilità di passare allo stato y al tempo 1 essendo partiti dallo stato x al tempo 0.

Per la proprietà di Markov, si ha

$$\begin{aligned}&P\{a < x_j < b; 1 \leq j < m; \\ &\quad x_m = b \mid 1 \leq m \leq M | x_0 = x, x_1 = y\} = \\ &= P\{a < x_j < b; 2 \leq j < m; \\ &\quad x_m = b, 2 \leq m \leq M | x_1 = y\} = \beta_{M-1}(y)\end{aligned}\quad (1.6.6)$$

Quindi

$$\beta_M(x) = \sum_y P_{xy} \beta_{M-1}(y)\quad (1.6.7)$$

e per $M \rightarrow \infty$

$$\beta(x) = \sum_y P_{xy} \beta(y) \quad (1.6.8)$$

ovvero β è invariante per la matrice di transizione e lo stesso vale per α .

Osservazione 1.6.3: si noti che $\alpha(x)$ e $\beta(x)$ non hanno nulla a che vedere con lo stato limite \bar{p} che è invece invariante per la matrice $A = P^+$.

Osservazione 1.6.4: si osservi che vale

$$\alpha(x) + \beta(x) = 1 \quad \forall x \quad (1.6.9)$$

purché gli stati siano tutti connessi; infatti in tal caso tutti gli stati tranne a e b sono inessenziali.

Questa relazione, che vale per una qualunque catena con due stati assorbenti, ci permette di ridurre la ricerca di $\alpha(x), \beta(x)$ ad una sola delle due funzioni, ad esempio $\beta(x)$ usando la (1.6.8) ed il fatto che $\beta(a) = 0, \beta(b) = 1$ per definizione.

Esempio 1.6.1: applichiamo il ragionamento sopra visto ad una catena omogenea di vita o morte, la cui matrice di transizione quando si cataloghi a come primo stato b come ultimo, è data da

$$P = \begin{vmatrix} 1 & 0 & 0 & 0 \\ q & r & p & 0 \\ 0 & q & r & p \\ & & & 0 & q & r & p \\ & & & 0 & 0 & 0 & 1 \end{vmatrix}.$$

In tal caso, ricordando la (1.6.8), vale

$$\beta(x) = p\beta(x+1) + r\beta(x) + q\beta(x-1) \quad a < x < b \quad (1.6.10)$$

$$\beta(a) = 0 \quad (\text{se la catena parte in } a \text{ non ne esce}) \quad (1.6.11)$$

$$\beta(b) = 1 \quad (\text{se la catena parte in } b \text{ non ne esce ed è sicuro che rimanga in } b) \quad (1.6.12)$$

Posto $r = 1 - p - q$ si trova

$$p[\beta(x+1) - \beta(x)] = q[\beta(x) - \beta(x-1)] \quad (1.6.13)$$

così che posto

$$\Delta_x = \beta(x+1) - \beta(x), \rho = \frac{q}{p} \quad (1.6.14)$$

si ha semplicemente

$$\Delta_x = \rho \Delta_{x-1} \quad (1.6.15)$$

Poiché per $x = a$ si ha

$$\Delta_a = \beta(a+1) - \beta(a) = \beta(a+1) \quad (1.6.16)$$

posto successivamente $x = a+1, a+2 \dots$ si trova

$$\begin{aligned} \Delta_{a+1} &= \rho \Delta_a = \rho \beta(a+1) \\ \Delta_{a+2} &= \rho^2 \Delta_a = \rho^2 \beta(a+1) \\ &\dots \\ \Delta_{b-1} &= \rho^{b-a-1} \Delta_a = \rho^{b-a-1} \beta(a+1) \end{aligned} \quad (1.6.17)$$

ovvero vale la relazione ricorsiva generale

$$\beta(x+1) - \beta(x) = \rho^{x-a} \Delta_a = \rho^{x-a} \beta(a+1) \quad (1.6.18)$$

Poiché

$$\begin{aligned} \beta(x+1) &= \beta(x+1) - \beta(x) + \beta(x) - \beta(x-1) + \dots + \\ &+ \beta(a+1) - \beta(a) = \\ &= \rho^{x-a} \beta(a+1) + \rho^{x-a-1} \beta(a+1) + \dots + \beta(a+1) \end{aligned} \quad (1.6.19)$$

ricordando che $\beta(a) = 0$, si ha che

$$\sum_{k=a}^x \rho^{k-a} \beta(a+1) = \beta(x+1) \quad (1.6.20)$$

applicando questa relazione per $x = b - 1$ e ricordando la (1.6.12), si trova

$$\sum_{k=a}^{b-1} \rho^{k-a} \beta(a+1) = \beta(b) = 1 \quad (1.6.21)$$

da cui discendono le relazioni

$$\beta(a+1) = \frac{1}{\sum_{k=a}^{b-1} \rho^{k-a}} \quad (1.6.22)$$

$$\beta(x+1) = \frac{\sum_{k=a}^x \rho^{k-a}}{\sum_{k=a}^{b-1} \rho^{k-a}} \frac{1 + \rho + \dots + \rho^{x-a}}{1 + \rho + \dots + \rho^{x-a}} = \frac{1 - \rho^{x+1-a}}{1 - \rho^{x+1-a}} \quad (1.6.23)$$

che risolvono completamente il problema posto per β ; α può essere ricavato dalla (1.6.9).

2 PROCESSI STOCASTICI A TEMPI DISCRETI: TRATTAZIONE GENERALE

2.1 Premessa e prime definizioni

In questo capitolo si intende studiare una classe di processi stocastici assai più estesa di quella delle catene di Markov che purtuttavia mantiene una decisiva semplificazione nella descrizione del suo comportamento in quanto il “processo” è rappresentato da una famiglia di variabili casuali (v.c.) ordinate per mezzo di un parametro n discreto

$$\underline{X} = \{X_n, n \in \mathcal{Z}\} . \quad (2.1.1)$$

Qui le singole componenti X_n sono in generale v.c. in \mathbb{R} e non limitate ad un numero discreto e finito di valori come nel caso delle catene.

Inoltre, per il tipo di scopo che si prefigge il presente studio, supporremo sempre che tutte le v.c. da noi usate abbiano media e varianza finite, così da poter usare il formalismo di $\mathcal{L}^2(\Omega)$ descritto nel quaderno Q4.

Dunque, in termini più formali, dato uno spazio di probabilità $\{\Omega, \mathcal{A}, P\}$ consideriamo una successione di v.c.

$$X_n = x_n(\omega), n \in \mathcal{Z} \quad (\text{oppure } n \in \mathcal{Z}^+ \equiv \{n \geq 0\}), \text{ con } X_n \in \mathcal{L}^2(\Omega) ;$$

un processo stocastico a parametro discreto è per noi il vettore $\underline{X} \equiv \{X_n\}$ con valori campionari $\underline{x} \equiv \{x_n\} \in \mathbb{R}^\infty$ e con la distribuzione di probabilità su \mathbb{R}^∞ , P_X definita da

$$P_X(B) = P\{\omega; \underline{x}(\omega) \in B\} \quad (2.1.2)$$

$\forall B \in \mathcal{B}$, l'algebra di Borel in \mathbb{R}^∞ definita come la minima σ -algebra che contiene tutti i cilindri (cfr. Q4 cap. 8).

Osservazione 2.1.1: una volta definita la distribuzione di probabilità P_X in \mathbb{R}^∞ , lo spazio di probabilità $\{\Omega, \mathcal{A}, P\}$ può in tutto e per tutto essere sostituito da $\{\mathbb{R}^\infty, \mathcal{B}, P_X\}$, finché ci si limita a considerare il processo stocastico \underline{X} o sue funzioni, riportandoci così ad una interpretazione di \underline{X} come una generalizzazione di una variabile N -dimensionale \underline{X}_N , passando da \mathbb{R}^N

a \mathbb{R}^∞ . Naturalmente questa operazione farà restringere la nostra attenzione da $\mathcal{L}^2(\Omega)$ a $\mathcal{L}^2(\underline{X})$ che è un sottospazio chiuso del precedente costituito da tutte le v.c. con varianza finita, funzioni di \underline{X} che sono appunto quelle che ci interessano; anzi come vedremo tra breve la nostra analisi, limitandosi all'ambito lineare, si restringerà ad un sottospazio assai più piccolo di $\mathcal{L}^2(\underline{X})$.

Avendo dotato \underline{X} della sua distribuzione P_X in \mathbb{R}^∞ è del tutto naturale definire l'operatore di media E per una $Y = g(\underline{X})$, ovvero

$$E_{\underline{X}}\{Y\} = \int_{\mathbb{R}^\infty} g(\underline{x}) dP_X(\underline{x}) . \quad (2.1.3)$$

Con tale operatore si possono definire i momenti di \underline{X} , ad esempio i momenti del 1° ordine

$$E\{X_i\} = \mu_i \quad (2.1.4)$$

o del 2° ordine

$$E\{X_i X_k\} = C_{ik} , \quad (2.1.5)$$

oppure anche la funzione caratteristica

$$g(\underline{t}) = E\{e^{i\underline{t}^+ \underline{x}}\} \quad (2.1.6)$$

$$(\underline{t}^+ = \{t_n, n \in \mathcal{Z}\}) .$$

Il vettore di \mathbb{R}^∞

$$\underline{\mu} = \{\mu_n, n \in \mathcal{Z}\} \quad (2.1.7)$$

è detto media del processo \underline{X} , mentre la matrice (o operatore, trattandosi di una matrice ∞ -dimensionale)

$$C \equiv \{C_{ik}, (i, k) \in \mathcal{Z} \otimes \mathcal{Z}\} \quad (2.1.8)$$

è detta matrice, o operatore, di covarianza del processo \underline{X} .

Diamo qualche primo esempio di processo stocastico a parametro discreto.

Esempio 2.1.1: il primo e più semplice esempio di processo stocastico è il cosiddetto rumore bianco (in senso stretto) che corrisponde ad una successione $\{X_n, n \in \mathcal{Z}\}$ di v.c. stocasticamente indipendenti ed identicamente distribuite.

Di fatto quando $n \in \mathcal{Z}^+$ questo modello generalizza la variabile campionaria di un campione bernoulliano al caso in cui si facciano infinite estrazioni indipendenti da una stessa v.c. X .

In questo caso è chiaro che

$$E\{X_n\} = \mu \quad (2.1.9)$$

$$E\{X_{n+k}X_n\} = \sigma_X^2 \delta_{k0} ; \quad (2.1.10)$$

dunque la media è un vettore con le componenti costanti, di solito si assume $\mu = 0$, mentre l'operatore di covarianza è proporzionale all'identità.

Esempio 2.1.2: (*processo di Wiener discreto*). Dato un processo di rumore bianco $\{\varepsilon_n\}$ a media nulla e con varianza σ_ε^2 costruiamo $\{X_n, n \in \mathcal{Z}^+\}$ tramite le formule ricorsive

$$\begin{cases} X_{n+1} = X_n + \varepsilon_{n+1} & (n \in \mathcal{Z}^+) \\ X_0 = 0 \end{cases} . \quad (2.1.11)$$

Quindi il senso di questo processo, detto di Wiener, è di essere un integrale (accumulatore) di rumore bianco nel tempo.

Lo schema (2.1.11) può essere risolto esplicitamente per X_{n+1} , dando

$$X_{n+1} = \sum_{k=0}^n \varepsilon_{k+1} . \quad (2.1.12)$$

Da qui si vede che

$$E\{X_{n+1}\} = 0 \quad (2.1.13)$$

$$E\{X_{n+k}X_n\} = n\sigma_\varepsilon^2, k \geq 0 \quad (2.1.14)$$

Si osservi che a causa della definizione (2.1.11) il processo di Wiener $\{X_n\}$ gode della proprietà di Markov

$$E\{X_{n+1}|X_n, X_{n-1} \dots\} = E\{X_{n+1}|X_n\} . \quad (2.1.15)$$

Notiamo anche che per il teorema centrale della statistica ci si può aspettare che X_n tenda a distribuirsi normalmente per $n \rightarrow \infty$.

Esempio 2.1.3: (processo di Poisson discreto). Sia $\{\varepsilon_n\}$ una successione di v.c. indipendenti del tipo testa o croce

$$\varepsilon_n = \begin{cases} 0 & 1 \\ q & p \end{cases} \quad n = 1, 2, \dots \quad (2.1.16)$$

e si ponga

$$X_n = \sum_{k=1}^n \varepsilon_k \quad n = 1, 2, \dots ; \quad (2.1.17)$$

dunque X_n è un contatore del numero di successi nel tempo (variabile) n .

Sarà

$$E\{X_n\} = np \quad (2.1.18)$$

$$E\{X_{n+k}X_n\} = npq, \quad k \geq 0 \quad (2.1.19)$$

Si osservi che in questo esempio, al contrario dei primi due, le v.c. X_n hanno un insieme di valori argomentali discreto; la peculiarità di questo processo è che tale insieme $\{0, 1, \dots, n\}$ varia al variare del tempo n . Per questo motivo, benché $\{X_n\}$ goda della proprietà di Markov, il processo non costituisce una catena di Markov.

Esempio 2.1.4: sia $\omega = U[-1/2, 1/2]$ una v.c. uniformemente distribuita su $[-1/2, 1/2]$, poniamo

$$x_n(\omega) = \sin \pi \omega n \quad n = 1, 2, \dots \quad (2.1.20)$$

Il processo \underline{X} corrispondente ha ovviamente media nulla

$$E\{X_n\} = \int_{-1/2}^{1/2} \sin \pi \omega n \, d\omega = 0 \quad (2.1.21)$$

e covarianza

$$\begin{aligned}
C_{n+k,n} &= E\{X_{n+k}X_n\} = & (2.1.22) \\
&= \frac{1}{2} \int_{-1/2}^{1/2} [\cos \pi \omega k - \cos \pi \omega (2n+k)] d\omega = \\
&= \begin{cases} \frac{1}{2} & k=0 \\ 0 & k=2\ell > 0 \\ \frac{(-1)^\ell}{\pi} \left[\frac{1}{2\ell+1} - \frac{(-1)^n}{2n+2\ell+1} \right] & k=2\ell+1 > 0 \end{cases}
\end{aligned}$$

Esempio 2.1.5: sia ancora $\omega = U[-1/2, 1/2]$ come nell'esempio 2.1.4 e si prenda

$$x_n(\omega) = \sin\left(n\frac{\pi}{2} + 2\pi\omega\right). \quad (2.1.23)$$

Il processo $\underline{X} = \{X_n\}$ ha media nulla

$$E\{X_n\} = \int_{-1/2}^{1/2} \sin\left(n\frac{\pi}{2} + 2\pi\omega\right) d\omega = 0 \quad (2.1.24)$$

e covarianza

$$\begin{aligned}
C_{n+k,n} &= E\{X_{n+k}X_n\} = & (2.1.25) \\
&= \frac{1}{2} \int_{-1/2}^{1/2} \left[\cos\left(k\frac{\pi}{2}\right) - \cos\left(n\pi + k\frac{\pi}{2} + 4\pi\omega\right) \right] d\omega = \\
&= \frac{1}{2} \cos\left(k\frac{\pi}{2}\right) = \begin{cases} (-1)^\ell & k=2\ell \geq 0 \\ 0 & k=2\ell+1 > 0 \end{cases}.
\end{aligned}$$

Esempio 2.1.6: (processi normali). Ricordiamo che dato un qualsiasi vettore $\underline{\mu} \in \mathbb{R}^\infty$ ed un operatore $C \equiv C_{nm}$ in \mathbb{R}^∞ definito positivo nel senso che

$$\begin{aligned}
\forall \underline{\lambda} \in \mathbb{R}^\infty, \quad \underline{\lambda}^+ C \underline{\lambda} &= \sum_n \sum_m \lambda_n \lambda_m C_{nm} < +\infty \\
\Rightarrow \underline{\lambda}^+ C \underline{\lambda} &\geq 0, & (2.1.26)
\end{aligned}$$

³Notiamo esplicitamente che se ad esempio $n, m \in \mathbb{Z}^+$ la condizione $\underline{\lambda}^+ C \underline{\lambda} < +\infty$ significa

$$\underline{\lambda}^+ C \underline{\lambda} = \lim_{N, M \rightarrow \infty} \sum_{n=0}^N \sum_{m=0}^M C_{nm} \lambda_n \lambda_m < +\infty;$$

la modifica di questo limite doppio nel caso che $n, m \in \mathbb{Z}$ è ovvia.

esiste in \mathbb{R}^∞ una distribuzione di probabilità P “normale con media $\underline{\mu}$ e covarianza C ”. La distribuzione P è normale nel senso che, supposto che il parametro n vari in \mathcal{Z}^+ ,

$$\forall N \in \mathcal{Z}^+ , \underline{X}_N \equiv \{X_0, X_1, \dots, X_N\}$$

è una v.c. normale con media $\underline{\mu}_N \equiv \{\mu_0, \mu_1 \dots \mu_N\}$ e con covarianza

$$C_N \equiv \{C_{n,m} ; 0 \leq n, m \leq N\} . \quad (2.1.27)$$

Il processo \underline{X} le cui componenti sono solo le componenti di \underline{x} , ovvero $X_n = x_n(\underline{x})$, è il processo normale con media $\underline{\mu}$ e covarianza C .

Questo schema ha grandi applicazioni quando si sappiano costruire modelli di operatori C che soddisfino la condizione di definita positività (2.1.26); ritorneremo più in là su questo argomento.

In Fig. 2.1 rappresentiamo una realizzazione con n da 0 a 100 per ognuno degli esempi proposti; in particolare il caso (f) corrisponde ad un processo gaussiano con

$$\mu_n = 0 , C_{n+k,n} = e^{-\alpha|k|} , \alpha = 0, 1 . \quad (2.1.28)$$

Naturalmente non è possibile creare un repertorio di esempi che siano rappresentativi di tutti i possibili processi stocastici, tuttavia è utile formulare qualche osservazione in merito al comportamento più o meno stazionario dei diversi processi.

Fig. 2.1.a

Fig. 2.1.b

Con stazionario intendiamo che un sottocampione, ad esempio di numerosità 10, preso in diverse epoche lungo la realizzazione presenta una significativa similarità statistica con gli altri; ad esempio ha circa la stessa media e la stessa varianza.

Fig. 2.1.c

Fig. 2.1.d

Fig. 2.1.e

Fig. 2.1.f

Così vediamo che nel caso (2.1.a), nonostante la realizzazione sia la più irregolare, la regolarità statistica è massima, invece il caso (2.1.b) è chiaramente non stazionario perché costituito da onde che si amplificano nel tempo, mentre (2.1.c) si allontana sistematicamente, cioè anche come media, dall'asse

dell'ascisse. Ancora i casi (2.1.e), (2.1.f) si presentano con un aspetto significativamente stazionario con però la differenza che (2.1.e) sembra più regolarmente periodico di (2.1.f); questa proprietà sono legate al fatto che la media sia costante μ e che la covarianza $C_{n+k,n}$ sia funzione della sola k e non di n , quindi anche la varianza $C_{n,n}$ sia costante, ovvero che la correlazione tra le variabili X_{n+k} ed X_n dipenda solo dalla loro distanza nel tempo k e non dall'epoca n

$$E\{(X_{n+k} - \mu)(X_n - \mu)\} = C_{n+k,n} = C(k) . \quad (2.1.29)$$

È anche importante notare che nel caso (2.1.e) la covarianza è puramente periodica in k mentre nel caso (2.1.f) essa tende a zero quando $k \rightarrow \infty$; questo fatto determina differenze salienti tra i due processi.

2.2 La descrizione hilbertiana dei processi discreti

L'obiettivo che ci proponiamo in questo paragrafo è di caratterizzare in modo compiuto le famiglie di funzioni lineari del processo \underline{X} che siano contenute in $\mathcal{L}^2(\underline{X})$.

Conveniamo di usare per gli elementi di $\mathcal{L}^2(\underline{X})$ la notazione semplice

$$\begin{aligned} X \in \mathcal{L}^2 & \quad \|x\|^2 = E\{X^2\} \\ X, Y \in \mathcal{L}^2 & \quad \langle X, Y \rangle = E\{XY\} , \end{aligned}$$

senza particolari indici per la norma ed il prodotto scalare.

Per comodità facciamo l'ipotesi che

$$\underline{\mu} = E\{\underline{X}\} = 0 ,$$

ciò che in ogni caso non riduce la generalità della discussione successiva.

Notiamo allora che secondo la convenzione fatta possiamo porre

$$\langle X_n, X_m \rangle = C_{nm} , \quad \langle \underline{X}, \underline{X}^+ \rangle = E\{\underline{X} \underline{X}^+\} = C . \quad (2.2.1)$$

Cominciamo osservando che se

$$\underline{\lambda} \in R_0 \subset \mathbb{R}^\infty \Leftrightarrow \underline{\lambda} \equiv \{\lambda_k ; \lambda_k \equiv 0 , |k| > N\} \quad (2.2.2)$$

esiste sempre in $\mathcal{L}^2(\underline{X})$ la v.c.

$$X = \underline{\lambda}^+ \underline{X} = \sum_{-N}^N \lambda_k X_k \in \mathcal{L}^2(\underline{X}) ; \quad (2.2.3)$$

poniamo per definizione

$$H_0 = Span \{X_k, k \in \mathcal{Z}\} = \{X = \underline{\lambda}^+ \underline{X} ; \underline{\lambda} \in R_0\} . \quad (2.2.4)$$

H_0 è ovviamente un sottospazio lineare di $\mathcal{L}^2(\underline{X})$; chiamiamo $H_{\underline{X}}$ (o $H(\underline{X})$) la chiusura di H_0 in $\mathcal{L}^2(\underline{X})$

$$H_{\underline{X}} = [H_0]_{\mathcal{L}^2(\underline{X})} . \quad (2.2.5)$$

Per prima cosa vogliamo caratterizzare l'insieme $\{\underline{\lambda} \in \mathbb{R}^\infty ; \underline{\lambda}^+ \underline{X} \in H_{\underline{X}}\}$; porremo poi

$$H_\ell \equiv \{X \in H_{\underline{X}} ; X = \underline{\lambda}^+ \underline{X}, \underline{\lambda} \in \mathbb{R}^\infty\} . \quad (2.2.6)$$

cioè l'insieme delle variabili in $H_{\underline{X}}$ che sono effettivamente funzioni lineari di \underline{X} .

Osservazione 2.2.1: notiamo fin da ora che H_ℓ sarà un sottospazio lineare di $H_{\underline{X}}$ ma non necessariamente un sottospazio chiuso. Anzi qualora H_ℓ fosse chiuso, deve necessariamente valere

$$H_\ell \equiv H_{\underline{X}}$$

poiché è sempre

$$H_0 \subseteq H_\ell \subseteq H_{\underline{X}} \quad (2.2.7)$$

ed H_0 è per definizione denso in $H_{\underline{X}}$.

Vale il seguente Teorema:

Teorema 2.2.1: sia $\underline{\lambda} \in \mathbb{R}^\infty$ tale che $\underline{\lambda}^+ \underline{X} = \lim_{N \rightarrow \infty} \sum_{k=-N}^N \lambda_k X_k \in H_{\underline{X}}$; allora è anche

$$0 \leq \underline{\lambda}^+ C \underline{\lambda} = \lim_{N, M \rightarrow \infty} \sum_{-N}^N \sum_{-M}^M \lambda_n \lambda_m C_{nm} < +\infty ; \quad (2.2.8)$$

viceversa sia

$$\underline{\lambda} \in H_C \equiv \{ \underline{\lambda} \in \mathbb{R}^\infty , \underline{\lambda}^+ C \underline{\lambda} < +\infty \} \quad (2.2.9)$$

allora \exists la v.c. $X = \underline{\lambda}^+ \underline{X} \in \mathcal{L}^2(\underline{X})$ e quindi $X \in H_{\underline{X}}$.

Dunque essenzialmente il teorema afferma che H_ℓ è l'immagine attraverso l'operatore lineare

$$U \underline{\lambda} = X \equiv \underline{\lambda}^+ \underline{X} \quad (2.2.10)$$

di H_C , definito come in (2.2.9).

Nota: per semplicità di scrittura ma senza perdere in generalità supponiamo che $n \in \mathbb{Z}^+$ così che le serie

$$\underline{\lambda}^+ \underline{X} = \lim_{N \rightarrow \infty} \sum_{n=1}^N \lambda_n X_n , \quad \underline{\lambda}^+ C \underline{\lambda} = \lim_{N, M \rightarrow \infty} \sum_{n=1}^N \sum_{m=1}^M \lambda_n \lambda_m C_{nm}$$

abbiamo solo l'estremo superiore variabile.

Inoltre a livello di notazione porremo, per ogni $\underline{\lambda} \in \mathbb{R}^\infty$ dato,

$$(\underline{\lambda}^N)^+ = \{ \lambda_1, \lambda_2, \dots, \lambda_N, 0, 0 \dots \} ; \quad (2.2.11)$$

il simbolo $\underline{\lambda}^N$, che riguarda una successione costruita su un preciso $\underline{\lambda}$ fisso, non va poi confuso con $\{\underline{\lambda}_N\}$ che viceversa indica una qualsiasi successione in \mathbb{R}^∞ di componenti $\{\lambda_{Nk}\}$.

□ Supponiamo dunque che $\underline{\lambda}$ sia tale che

$$X = \underline{\lambda}^+ \underline{X} = \lim_{N \rightarrow \infty} (\underline{\lambda}^N)^+ \underline{X} \in H_{\underline{X}} ; \quad (2.2.12)$$

ciò ha tre conseguenze:

$$\forall N, \|\underline{\lambda}^N\|^2 = C\underline{\lambda}^N \leq a^2 < +\infty, \quad (2.2.13)$$

$$\forall N, M |(\underline{\lambda}^N)C\underline{\lambda}^M| = | \langle \underline{\lambda}^N, \underline{\lambda}^M \rangle | \leq \quad (2.2.14)$$

$$\leq \|\underline{\lambda}^N\| \cdot \|\underline{\lambda}^M\| \leq a^2 < +\infty,$$

$$\forall N > N_\varepsilon \text{ opportuno, } \forall p > 0$$

$$= \|\underline{\lambda}^{N+p} - \underline{\lambda}^N\|^2 = \|\underline{\lambda}^{N+p} - \underline{\lambda}^N\|^2 = \quad (2.2.15)$$

$$= (\underline{\lambda}^{N+p} - \underline{\lambda}^N)C(\underline{\lambda}^{N+p} - \underline{\lambda}^N) < \varepsilon^2 .$$

Ma allora, fissato ε , $\forall N, M > N_\varepsilon$ e $\forall p, q > 0$,

$$\left| \sum_{n=1}^{N+p} \sum_{m=1}^{M+q} \lambda_n \lambda_m C_{nm} - \sum_{n=1}^N \sum_{m=1}^M \lambda_n \lambda_m C_{nm} \right| =$$

$$= |(\underline{\lambda}^{N+p})C\underline{\lambda}^{M+q} - (\underline{\lambda}^N)C\underline{\lambda}^M| =$$

$$| \langle \underline{\lambda}^{N+p}, \underline{\lambda}^{M+q} \rangle - \langle \underline{\lambda}^N, \underline{\lambda}^M \rangle | \leq$$

$$\leq | \langle \underline{\lambda}^{N+p} - \underline{\lambda}^N, \underline{\lambda}^{M+q} \rangle | + | \langle \underline{\lambda}^N, \underline{\lambda}^{M+q} - \underline{\lambda}^M \rangle | \leq$$

$$\leq \|\underline{\lambda}^{N+p} - \underline{\lambda}^N\| \cdot \|\underline{\lambda}^{M+q}\| + \|\underline{\lambda}^N\| \cdot \|\underline{\lambda}^{M+q} - \underline{\lambda}^M\| \leq$$

$$\leq 2a \cdot \varepsilon$$

dove si è usata la diseguaglianza di Schwarz, oltre alle (2.2.13), (2.2.14), (2.2.15); dunque esiste il limite (2.2.8) ovvero $\underline{\lambda} \in H_C$.

Viceversa se $\underline{\lambda} \in H_C$, esiste il limite

$$\lim_{N, M \rightarrow \infty} (\underline{\lambda}^N)C\underline{\lambda}^M = \underline{\lambda}^+C\underline{\lambda}; \quad (2.2.16)$$

ma allora, dall'identità

$$\begin{aligned} \| (\underline{\lambda}^N)^+ \underline{X} - (\underline{\lambda}^M)^+ \underline{X} \|^2 &= (\underline{\lambda}^N)^+ C \underline{\lambda}^N + (\underline{\lambda}^M)^+ C \underline{\lambda}^M - \\ &- 2(\underline{\lambda}^N)^+ C \underline{\lambda}^M , \end{aligned} \quad (2.2.17)$$

prendendo il limite per N ed $M \rightarrow \infty$ ed usando la (2.2.16) si ha

$$\lim_{N, M \rightarrow \infty} \| (\underline{\lambda}^N)^+ \underline{X} - (\underline{\lambda}^M)^+ \underline{X} \|^2 = 2\underline{\lambda}^+ C \underline{\lambda} - 2\underline{\lambda}^+ C \underline{\lambda} = 0 ,$$

così che $\{(\underline{\lambda}^N)^+ \underline{X}\}$ è una successione di Cauchy e dunque $\exists X \in H_{\underline{X}}$ tale che $X = \underline{\lambda}^+ \underline{X} \in H_{\underline{X}} \subset \mathcal{L}^2(\underline{X})$. \square

Notiamo tra l'altro che H_C è uno spazio lineare in quanto

$$\begin{aligned} \underline{\lambda}, \underline{\gamma} \in H_C &\Rightarrow \underline{\lambda}^+ \underline{X}, \underline{\gamma}^+ \underline{X} \in H_{\underline{X}} \Rightarrow (\underline{\lambda} + \underline{\gamma})^+ \underline{X} \in H_{\underline{X}} \Rightarrow \\ &\Rightarrow \underline{\lambda} + \underline{\gamma} \in H_C . \end{aligned}$$

Nella dimostrazione del Teorema 2.2.1 abbiamo più volte usato un'utile identità che segnaliamo sotto forma di Corollario.

Corollario 2.2.1: *l'operatore lineare $U : H_C \rightarrow H_\ell$ definito da (2.2.10) è una isometria se in H_C si introduce la seminorma ⁴*

$$|\underline{\lambda}|_C^2 = \underline{\lambda}^+ C \underline{\lambda} . \quad (2.2.18)$$

\square Infatti se $\underline{\lambda} \in H_C$ e quindi $X \in H_\ell$ è chiaramente

$$\begin{aligned} |\underline{\lambda}|_C^2 &= \underline{\lambda}^+ C \underline{\lambda} = \lim_{N \rightarrow \infty} (\underline{\lambda}^N)^+ C \underline{\lambda}^N = \lim_{N \rightarrow \infty} \| (\underline{\lambda}^N)^+ \underline{X} \|^2 = \\ &= \| \underline{\lambda}^+ \underline{X} \|^2 = \| X \|^2 \geq 0 . \end{aligned} \quad (2.2.19)$$

\square

Dunque l'operatore C è sempre semipositivo.

Definizione 2.2.1: *diciamo che l'operatore C è positivo, $C > 0$, se*

$$\underline{\lambda} \in H_C , C \underline{\lambda} = 0 \Rightarrow \underline{\lambda} = 0 . \quad (2.2.20)$$

⁴Ricordiamo che una seminorma ha le stesse proprietà di una norma senonché non è detto che valga $|\underline{\lambda}|_C = 0, \Rightarrow \underline{\lambda} = 0$.

Si può dimostrare che tale definizione è equivalente a

$$\underline{\lambda} \in H_C, \underline{\lambda}^+ C \underline{\lambda} = 0 \Rightarrow \underline{\lambda} = 0. \quad (2.2.21)$$

Notiamo che la (2.2.20) è consistente in quanto se $\underline{\lambda} \in H_C$, allora $C \underline{\lambda} \in \mathbb{R}^\infty$ per cui ha senso porre $C \underline{\lambda} = 0$. In effetti poiché $\forall k$

$$\underline{\delta}_k = \{\delta_{k\ell}\} = \{0, \dots, 0, \overset{k}{1}, 0 \dots\}^+ \in R_0 \subset H_C$$

esiste finito $\forall k$

$$\underline{\delta}_k^+ C \underline{\lambda} = (C \underline{\lambda})_k,$$

cioè $C \underline{\lambda} \in \mathbb{R}^\infty$.⁵

Osservazione 2.2.2: quando le (2.2.20) o (2.2.21) sono soddisfatte, $|\underline{\lambda}|_C$ data dalla (2.2.18) diventa una vera norma. Inoltre H_C diventa uno spazio pre-Hilbertiano ponendo

$$(\underline{\lambda}, \underline{\gamma})_C = \underline{\lambda}^+ C \underline{\gamma}; \quad (2.2.22)$$

la definizione (2.2.20) è coerente in quanto se $\underline{\lambda}, \underline{\gamma} \in H_C \Rightarrow (\underline{\lambda} + \underline{\gamma}), (\underline{\lambda} - \underline{\gamma}) \in H_C$ e quindi è definita

$$\underline{\lambda}^+ C \underline{\gamma} = \frac{1}{4} \{(\underline{\lambda} + \underline{\gamma})^+ C (\underline{\lambda} + \underline{\gamma}) - (\underline{\lambda} - \underline{\gamma})^+ C (\underline{\lambda} - \underline{\gamma})\}. \quad (2.2.23)$$

Osservazione 2.2.3: se C non è positiva il processo \underline{X} soffre di una patologia che d'ora in avanti escluderemo sempre, ovvero esiste un tempo \bar{n} tale che $X_{\bar{n}}$ sia una funzione lineare di $\{X_k, k \neq \bar{n}\}$. Infatti se $\underline{\lambda}$ è tale che

$$\underline{\lambda} \neq 0, \underline{\lambda}^+ C \underline{\lambda} = \|\underline{\lambda}^+ \underline{X}\|^2 = 0, \quad (2.2.24)$$

$\exists \bar{n}, \lambda_{\bar{n}} \neq 0$ e quindi si può porre

$$X_{\bar{n}} = -\frac{1}{\lambda_{\bar{n}}} \sum_{k \neq \bar{n}} \lambda_k X_k, \quad (2.2.25)$$

⁵La (2.2.21) garantisce che esiste l'operatore C^{-1} , definito sul range di C , così che $C \underline{\lambda} = \underline{\mu} \Rightarrow \underline{\lambda} = C^{-1} \underline{\mu}$; tale relazione però non ci dice che C^{-1} è esprimibile come una matrice; ciò è vero solo se il range di C contiene R_0 .

la serie a secondo membro essendo ancora ovviamente convergente. La (2.2.25) ci dice che conoscendo $X_k, k \neq \bar{n}$ per una certa realizzazione è possibile predire quasi certamente il valore di $X_{\bar{n}}$.

Poiché a noi interessa solo il caso in cui questa prevedibilità quasi certa non è possibile, escluderemo sempre che possa valere la (2.2.24).

Notiamo anche che (2.2.25) è conseguenza del fatto che

$$X_{\bar{n}} \in [Span \{X_k, k \neq \bar{n}\}]_{H_{\underline{X}}} ; \quad (2.2.26)$$

questa patologia può verificarsi anche in casi in cui $C > 0$, come vedremo nel seguente controesempio, perciò sarà proprio la (2.2.26) che considereremo come condizione che non si verifichi mai.

Esempio 2.2.1: sia $\{\xi_n, n \geq 0\}$ un processo di rumore bianco con varianza unitaria; poniamo

$$X_0 = \xi_0, \quad X_k = \xi_0 + \xi_k .$$

L'operatore di covarianza di \underline{X} è

$$C = \begin{vmatrix} 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot \\ 1 & 2 & 1 & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & 2 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}$$

così che

$$C_{\underline{\lambda}} = \begin{vmatrix} \lambda_0 + \sum_{k=1}^{+\infty} \lambda_k \\ (\lambda_0 + \sum_{k=1}^{+\infty} \lambda_k) + \lambda_1 \\ (\lambda_0 + \sum_{k=1}^{+\infty} \lambda_k) + \lambda_2 \\ \dots \end{vmatrix} = 0$$

implica

$$\lambda_0 + \sum_{k=1}^{+\infty} \lambda_k = 0 \Rightarrow \lambda_1 = \lambda_2 = \dots = 0 \Rightarrow \lambda_0 = 0 \Rightarrow \underline{\lambda} = 0 .$$

Dunque C è positiva e tuttavia

$$X_o \in [Span \{X_k, x \geq 1\}]_{H_{\underline{X}}}$$

in quanto

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N X_k = \lim_{N \rightarrow \infty} X_0 + \frac{1}{N} \sum_{k=1}^N \xi_k = X_0 .$$

Da questo controesempio appare chiaro che occorrerà cercare una condizione più forte su C per giungere a garantirci che tutti gli elementi di $H_{\underline{X}}$ siano esprimibili come funzioni lineari di \underline{X} ; ora tuttavia sottolineiamo qual è la conseguenza dell'ipotesi $C > 0$.

Teorema 2.2.2: *se $C > 0$, l'operatore $U \equiv \{C\underline{\lambda} = \underline{\lambda}^+ \underline{X}\}$ tra H_C ed H_ℓ è biunivoco.*

□ Infatti $\underline{\lambda} \rightarrow X = \underline{\lambda}^+ \underline{X}$ è univoco per l'unicità del limite di $(\underline{\lambda}^N)^+ \underline{X}$; viceversa se

$$X = \underline{\lambda}^+ \underline{X} = U\underline{\lambda} = 0$$

è anche, ricordando che il prodotto scalare è continuo rispetto ai propri elementi,

$$0 = \langle \underline{X}, X \rangle = \langle \underline{X}, \underline{\lambda}^+ \underline{X} \rangle = C\underline{\lambda} \Rightarrow \underline{\lambda} = 0 .$$

Dunque U è invertibile e perciò biunivoco. □

Corollario 2.2.2: *se $C > 0$ e se H_C è uno spazio di Hilbert, ovvero se tale spazio è completo, allora*

$$H_\ell \equiv H_{\underline{X}} . \quad (2.2.27)$$

□ In effetti se U è un'isometria biunivoca ogni successione $Y_n = \underline{\lambda}_n^+ \underline{X}$ di Cauchy in H_ℓ è generata da una successione $\{\underline{\lambda}_n\}$ di Cauchy in H_C , che quindi ammette limite $\underline{\lambda}$ in H_C ; ma allora $Y = \underline{\lambda}^+ \underline{X}$ è in H_ℓ ed è il limite di Y_n in quanto

$$\|Y - Y_n\|^2 = |\underline{\lambda} - \underline{\lambda}_n|_C \rightarrow 0 .$$

Perciò H_ℓ è chiuso in $H_{\underline{X}}$ e quindi coincide con esso per la Osservazione 2.2.1.

□

L'importanza di questo corollario sta nel fatto che esso afferma che la proprietà (2.2.27) può essere verificata guardando solo la struttura della matrice di covarianza C e lo spazio H_C da essa derivato.

Cerchiamo ora di vedere cosa possiamo guadagnare introducendo la condizione più restrittiva che (2.2.26) non sia mai verificata.

A tale scopo poniamo

$$\begin{aligned} H_{(n)} &\equiv [\text{Span} \{X_k, k \leq n\}], H^{(n)} \equiv [\text{Span} \{X_k, k > n\}], \\ H_{(-,n)} &\equiv [\text{Span} \{X_k, k \neq n\}]. \end{aligned} \quad (2.2.28)$$

Inoltre ricordiamo la definizione di successione biortogonale associata (che chiameremo b.o.a.) alla successione $\{X_n\}$.

Definizione 2.2.2: $\{\xi_n\}$ è una successione b.o.a. a $\{X_n\}$ se

$$\langle \xi_j, X_k \rangle = \delta_{jk}. \quad (2.2.29)$$

Teorema 2.2.3: se il processo \underline{X} non può mai essere predetto quasi certamente, cioè se

$$X_n \notin H_{(-,n)}, \quad \forall n, \quad (2.2.30)$$

allora esiste la successione $\{\xi_n\}$ b.o.a. a $\{X_n\}$.

□ Notiamo che $H_{(-,n)}$ è sottospazio chiuso di $H_{\underline{X}}$, perciò se è vera la (2.2.30) esisterà il vettore \overline{X}_n , proiezione ortogonale di X_n su $H_{(-,n)}$ e per di più

$$X_n \neq \overline{X}_n \Rightarrow \|X_n - \overline{X}_n\| > 0, \Rightarrow \|X_n\| > \|\overline{X}_n\|.$$

Poniamo allora

$$\xi_n = \frac{X_n - \overline{X}_n}{\|X_n - \overline{X}_n\|} = \frac{X_n - \overline{X}_n}{\|X_n\|^2 - \|\overline{X}_n\|^2}; \quad (2.2.31)$$

risulta

$$\forall m \neq n \quad \langle \xi_n, X_m \rangle = 0$$

perché per definizione $\xi_n \perp H_{(-,n)}$, $X_m \in H_{(-,n)}$, mentre

$$\langle \xi_n, X_n \rangle = \frac{\|X_n\|^2 - \langle X_n, \overline{X}_n \rangle}{\|X_n\|^2 - \|\overline{X}_n\|^2} = 1,$$

così che $\{\xi_n\}$ è la successione cercata. □

Notiamo che in base all'Osservazione 2.2.3 se esiste $\{\xi_n\}$ b.o.a. ad $\{X_n\}$, l'operatore C non può che essere positivo, così che, come si è già detto, (2.2.30) è una condizione più stringente di $C > 0$.

Purtroppo l'esistenza di $\{\xi_n\}$ non è ancora sufficiente per garantire la (2.2.27) come il seguente controesempio, simile all'Esempio 2.2.1, dimostra.

Esempio 2.2.2: sia $\{\xi_n, n \geq 0\}$ un rumore bianco con varianza unitaria; poniamo

$$X_n = \xi_0 + \xi_n, \quad n \geq 1;$$

è chiaro che $\{\xi_n, n \geq 1\}$ stessa è una successione b.o.a. a $\{X_n\}$. Tuttavia il vettore ξ_0 , che sta in $H_{\underline{X}}$ in quanto

$$\xi_0 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N X_n,$$

non può essere in H_ℓ . Infatti se

$$k \geq 1, \quad \xi_0 = \sum_{k=1}^{+\infty} \lambda_k X_k = \left(\sum_{k=1}^{+\infty} \lambda_k \right) \xi_0 + \sum_{k=1}^{+\infty} \lambda_k \xi_k,$$

allora è anche, preso il prodotto scalare con ξ_j ,

$$\lambda_j = \langle \xi_0, \xi_j \rangle = 0 \Rightarrow \xi_0 = 0$$

il che è assurdo per la condizione $\|\xi_0\| = 1$.

Osservazione 2.2.4: se $X \in H_\ell$, $X = \underline{\lambda}^+ \underline{X}$, allora $\{\underline{\xi}_n\}$ b.o.a. a $\{X_n\}$ permette di costruire direttamente $\underline{\lambda}$ da X , cioè realizza l'operatore U^{-1} .

Infatti, se ad esempio $n \in \mathcal{Z}^+$,

$$\langle X, \xi_k \rangle = \lim_{N \rightarrow \infty} \langle \sum_{n=0}^N \lambda_n X_n, \xi_n \rangle = \lambda_k; \quad (2.2.32)$$

se $n \in \mathcal{Z}$ la dimostrazione è analoga.

Per questo motivo $\{\xi_n\}$ è anche chiamata successione dei coefficienti biortogonali associati a $\{X_n\}$.

Osservazione 2.2.5: quando risulti $H_\ell = H_{\underline{X}}$ si avrà che $\forall X \in H_{\underline{X}}, \exists \underline{\lambda} \in H_C; X = \underline{\lambda}^+ \underline{X}$.

In questo caso si dice che $\{X_n\} \equiv \underline{X}$ è una base di Schauder dello spazio di Hilbert $H_{\underline{X}}$. Tuttavia è noto che per una semplice base di Schauder la convergenza della serie $X = \sum_n \lambda_n X_n$ è condizionata all'ordine in cui vengono presi gli X_n . Se si vuole che ciò non avvenga occorre richiedere che $\{X_n\}$ sia una cosiddetta base di Riesz, ovvero che valga una delle seguenti due condizioni, che sono in un certo senso simili alla condizione di convergenza assoluta per una serie numerica:

$$\sum \lambda_n X_n \in H_{\underline{X}} \Rightarrow \sum |\lambda_n| X_n \in H_{\underline{X}}, \quad (2.2.33)$$

$$\sum \lambda_n X_n \in H_{\underline{X}} \Rightarrow \sum \varepsilon_n \lambda_n X_n \in H_{\underline{X}} \text{ q.c.} \quad (2.2.34)$$

dove ε_n è una successione di ± 1 presi a caso ed in modo indipendente.

È facile vedere che la (2.2.34) implica la condizione equivalente

$$\underline{\lambda}^+ C \underline{\lambda} < +\infty \Rightarrow \sum_n \lambda_n^2 C_{nn} < +\infty; \quad (2.2.35)$$

la condizione (2.2.35) è essenzialmente solo una condizione su C . Introduciamo una nozione di equivalenza tra due operatori C, D positivi in \mathbb{R}^∞ .

Definizione 2.2.3: dati $C > 0, D > 0$ diciamo che C e D sono equivalenti tra loro, cioè che indicheremo con $C \asymp D$, se

$$\alpha \underline{\lambda}^+ D \underline{\lambda} \leq \underline{\lambda}^+ C \underline{\lambda} \leq \beta \underline{\lambda}^+ D \underline{\lambda}, \quad \beta \geq \alpha > 0. \quad (2.2.36)$$

È facile vedere che (2.2.36) è una relazione di equivalenza; in particolare il caso $\beta = \alpha > 0$ corrisponde proprio a $C = D$.

Poiché la relazione tra operatori positivi A, B

$$A \leq B \Rightarrow \underline{\lambda}^+ A \underline{\lambda} \leq \underline{\lambda}^+ B \underline{\lambda} \quad \forall \underline{\lambda} \text{ per cui } \exists \underline{\lambda}^+ A \underline{\lambda},$$

è una relazione di ordine parziale, riportiamo la (2.2.36) anche nella forma equivalente

$$\alpha D \leq C \leq \beta D. \quad (2.2.37)$$

Lemma 2.2.1: *se $C \asymp D$ valgono le seguenti proprietà:*

- 1) $\underline{\lambda} \in H_C \Leftrightarrow \underline{\lambda} \in H_D$
- 2) se $\{\underline{\lambda}_n\}$ è fondamentale in H_C lo è anche in H_D e viceversa
- 3) H_C è uno spazio di Hilbert se lo è H_D e viceversa, inoltre $C \asymp D$ anche se (2.2.36) è verificata solo $\forall \underline{\lambda} \in R_0$.

□ Le 1) e 2) sono immediate.

Se H_C è di Hilbert, cioè completo, e $\{\underline{\lambda}_n\} \in H_C$ è fondamentale $\exists \underline{\lambda} \in H_C$ tale che $|\underline{\lambda} - \underline{\lambda}_n|_C \rightarrow 0$; ma allora $\underline{\lambda} \in H_D$ per la 1) e inoltre

$$|\underline{\lambda} - \underline{\lambda}_n|_D \leq \frac{1}{\alpha} |\underline{\lambda} - \underline{\lambda}_n|_C \rightarrow 0$$

cioè anche H_D è completo e perciò di Hilbert.

Infine se (2.2.36) vale $\forall \underline{\lambda} \in \mathbb{R}_0$ preso un qualunque $\underline{\lambda} \in H_D$ è $|\underline{\lambda} - \underline{\lambda}^N|_D \rightarrow 0$ per il Teorema 2.2.1; ma allora è anche $|\underline{\lambda} - \underline{\lambda}^N|_C \rightarrow 0$ per la (2.2.36) e quindi la (2.2.36) stessa vale $\forall \underline{\lambda} \in H_D$. Lo stesso vale $\forall \underline{\lambda} \in H_C$. □

Dunque essenzialmente il Lemma 2.2.1 afferma che H_C ed H_D sono insiemisticamente coincidenti e topologicamente equivalenti.

Lemma 2.2.2: *se D è un operatore matriciale diagonale*

$$D_{nm} = D_n \delta_{nm} , \quad D_n > 0 \quad \forall n ,$$

H_D è sempre uno spazio di Hilbert.

□ In effetti in questo caso, posto

$$D^{1/2} = \{ \sqrt{D_n} \delta_{nm} \} \tag{2.2.38}$$

si vede che

$$\underline{\lambda} \in H_D , \quad \underline{\lambda}^+ D \underline{\lambda} < +\infty \Leftrightarrow \underline{\gamma} = D^{1/2} \underline{\lambda} \in \ell^2 , \quad \underline{\gamma}^+ \underline{\gamma} = \underline{\lambda}^+ D \underline{\lambda} < +\infty .$$

Dunque $D^{1/2}$ è una isometria invertibile (in quanto $D_n > 0$) tra H_D ed ℓ^2 ; ma allora poiché ℓ^2 è uno spazio di Hilbert e perciò chiuso lo è anche H_D .

Con ciò abbiamo dimostrato tutte le proprietà necessarie per trarre le conclusioni che raccogliamo nel seguente teorema:

Teorema 2.2.4: *sia D l'operatore diagonale costruito con la diagonale principale di C ,*

$$D_{nm} = C_{nm}\delta_{nm} ; \quad (2.2.39)$$

se $C \asymp D$ valgono le seguenti proprietà:

- 1) H_C è completo e dunque uno spazio di Hilbert,
- 2) $H_\ell \equiv H_{\underline{X}}$ e quindi

$$\forall X \in H_{\underline{X}}, \exists \underline{\lambda} \in H_C, X = \underline{\lambda}^+ \underline{X},$$

così che $\{X_n\}$ è una base di (Schauder) di $H_{\underline{X}}$,

- 3) $\{X_n\}$ è una base di Riesz di $H_{\underline{X}}$,
- 4) $\exists \{\xi_n\}$ b.o.a. a $\{X_n\}$ e tale successione è completa in $H_{\underline{X}}$.
- 5) l'operatore C^{-1} è un operatore matriciale in \mathbb{R}^∞ .

□ I punti 1), 2) e 3) sono già stati provati. Quanto al punto 4), che è poi connesso con il concetto di regolarità del processo che vedremo poi, l'esistenza di $\{\xi_n\}$ deriva dall'osservazione che un qualunque elemento di $H_{(-,n)}$ può essere scritto come

$$X = \sum_{k \neq n} \lambda_k X_k \in H_{(-,n)}, \underline{\lambda} \in H_C ; \quad (2.2.40)$$

perché $\{X_k, k \neq n\}$ è a sua volta una base di Riesz di $H_{(-,n)}$.

Ma allora anche \overline{X}_n avrà tale rappresentazione e quindi

$$\begin{aligned} \|X_n - \overline{X}_n\|^2 &\geq \alpha [C_{nn} + \sum_{k \neq n} \lambda_k^2 C_{kk}] \geq \\ &\geq \alpha C_{nn} > 0, \end{aligned}$$

cioè

$$X_n \notin H_{(-,n)} \quad \forall n$$

come richiesto dal Teorema 2.2.3.

Inoltre

$$\begin{aligned} X \in H_{\underline{\lambda}} &\Rightarrow X = \underline{\lambda}^+ \underline{X} ; \langle X, \xi_k \rangle = 0 , \forall k \Rightarrow \lambda_k = 0 , \forall k \\ &\Rightarrow \underline{\lambda} = 0 \Rightarrow X = 0 , \end{aligned}$$

cioè $\{\xi_k\}$ è completo in $H_{\underline{\lambda}}$.

Passiamo ora al punto 5); dapprima supponiamo che C abbia la diagonale identica a 1, così che C diventa una matrice di correlazione

$$C_{n,n} = 1 \Rightarrow D = I , H_D = \ell^2 , C = R . \quad (2.2.41)$$

In questo caso la condizione $C \asymp D$ diventa $R \asymp I$ e quindi $H_R \equiv H_D \equiv \ell^2$ come insieme. In primo luogo R trasforma ℓ^2 in sé; infatti

$$\begin{aligned} \underline{\lambda} \in \ell^2 \leq H_R, \quad |R\underline{\lambda}|^2 &= \underline{\lambda}^+ R^2 \underline{\lambda} = \underline{\lambda}^+ R^{1/2} R R^{1/2} \underline{\lambda} = \\ &\leq \beta |R^{1/2} \underline{\lambda}|_{\ell^2}^2 = \beta \underline{\lambda}^+ R \underline{\lambda} = \beta |\underline{\lambda}|_R^2 \leq \beta^2 |\underline{\lambda}|_{\ell^2}^2 ; \end{aligned} \quad (2.2.42)$$

tra l'altro (2.2.42) dimostra che se $\underline{\lambda}_n \rightarrow \underline{\lambda}$ in H_R allora $R\underline{\lambda}_n \rightarrow R\underline{\lambda}$ in ℓ^2 , cioè R è un operatore continuo in ℓ^2 . Il range dell'operatore R è denso in ℓ^2 ; infatti, essendo R simmetrica

$$\underline{\gamma} \in \ell^2, \quad \underline{\gamma}^+ R \underline{\lambda} = 0 \quad \forall \underline{\lambda} \in H_R \equiv \ell^2 \Rightarrow R \underline{\gamma} = 0 \Rightarrow \underline{\gamma} = 0 .$$

Inoltre il range di R è anche chiuso in ℓ^2 . Infatti sia $\underline{\mu}_n = R\underline{\lambda}_n$, $\underline{\mu}_n \rightarrow \underline{\mu}$ in ℓ^2 ; allora

$$\begin{aligned} |\underline{\lambda}_n - \underline{\lambda}_m|_R^2 &= (\underline{\lambda}_n - \underline{\lambda}_m)^+ (\underline{\mu}_n - \underline{\mu}_m) \leq |\underline{\lambda}_n - \underline{\lambda}_m|_{\ell^2} \cdot |\underline{\mu}_n - \underline{\mu}_m|_{\ell^2} \leq \\ &\leq \frac{1}{\alpha} |\underline{\lambda}_n - \underline{\lambda}_m|_R |\underline{\mu}_n - \underline{\mu}_m|_{\ell^2} , \end{aligned}$$

cioè

$$|\underline{\lambda}_n - \underline{\lambda}_m|_R \leq \frac{1}{\alpha} |\underline{\mu}_n - \underline{\mu}_m|_{\ell^2} ,$$

così che $\underline{\lambda}_n$ è convergente ad un qualche limite $\underline{\lambda} \in H_R$. Ma allora

$$\underline{\mu} = \lim \underline{\mu}_n = \lim R\underline{\lambda}_n = R\underline{\lambda} ,$$

cioè $\underline{\mu}$ appartiene al range di R che risulta chiuso. Pertanto il range di R deve essere tutto ℓ^2 e quindi $\text{range } \{R\} \supset R_0$ così che se $\underline{\delta}_k = \{\delta_{k\ell}\}$ esistono in ℓ^2 i vettori

$$\underline{S}_k = R^{-1}\underline{\delta}_k = \{S_{k\ell}\} \quad \left(\sum_{\ell} S_{k\ell}^2 < +\infty \right);$$

ma allora se $\underline{\mu} = \sum \mu_k \underline{\delta}_k \in \ell^2$, $(\sum \mu_k^2 < +\infty)$ e $\underline{\mu} = R\underline{\lambda}$, è anche

$$\underline{\lambda} = \sum \mu_k \underline{S}_k; \quad \lambda_{\ell} = \sum_{\ell} S_{k\ell} \mu_k,$$

cioè R^{-1} ha una rappresentazione come matrice.

Tra l'altro da $R\underline{\delta}_k \equiv \{R_{k\ell}\} \in \ell^2$, $\forall k$ ed $\underline{S}_k = R^{-1}\underline{\delta}_k = \{S_{k\ell}\} \in \ell^2$, $\forall k$, discende che tanto la matrice R quanto $R^{-1} = S$ hanno righe (colonne) in ℓ^2 .

Infine per una C generale, soddisfacente $C \asymp D$, osserviamo che si può porre

$$C = D^{1/2} R D^{1/2}$$

con R matrice di correlazione; ma allora

$$C^{-1} = D^{-1/2} R^{-1} D^{-1/2}$$

così che l'operatore C^{-1} è associato alla matrice

$$(C^{-1})_{n,m} = C_{nn}^{-1/2} S_{nm} C_{mm}^{-1/2}. \quad (2.2.43)$$

□

Osservazione 2.2.6: supponiamo che

$$\forall n, \quad \underline{D} \leq C_{nn} = \|X_n\|^2 \leq \tilde{D}; \quad (2.2.44)$$

allora

$$C \asymp D \Leftrightarrow C \asymp I \quad (2.2.45)$$

e se la (2.2.45) è soddisfatto è anche $H_C \equiv \ell^2$; in questo caso C è detto positivo in senso stretto.

In effetti se vale (2.2.44) è

$$\underset{\sim}{D} \underset{\sim}{\lambda}^+ \underset{\sim}{\lambda} \leq \underset{\sim}{\lambda}^+ D \underset{\sim}{\lambda} \leq \underset{\sim}{D} \underset{\sim}{\lambda}^+ \underset{\sim}{\lambda} \Rightarrow D \asymp I ,$$

così che $C \asymp D \Rightarrow C \asymp I$ e viceversa poiché \asymp è una relazione di equivalenza.

Dunque col Teorema 2.2.4 giungiamo alla conclusione che la condizione $C \asymp D$, complementata magari dalla (2.2.44), ci permette di procedere agevolmente ad un calcolo lineare in $H_{\underline{X}}$ mediante la rappresentazione di un qualunque X come $X = \underline{\lambda}^+ \underline{X}$; resta naturalmente aperta la questione di come si possa verificare tale condizione esaminando solo la matrice C . Torneremo in modo più soddisfacente su queste questioni analizzando i cosiddetti processi stazionari; qui ci limitiamo a dare una condizione sufficiente assai semplice anche se ben lontana dall'essere necessaria.

Lemma 2.2.3: *si ponga*

$$\rho_k = \sup_n \frac{|C_{n+k,n}|}{\sqrt{C_{n+k,n+k} C_{nn}}} ; \quad (2.2.46)$$

se

$$0 < \sum_k \rho_k < \frac{1}{2} \quad (2.2.47)$$

allora $C \asymp D$ con

$$\alpha = 1 - 2 \sum_k \rho_k , \quad \beta = 1 + 2 \sum_k \rho_k . \quad (2.2.48)$$

□ In effetti notiamo che, se ad esempio $n \in \mathcal{Z}$,

$$\begin{aligned} \underline{\lambda}^+ C \underline{\lambda} &= \sum_n \lambda_n^2 C_{n,n} + 2 \sum_{k=1}^{+\infty} \sum_n C_{n+k,n} \lambda_{n+k} \lambda_n \leq \\ &\leq \sum_n C_{n,n} \lambda_n^2 + 2 \sum_{k=1}^{+\infty} \sum_n |C_{n+k,n}| |\lambda_{n+k}| |\lambda_n| \leq \end{aligned}$$

$$\begin{aligned}
&\leq \sum_n C_{n,n} \lambda_n^2 + 2 \sum_{k=1}^{+\infty} \rho_k \sum_n \sqrt{C_{n+k,n+k}} |\lambda_{n+k}| \sqrt{C_{n,n}} |\lambda_n| \leq \\
&\leq \sum_n C_{n,n} \lambda_n^2 + 2 \sum_{k=1}^{+\infty} \rho_k \sqrt{\sum_n C_{n+k,n+k} \lambda_{n+k}^2 \left(\sum_n C_{n,n} \lambda_n^2 \right)} = \\
&= \underline{\lambda}^+ D \underline{\lambda} \left(1 + 2 \sum_{k=1}^{+\infty} \rho_k \right) .
\end{aligned}$$

Analogamente

$$\begin{aligned}
\underline{\lambda}^+ C \underline{\lambda} &\geq \sum_n C_{n,n} \lambda_n^2 - 2 \sum_{k=1}^{+\infty} \rho_k \sqrt{C_{n+k,n+k}} |\lambda_{n,n}| \sqrt{C_{n,n}} \geq \\
&\geq \left(1 - 2 \sum_{k=1}^{+\infty} \rho_k \right) \underline{\lambda}^+ D \underline{\lambda} .
\end{aligned}$$

□

Come si vede questo Lemma è basato su una maggiorazione abbastanza grezza, ciononostante esso serve a provare che comunque se la correlazione tra X_{n+k} ed X_n va a zero abbastanza rapidamente con $k \rightarrow \infty$ e per ogni n , allora la nostra condizione può essere soddisfatta.

Concludiamo il paragrafo stabilendo alcune proprietà utili dello spazio $H_{\underline{X}}$ quando valga la condizione (2.2.45), $C \asymp I$.

Lemma 2.2.4: *se $C \asymp I$ allora*

- 1) $\underline{\xi} = C^{-1} \underline{X}$, $C_{\underline{\xi}\underline{\xi}} = C^{-1}$
- 2) \underline{X} è b.o.a. a $\underline{\xi}$
- 3) $\forall X \in H_{\underline{X}}$ vale

$$X = \underline{X}^+ \langle \underline{\xi}, X \rangle = \underline{\xi}^+ \langle \underline{X}, X \rangle . \quad (2.2.49)$$

□ In primo luogo notiamo che se $C_{lm}^{(-1)}$ sono gli elementi della rappresentazione matriciale di C^{-1} , chiamato \underline{c}_k^- il vettore corrispondente alla colonna di posto

k in C^{-1} , si ha $\underline{c}_k^- \in \ell^2 \equiv H_C$ così che posto

$$\xi_k = (\underline{c}_k^-)^+ \underline{X} \Rightarrow \xi_k \in H_{\underline{X}},$$

cioè $\underline{\xi}$, definito come

$$\underline{\xi} = C^{-1} \underline{X} \in H_{\underline{X}},$$

è un vettore in $H_{\underline{X}}$. Inoltre

$$\langle \underline{\xi}, \underline{X}^+ \rangle = C^{-1} \langle \underline{X}, \underline{X}^+ \rangle = C^{-1} C = I, \quad (2.2.50)$$

cioè $\underline{\xi}$ è b.o.a. a \underline{X} . Ancora

$$C_{\xi_k \xi_k} = (\underline{c}_k^-)^+ C \underline{c}_k^- = \underline{\delta}_k^+ \underline{c}_k^- = (C^{-1})_{\ell k}$$

ed il punto 1) è provato.

Quanto al punto 2) basta trasporre (2.2.50)

$$\langle \underline{X}, \underline{\xi}^+ \rangle = I \quad (2.2.51)$$

per vedere che \underline{X} è b.o.a. a $\underline{\xi}$.

Infine notiamo che se $C \asymp I$, anche $C_{\underline{\xi}\underline{\xi}} = C^{-1} \asymp I$ e quindi

$$\forall X \in H_{\underline{X}}, \exists \underline{\gamma}, X = \underline{\xi}^+ \underline{\gamma}, \underline{\gamma} = \langle \underline{X}, X \rangle;$$

anche 3) è allora provato. \square

Lemma 2.2.5: *sotto le condizioni del Teorema 2.2.4 è*

$$U^* = U^{-1}. \quad (2.2.52)$$

\square Ricordiamo che $U^{-1}; H_{\underline{X}} \rightarrow H_C$ è per definizione tale che $\forall X \equiv \underline{X}^+ \underline{\lambda}$

$$U^{-1} X = U^{-1}(\underline{X}^+ \underline{\lambda}) \equiv \underline{\lambda}.$$

Poniamo $\underline{\gamma} = U^* X = U^*(\underline{X}^+ \underline{\lambda})$ e proviamo che $\underline{\gamma} \equiv \underline{\lambda}$. Ricordiamo che per definizione di U^* , $\forall \underline{\mu} \in H_C$

$$(U^+ X, \underline{\mu})_C = \langle X, U \underline{\mu} \rangle = \langle X, \underline{X}^+ \underline{\mu} \rangle;$$

ma allora

$$(\underline{\gamma}, \underline{\mu})_C = (U^+ X, \underline{\mu})_C = \langle \underline{\lambda}^+ \underline{X}, \underline{X}^+ \underline{\mu} \rangle = (\underline{\lambda}, \underline{\mu})_C$$

e quindi $\underline{\gamma} \equiv \underline{\lambda}$. \square

2.3 La predizione lineare ottimale (caso finito)

Gli strumenti teorici che abbiamo sviluppato nel §2.2 permettono di formulare e risolvere in modo assai semplice il problema di predire linearmente una variabile $Y \in H_{\underline{X}}$ a partire dalla conoscenza di altre variabili $\{Y_\ell, \ell = 1, \dots, m\} \in H_{\underline{X}}$, che consideriamo come osservabili. La teoria qui presentata è essenzialmente la versione a temi discreti di quella sviluppata negli anni '40 e '50 da A. N. Kolmogorov e N. Wiener.

Supponiamo dunque che siano soddisfatte le condizioni del Teorema 2.2.4. Siano poi:

$$Y_\ell = \underline{a}_\ell^+ \underline{X} \quad \ell = 1, \dots, m; \underline{a}_\ell \in H_C \quad (2.3.1)$$

delle variabili osservabili che possiamo anche considerare come funzionali lineari del processo; notiamo che considerando $\underline{Y} = \{Y_\ell\}$ come una v.c. in \mathbb{R}^m , la (2.3.1) può essere scritta come

$$\begin{aligned} \underline{Y} &= A \underline{X} \\ A &= \{a_{\ell k}, \ell = 1, \dots, m, k \in \mathcal{Z}\} . \end{aligned} \quad (2.3.2)$$

Ricordiamo che tutte le variabili già considerate sono per ipotesi a media nulla, cioè $E\{\underline{Y}\} = 0$.

Inoltre la matrice di covarianza di \underline{Y} è data da

$$C_{\underline{Y} \underline{Y}} = E\{\underline{Y} \underline{Y}^+\} = \{E(Y_\ell Y_k)\} = \{\underline{a}_\ell^+ C \underline{a}_k\} = A C A^+ \quad (2.3.3)$$

Sia ora Y un'altra variabile di $H_{\underline{X}}$ che potrà perciò essere espressa come funzione lineare di \underline{X} ,

$$Y = \underline{a}^+ \underline{X} ; \quad (2.3.4)$$

il problema che ci poniamo è di trovare la migliore approssimazione di Y mediante una combinazione lineare del vettore delle osservabili \underline{Y} . Notiamo che anche per Y si ha $E\{Y\} = 0$,

$$\sigma_Y^2 = E\{Y^2\} = \underline{a}^+ C \underline{a} \quad (2.3.5)$$

e inoltre

$$C_{\underline{Y} Y} = E\{\underline{Y} Y\} = E\{A \underline{X} \underline{X}^+ \underline{a}\} = A C \underline{a} . \quad (2.3.6)$$

Osserviamo che tutte le operazioni che coinvolgono la matrice A sono corrette perché coinvolgono un numero finito di prodotti di vettori in H_C con l'operatore C .

Ora si può anche notare che il problema posto in realtà si riduce ad un problema finito-dimensionale di regressione lineare della variabile Y sulle $\{Y_\ell\}$; tale problema visto in $H_{\underline{X}}$ è inquadrato nel sottospazio a $m + 1$ dimensioni $Span\{Y, Y_\ell\}$ e consiste nel cercare il minimo della distanza

$$\begin{aligned} \|Y - \underline{\lambda}^+ \underline{Y}\|^2 &= \left\| Y - \sum_{\ell=1}^m \lambda_\ell Y_\ell \right\|^2 = & (2.3.7) \\ &= E\{(Y - \underline{\lambda}^+ \underline{Y})^2\} = \sigma_Y^2 - 2\underline{\lambda}^+ C_{\underline{Y} \underline{Y}} + \underline{\lambda}^+ C_{\underline{Y} Y} \underline{\lambda} = . \end{aligned}$$

Tale minimo è raggiunto dallo stimatore ⁶

$$\hat{Y} = \hat{\underline{\lambda}}^+ \underline{Y} = C_{Y \underline{Y}} C_{\underline{Y} \underline{Y}}^{-1} \underline{Y} \quad (2.3.8)$$

che geometricamente rappresenta la proiezione ortogonale di Y su $Span\{Y_\ell, \}$; chiameremo \hat{Y} lo stimatore di Wiener-Kolmogorov (W.K.) di Y .

L'errore quadratico medio associato a questo stimatore è dato da

$$\begin{aligned} \mathcal{E}^2 &= E\{(Y - \hat{Y})^2\} = \|Y - \hat{Y}\|^2 = & (2.3.9) \\ &= \|Y\|^2 - \|\hat{Y}\|^2 = \sigma_Y^2 - C_{Y \underline{Y}} C_{\underline{Y} \underline{Y}}^{-1} C_{\underline{Y} Y} . \end{aligned}$$

Come si vede, data l'interpretazione della distanza in $H_{\underline{X}}$, il nostro stimatore \hat{Y} minimizza l'errore quadratico medio di stima tra tutti gli stimatori lineari e dunque in questo senso è ottimale dal punto di vista statistico.

Osservazione 2.3.1: nella soluzione (2.3.8) (2.3.9) del problema di stima posto, l'aspetto infinito dimensionale compare solo nella costruzione di σ_Y^2 , $C_{\underline{Y} \underline{Y}}$ e $C_{\underline{Y} Y}$ cioè nelle regole di propagazione della varianza-covarianza da C alle nuove variabili Y, Y_ℓ ; dunque il Teorema 2.2.4 qui ha lo scopo di assicurare che qualunque siano $Y, Y_\ell \in H_{\underline{X}}$ esistono i corrispondenti vettori di coefficienti $\underline{a}, \underline{a}_\ell$ dai quali varianza e covarianza di Y, \underline{Y} possono essere calcolate quando però $\underline{a}, \underline{a}_\ell$ siano già noti e magari siano finito-dimensionali, così che

⁶Si osservi che è $C_{Y \underline{Y}} = C_{\underline{Y} Y}^+$.

le (2.3.2) (2.3.5) (2.3.6) siano sempre ammissibili. Il problema posto ammette sempre la rappresentazione (2.3.7). Quanto alla soluzione (2.3.8), (2.3.9) questa esiste ed è unica chiaramente solo se $C_{\underline{Y} \underline{Y}}$ è una matrice invertibile. Notiamo che da (2.3.3) risulta

$$\underline{\lambda}^+ C_{\underline{Y} \underline{Y}} \underline{\lambda} = \sum_{\ell, k} \lambda_\ell \lambda_k \underline{a}_\ell^+ C \underline{a}_k = \left(\sum_{\ell} \lambda_\ell \underline{a}_\ell \right)^+ C \left(\sum_k \lambda_k \underline{a}_k \right)$$

e tale espressione, per le ipotesi su C , può essere nulla solo se $\sum \lambda_\ell \underline{a}_\ell \equiv 0$, cioè se i funzionali che definiscono le osservabili Y_ℓ non sono tra loro linearmente indipendenti. Pertanto facendo l'ipotesi che le osservabili siano sempre linearmente indipendenti la matrice (finita e simmetrica) $C_{\underline{Y} \underline{Y}}$ risulta positiva e quindi invertibile, così che (2.3.8) e (2.3.9) hanno senso.

Osserviamo anche che se $C_{\underline{Y} \underline{Y}}$ è positivo, lo è anche $C_{\underline{Y} \underline{Y}}^{-1}$ così che dalla (2.3.9) risulta

$$\mathcal{E}^2 < \sigma_Y^2, \quad (2.3.10)$$

il che rende significativo usare lo stimatore di W.K., \hat{Y} .

Osservazione 2.3.2: poiché tra H_C ed $H_{\underline{X}}$ esiste un'isometria biunivoca U , con inverso U^* , il problema di minimizzare $\| Y - \underline{\lambda}^+ \underline{Y} \|$ ha un'immagine isometrica in H_C , cioè

$$\begin{aligned} U^* Y = \underline{a}, \quad U^*(\underline{\lambda}^+ \underline{Y}) &= U^*[(A^+ \underline{\lambda})^+ \underline{X}] = A^+ \underline{\lambda} \\ |\underline{a} - A^+ \underline{\lambda}|_C &= \min; \end{aligned}$$

è questo un problema di minimi quadrati in H_C con soluzione

$$\hat{\underline{\lambda}} = (ACA^+)^{-1} AC \underline{a} = C_{\underline{Y} \underline{Y}}^{-1} C_{\underline{Y} Y}$$

ovvero

$$\hat{Y} = UA^+ \hat{\underline{\lambda}} = (A^+ \underline{\lambda})^+ \underline{X} = C_{\underline{Y} \underline{Y}}^{-1} C_{\underline{Y} Y} Y$$

Osservazione 2.3.3: quanto visto per la predizione di una singola variabile Y può essere facilmente generalizzato alla predizione di un vettore \underline{V} , di variabili in $H_{\underline{X}}$. Dunque ora l'informazione di covarianza di \underline{V} e di cross-covarianza con le osservabili \underline{Y} sarà data dalle seguenti formule

$$\underline{V} = B \underline{X} \quad (V_k = b_k^+ \underline{X}) \Rightarrow C_{\underline{V} \underline{V}} = BCB^+, \quad C_{\underline{V} \underline{Y}} = BCA^+.$$

Per trovare lo stimatore ottimale di \underline{V} , definiamo

$$Y = \underline{b}^+ \underline{V} \quad (\sigma_Y^2 = \underline{b}^+ C_{\underline{V}} \underline{V} \underline{b}), \quad C_{\underline{Y} \ \underline{Y}} = ACB^+ \underline{b} = C_{\underline{Y} \ \underline{V} \underline{b}}$$

ed applichiamo le formule già viste per lo stimatore W.K., \hat{Y} ; come vedremo risulterà

$$\hat{Y} = \underline{b}^+ \hat{V}$$

con \hat{V} indipendente da \underline{b} e con ciò il \hat{V} così trovato sarà uniformemente ottimale.

Ora

$$\hat{Y} = C_{\underline{Y} \ \underline{Y}} C_{\underline{Y} \ \underline{Y}}^{-1} \underline{Y} = \underline{b}^+ C_{\underline{V} \ \underline{Y}} C_{\underline{Y} \ \underline{Y}}^{-1} \underline{Y}$$

da cui appare chiaro che

$$\hat{V} = C_{\underline{V} \ \underline{Y}} C_{\underline{Y} \ \underline{Y}}^{-1} \underline{Y}. \quad (2.3.11)$$

Notiamo anche che da

$$\begin{aligned} \mathcal{E}^2 &= \sigma_Y^2 - C_{\underline{Y} \ \underline{Y}} C_{\underline{Y} \ \underline{Y}}^{-1} C_{\underline{Y} \ \underline{Y}} = \\ &= \underline{b}^+ C_{\underline{V} \ \underline{V} \underline{b}} - \underline{b}^+ C_{\underline{V} \ \underline{Y}} C_{\underline{Y} \ \underline{Y}}^{-1} C_{\underline{Y} \ \underline{V} \underline{b}} \end{aligned}$$

appare pure chiaro che la matrice di covarianza degli errori di stima è data da

$$\begin{aligned} E &= E\{(\underline{V} - \hat{V})(\underline{V} - \hat{V})^+\} = \\ &= C_{\underline{V} \ \underline{V}} - C_{\underline{V} \ \underline{Y}} C_{\underline{Y} \ \underline{Y}}^{-1} C_{\underline{Y} \ \underline{V}}. \end{aligned} \quad (2.3.12)$$

Osserviamo anche che un qualsiasi altro stimatore lineare \tilde{V} di \underline{V} ha una matrice di covarianza degli errori \tilde{E} più grande della E data dalla (2.3.12).

Osservazione 2.3.4: supponiamo che \underline{X} sia un processo normale; allora ogni variabile $X \in H_{\underline{X}}$ è normale perché limite in media quadratica di variabili normali. In questo caso come è noto la variabile di regressione $E\{Y|\underline{Y}\}$ è una funzione lineare di \underline{Y} e quindi anche il nostro stimatore W.K., \hat{Y} è ottimale in tutto $\mathcal{L}^2(\Omega)$, cioè è la funzione di \underline{Y} in $\mathcal{L}^2(\Omega)$ che dà luogo alla minima varianza dell'errore di stima.

Esempio 2.3.1: sia $\{X_n; n \in \mathcal{Z}^+, X_0 = 0\}$ il processo di Wiener descritto nell'Esempio 2.1.2, così che

$$E\{X_n\} = 0 \quad E\{X_{n+k}X_n\} = n\sigma_\varepsilon^2 ;$$

supponiamo per semplicità che $\sigma_\varepsilon^2 = 1$.

Supponiamo che si sia osservato il processo al tempo $n = 3$ ottenendo il valore

$$X_3 = 6 ;$$

vogliamo trovare le predizioni ottimali \hat{X}_n ed i loro e.q.m di stima.

Notiamo che

$$C_{n3} = E\{X_n X_3\} = \begin{cases} n & n < 3 \\ 3 & n \geq 3 \end{cases}$$

così che

$$\hat{X}_n = C_{n3}C_{33}^{-1} \cdot X_3 = \begin{cases} 2n & n < 3 \\ 6 & n \geq 3 \end{cases}$$

Inoltre

$$\mathcal{E}_n^2 = C_{nn} - C_{n3}C_{33}^{-1}C_{3n} = \begin{cases} n - \frac{n^2}{3} & n < 3 \\ n - 3 & n \geq 3 \end{cases}$$

Si noti che giustamente $\mathcal{E}_3 = 0$.

In Fig. 2.3.1 sono riportate sia la curva delle predizioni $\{\hat{X}_n\}$, che dell'errore di predizione $\{\mathcal{E}_n\}$.

Fig. 2.3.1

$$\widehat{X}_n = \text{---} , \mathcal{E}_n = \text{---} .$$

Esempio 2.3.2: sia $\{X_n, n \in \mathcal{Z}\}$ un processo a media nulla e con covarianza

$$E\{X_{n+m}X_n\} = 4e^{-\frac{|k|}{2}} .$$

Si sia osservato il vettore

$$\underline{Y} = \begin{pmatrix} X_{-1} \\ X_1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

e si voglia stimare

$$Y = X_n$$

con il suo e.q.m. \mathcal{E}_n .

Fig. 2.3.2

$$\widehat{X}_n = \text{---} , \mathcal{E}_n = \text{---} .$$

Risulta

$$C_{\underline{Y}\underline{Y}} = 4 \begin{pmatrix} 1 & e^{-1} \\ e^{-1} & 1 \end{pmatrix} ; C_{\underline{Y}\underline{Y}}^{-1} = \frac{1}{4} \frac{e^2}{e^2 - 1} \begin{pmatrix} 1 & -e^{-1} \\ -e^{-1} & 1 \end{pmatrix} ;$$

$$C_{\underline{Y}\underline{Y}} = 4 \begin{vmatrix} e^{-\frac{|n+1|}{2}} \\ e^{-\frac{|n-1|}{2}} \end{vmatrix}$$

così che

$$\widehat{Y} = \widehat{X}_n = C_{Y \underline{Y}} C_{\underline{Y}\underline{Y}}^{-1} \begin{vmatrix} X_{-1} \\ X_1 \end{vmatrix} = \frac{e^2}{e^2 - 1} \left[e^{-\frac{|n+1|}{2}} - e^{-\frac{|n-1|}{2}} \right] ;$$

inoltre

$$\mathcal{E}_n^2 = 4 - 4 \frac{e^2}{e^2 - 1} \left(e^{-|n+1|} + e^{-|n-1|} - 2e^{-\frac{|n+1|}{2} - \frac{|n-1|}{2} - 1} \right) .$$

Stime \widehat{X}_n ed e.q.m. \mathcal{E}_n sono rappresentati in Fig. 2.3.2.

Esempio 2.3.3: sia $X_n, n \in \mathcal{Z}$ un processo a media nulla e con varianza

$$C_k = e^{-\frac{k^2}{2}} .$$

Si è osservato il funzionale del processo

$$Y = \frac{1}{3}(X_{-1} + X_0 + X_1) = 1$$

e si vuole stimare il vettore

$$\underline{V} = \begin{vmatrix} X_{-1} \\ X_0 \\ X_1 \end{vmatrix}$$

e la matrice di covarianza E del vettore degli errori di stima, $\underline{V} - \widehat{\underline{V}}$.

Si noti che

$$C_{\underline{V}\underline{V}} = 4 \begin{vmatrix} 1 & e^{-1/2} & e^{-2} \\ e^{-1/2} & 1 & e^{-1/2} \\ e^{-2} & e^{-1/2} & 1 \end{vmatrix}$$

$$C_{Y\underline{V}} = \frac{4}{9} [3 + 4e^{-1/2} + 2e^{-2}]$$

$$C_{\underline{V}Y} = \frac{4}{3} \begin{vmatrix} 1 + e^{-1/2} + e^{-2} \\ 1 + 2e^{-1/2} \\ 1 + e^{-1/2} + e^{-2} \end{vmatrix}$$

così che

$$\hat{\underline{V}} = C_{\underline{V} \underline{Y}} = C_{\underline{Y} \underline{Y}}^{-1} = \begin{vmatrix} 0.9173 \\ 1.1654 \\ 0.9173 \end{vmatrix}$$

$$E = C_{\underline{V} \underline{V}} - C_{\underline{V} \underline{Y}} C_{\underline{Y} \underline{Y}}^{-1} = C_{\underline{Y} \underline{Y}} \begin{vmatrix} 1.8696 & -0.2806 & -1.5894 \\ -0.2806 & 0.5611 & -0.2806 \\ -1.5894 & -0.2806 & 1.8696 \end{vmatrix} ;$$

l'ovvia simmetria della soluzione è dovuta al fatto che il funzionale d'osservazione è simmetrico rispetto all'origine $n = 0$.

Si noti anche che mentre la matrice

$$C_{\hat{\underline{V}} \hat{\underline{V}}} = C_{\underline{V} \underline{Y}} C_{\underline{Y} \underline{Y}} C_{\underline{Y} \underline{V}}$$

è necessariamente singolare in quanto si sono stimati 3 funzionali (le componenti di \underline{V}) da una sola osservazione, al contrario la matrice E non risulta singolare.

Osservazione 2.3.5: osserviamo che la soluzione al problema di dare la migliore stima lineare di un vettore stocastico \underline{Y} , noto un altro vettore stocastico \underline{V} , fornita dalle (2.3.11) e (2.3.12), è in sostanza la soluzione di un problema finito-dimensionale per la quale, una volta assunto $E\{\underline{V}\} = 0, E\{\underline{Y}\} = 0$, occorre conoscere solo le covarianze e cross-covarianze

$$C_{\underline{Y} \underline{Y}}, C_{\underline{V} \underline{Y}} = C_{\underline{Y} \underline{V}}^+, C_{\underline{V} \underline{V}}. \quad (2.3.13)$$

La connessione tra la teoria dei processi stocastici e le applicazioni di questo paragrafo sta tutta nelle regole di propagazione.

$$\begin{aligned} \underline{Y} = A\underline{X}(Y_n = \underline{a}_n^+ \underline{X}, \underline{a}_n \in H_C) \\ \underline{V} = B\underline{X}(V_m = \underline{b}_m^+ \underline{X}, \underline{b}_m \in H_C) \end{aligned} \Rightarrow \begin{cases} C_{\underline{Y} \underline{Y}} = ACA^+ \\ C_{\underline{Y} \underline{V}} = ACB^+ \\ C_{\underline{V} \underline{V}} = BCB^+ \end{cases} \quad (2.3.14)$$

Estendiamo allora il caso esaminato ad un problema che pur essendo simile alla predizione di funzionali di \underline{X} da altri noti, ne differisce in quanto introduce un rumore \underline{v} nel modello d'osservazione.

Dunque la variabile \underline{Y} delle osservabili sarà legata al processo \underline{X} dal modello lineare

$$\underline{Y} = A\underline{X} + \underline{\nu} \quad (Y_\ell = a_\ell^+ \underline{X} + \nu_\ell) \quad (2.3.15)$$

con

$$E\{\underline{\nu}\} = 0, \quad E\{\underline{\nu} \underline{\nu}^+\} = C_{\underline{\nu} \underline{\nu}} \quad (2.3.16)$$

$$E\{\underline{X} \underline{\nu}^+\} = 0; \quad (2.3.17)$$

$\underline{\nu}$, che rappresenta qui l'errore di osservazione di \underline{Y} , è a media nulla e spesso (ma non sempre) è un rumore bianco così che $C_{\underline{\nu} \underline{\nu}} = \sigma_\nu^2 I$; inoltre per semplicità $\underline{\nu}$ è assunto linearmente indipendente dal processo \underline{X} , così che vale la (2.3.17).

Da (2.3.15), (2.3.16) e (2.3.17) si ricava

$$C_{\underline{Y} \underline{Y}} = ACA^+ + C_{\underline{\nu} \underline{\nu}}, \quad C_{\underline{Y} \underline{X}} = ACB^+, \quad (2.3.18)$$

inoltre è ovviamente $C_{\underline{X} \underline{X}} = BCB^+$ come già nella (2.3.14).

Una volta specificata la (2.3.18) si trova direttamente come stimatore W.K. di \underline{Y} ,

$$\hat{\underline{Y}} = C_{\underline{Y} \underline{Y}} \cdot C_{\underline{Y} \underline{Y}}^{-1} \underline{Y} = (BCA^+)(ACA^+ + C_{\underline{\nu} \underline{\nu}})^{-1} \underline{Y} \quad (2.3.19)$$

con matrice degli errori di stima

$$\begin{aligned} E &= C_{\underline{\nu} \underline{\nu}} - C_{\underline{Y} \underline{Y}} C_{\underline{Y} \underline{Y}}^{-1} = C_{\underline{Y} \underline{Y}} = \\ &= BCB^+ - BCA^+[ACA^+ + C_{\underline{\nu} \underline{\nu}}]^{-1} ACB^+. \end{aligned} \quad (2.3.20)$$

Queste formule quando siano applicate al problema particolare,

$$\begin{cases} Y_\ell = X_\ell + \nu_\ell & \ell = 1, 2, \dots, N \\ V_\ell = X_\ell \end{cases} \quad (2.3.21)$$

cioè alla ricerca delle stime ottimali \hat{X}_ℓ di X_ℓ a partire da osservazioni con noise, danno

$$\hat{X}_\ell = \sum_{k,j=1}^N C_{\ell k} \cdot (C + C_{\underline{\nu} \underline{\nu}})^{(-1)}_{kj} Y_j \quad \ell = 1, 2, \dots, N$$

ovvero

$$\hat{\underline{X}} = C^{(N)}[C^{(N)} + C_{\underline{\nu} \underline{\nu}}]^{-1} \underline{Y}, \quad (2.3.22)$$

dove

$$C^{(N)} = \{C_{\ell k}; \ell, k = 1, 2, \dots, N\}. \quad (2.3.23)$$

Inoltre risulta

$$E = C^{(N)} - C^{(N)}[C^{(N)} + C_{\underline{\nu} \underline{\nu}}]^{-1} C^{(N)}. \quad (2.3.24)$$

Il problema in questo caso è noto come problema di filtraggio del noise e la soluzione (2.3.22), (2.3.24) è nota come filtro ottimale di Wiener.

Osserviamo anche che quando $C_{\underline{\nu} \underline{\nu}} = 0$, cioè quando non ci sia noise di osservazione, risulta $\hat{\underline{X}} = \underline{Y}$ e $E = 0$, come è logico aspettarsi.

Osserviamo ancora che lo stesso schema, con opportune modifiche della matrice di covarianza, vale anche quando i “tempi” d’osservazione ℓ non siano regolarmente spazati come in (2.3.21).

Infine notiamo che la (2.3.24) può essere anche scritta come

$$E = C_{\underline{\nu} \underline{\nu}} - C_{\underline{\nu} \underline{\nu}}[C^{(N)} + C_{\underline{\nu} \underline{\nu}}]^{-1} C_{\underline{\nu} \underline{\nu}}, \quad (2.3.25)$$

formula che può risultare computazionalmente più semplice, in particolare quando $C_{\underline{\nu} \underline{\nu}} = \sigma_{\nu}^2 I$.

Esempio 2.3.4: sia $\{X_n, n \in \mathcal{Z}\}$ un processo a media nulla e con matrice di covarianza

$$C_{n+k, n} = \begin{cases} 4 - |k| & |k| \leq 3 \\ 0 & |k| > 3 \end{cases}.$$

Supponiamo che si siano osservati i valori di X_{-1}, X_1 con noise bianco, $\sigma_{\nu} = 1$ e indipendente dagli X ; il modello d’osservazione ed i valori osservati risultano

$$\begin{aligned} Y_1 &= X_{-1} + \nu_{-1} & Y_{01} &= 2 \\ Y_2 &= X_1 + \nu_1 & Y_{02} &= 2 \end{aligned}.$$

Si vuole la predizione W.K. di X_n, \hat{X}_n ed il suo errore quadratico medio \mathcal{E}_n . Per l'evidente simmetria del problema ci limiteremo a stimare per $n \geq 0$. Si noti che in questo caso $V = X_n$,

$$C_{\underline{Y} \underline{Y}} = \begin{vmatrix} 4 & 2 \\ 2 & 4 \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \quad C_{\underline{Y} \underline{Y}}^{-1} = \frac{1}{21} \begin{vmatrix} 5 & -2 \\ -2 & 5 \end{vmatrix},$$

$$C_{V V} = 4,$$

$$n = \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad \geq 5$$

$$C_{\underline{Y} V} = \begin{vmatrix} 3 \\ 3 \end{vmatrix} \begin{vmatrix} 2 \\ 4 \end{vmatrix} \begin{vmatrix} 1 \\ 3 \end{vmatrix} \begin{vmatrix} 0 \\ 2 \end{vmatrix} \begin{vmatrix} 0 \\ 1 \end{vmatrix} \begin{vmatrix} 0 \\ 0 \end{vmatrix}$$

così che risulta

$\hat{X}_n =$	$\frac{36}{21}$	$\frac{36}{21}$	$\frac{24}{21}$	$\frac{12}{21}$	$\frac{6}{21}$	0
$\mathcal{E}_n =$	1.1952	0.8729	1.4800	1.6762	1.9396	2.000

Come si vede dove la correlazione con i funzionali osservati va a 0, cioè per $n \geq 5$, in mancanza di ulteriore informazione, la predizione va pure a 0 e l'errore di stima va a 2, cioè al valore quadratico medio del processo.

Fig. 2.3.3

$$\hat{X}_n = \text{---}\times\text{---} , \mathcal{E}_n = \text{---}\ast\text{---} , Y_0 = \circ .$$

2.4 Operatori lineari in uno spazio di Hilbert e loro rappresentazione matriciale. Propagazione della covarianza

Ricordiamo in primo luogo alcuni elementi di teoria degli operatori.

L'operatore lineare \mathcal{A} in $H_{\underline{X}}$ (cioè tale che $\mathcal{A}X = Y \in H_{\underline{X}}, \forall X \in H_{\underline{X}}$) è limitato se

$$\|\mathcal{A}X\| \leq a \|X\| ; \tag{2.4.1}$$

in questo caso naturalmente \mathcal{A} è continuo in tutto $H_{\underline{X}}$ ed inoltre si può porre

$$\|\mathcal{A}\| = \sup_X \frac{\|\mathcal{A}X\|}{\|X\|} = \inf\{a ; \|\mathcal{A}X\| \leq a \|X\|\} . \tag{2.4.2}$$

\mathcal{A} ammette inverso se

$$\mathcal{A}X = 0 \Rightarrow X = 0 ; \tag{2.4.3}$$

inoltre \mathcal{A}^{-1} è limitato se e solo se

$$\| \mathcal{A}X \| \geq b^{-1} \| X \| \quad (2.4.4)$$

con b opportuna costante maggiore di zero. In effetti se $Y = \mathcal{A}X$, $X = \mathcal{A}^{-1}Y$, la (2.4.4) implica che

$$\| X \| = \| \mathcal{A}^{-1}Y \| \leq b \| Y \| , \quad (2.4.5)$$

così che

$$\| \mathcal{A}^{-1} \| \leq b . \quad (2.4.6)$$

Sia ora $\underline{X} \equiv \{X_n\}$ una base da cui $H_{\underline{X}}$ è generato; ad ogni operatore \mathcal{A} lineare limitato possiamo associare la successione

$$\underline{Y} = \{Y_n\} \equiv \mathcal{A}\underline{X} \equiv \{\mathcal{A}X_n\} ; \quad (2.4.7)$$

poiché \underline{X} è una base in $H_{\underline{X}}$ si ha per un opportuno vettore \underline{a}_n ,

$$Y_n = \underline{a}_n^+ \underline{X} = (AX)_n , \quad (2.4.8)$$

cioè \underline{a}_n^+ è la riga n -esima di una matrice infinito-dimensionale A tale per cui

$$\underline{Y} = A\underline{X} \equiv \mathcal{A}\underline{X} \quad (2.4.9)$$

Osservazione 2.4.1: se \underline{X} è una base di Riesz e $C \asymp I$, \underline{a}_n^+ risulta essere un vettore in ℓ^2 così che le righe di A sono tutte in ℓ^2 .

La conoscenza di A è in un certo senso equivalente a quello di \mathcal{A} in quanto ci permette di ricostruire come \mathcal{A} agisce su un elemento qualunque di $H_{\underline{X}}$; infatti se

$$X = \underline{\lambda}^+ \underline{X} \in H_{\underline{X}} (\underline{\lambda} \in \ell^2)$$

allora

$$\begin{aligned} \mathcal{A}X &= \underline{\lambda}^+ \mathcal{A}\underline{X} = \underline{\lambda}^+ A\underline{X} = \\ &= (A^+ \underline{\lambda})^+ \underline{X} , \end{aligned}$$

ovvero

$$\mathcal{A}X = Y ; X = \underline{\lambda}^+ \underline{X}, Y = \underline{\mu}^+ \underline{X} \Rightarrow A^+ \underline{\lambda} = \underline{\mu} . \quad (2.4.10)$$

Lemma 2.4.1: se \mathcal{A} è limitato in $H_{\underline{X}}$, A^+ è limitato in H_C .

□ Essendo \mathcal{A} limitato per ipotesi, è pure

$$\|A^2\| = |\underline{\mu}|_C^2 = |A^+\underline{\lambda}|_C^2 \leq a^2 \|X\| = a^2 |\underline{\lambda}|_C^2, \quad (2.4.11)$$

così che A^+ è un operatore limitato in H_C . □

Osservazione 2.4.2: si noti che vale anche il viceversa del Lemma 2.4.1, ovvero se A^+ è un operatore limitato in H_C , allora

$$X = \underline{\lambda}^+ \underline{X}, \quad Y = (A^+ \underline{\lambda})^+ \underline{X}; \quad \mathcal{A}X = Y$$

definisce un operatore limitato in $H_{\underline{X}}$, in quanto

$$\|\mathcal{A}X\| = |A^+ \underline{\lambda}|_C \leq a |\underline{\lambda}|_C = a \|X\|$$

Vale dunque il seguente Teorema:

Teorema 2.4.1: la corrispondenza

$$\mathcal{A} \Leftrightarrow A^+$$

è una rappresentazione algebrica, isometrica rispetto alle norme uniformi di \mathcal{A} ed A^+ rispettivamente in $H_{\underline{X}}$ e in H_C .

□ In effetti se \mathcal{A}, \mathcal{B} sono limitati in $H_{\underline{X}}$ e A^+, B^+ sono le rispettive rappresentazioni matriciali, $X = \underline{\lambda}^+ \underline{X}$

$$\mathcal{B}\mathcal{A}X = \mathcal{B}(A^+ \underline{\lambda})^+ \underline{X} = (B^+ A^+ \underline{\lambda})^+ \underline{X}$$

cioè

$$\mathcal{B}\mathcal{A} \Leftrightarrow B^+ A^+ . \quad (2.4.12)$$

Inoltre, $X = \underline{\lambda}^+ \underline{X}$

$$\|\mathcal{A}\| = \sup_X \frac{\|\mathcal{A}X\|}{\|X\|} = \sup_{\underline{\lambda}} \frac{|A^+ \underline{\lambda}|_C}{|\underline{\lambda}|_C} = |A^+| \quad (2.4.13)$$

□

Osservazione 2.4.3: come si è visto dal Lemma 2.4.1 se \mathcal{A} è un operatore limitato in $H_{\underline{X}}$, A^+ è limitato in H_C ; in particolare posto

$$\underline{e}_k^+ A^+ \underline{\lambda} = (A \underline{e}_k)^+ \underline{\lambda} = \varphi_k(\underline{\lambda}), \quad (2.4.14)$$

è chiaro che φ_k è un funzionale lineare e limitato su H_C . Se ora supponiamo che $C \asymp I$, si ha che $H_C \equiv \ell_2$ e quindi φ_k ha come rappresentazione un prodotto scalare in ℓ_2 , per il Teorema di Riesz, $\varphi_k(\underline{\lambda}) = \underline{b}_k^+ \underline{\lambda}$ ($\underline{b}_k \in \ell^2$), che confrontata con la (2.4.14) mostra che $A \underline{e}_k = \text{col}_k A \in \ell^2$.

Questa osservazione completa l'Osservazione 2.4.1 sulla struttura di A . notiamo anche che vale il seguente Lemma:

Lemma 2.4.2: *se A^+ è limitato in H_C e se $C \asymp I$, così che $H_C \equiv \ell^2$, allora anche A è limitato in H_C .*

□ Poiché $H_C \equiv \ell^2$ la verifica della limitatezza di A può essere fatta direttamente in ℓ^2 . Basta osservare che

$$|A \underline{\mu}|_{\ell^2} = \sup_{|\underline{\lambda}|=1} (A \underline{\mu})^+ \underline{\lambda} = \sup_{|\underline{\lambda}|=1} \underline{\mu}^+ A^+ \underline{\lambda} \leq |A^+| |\underline{\mu}|_{\ell^2} \quad (2.4.15)$$

□

Avendo concluso questa parte generale sulla rappresentazione matriciale degli operatori limitati, possiamo ora passare ai due risultati principali del paragrafo. Intanto notiamo che ogni operatore limitato \mathcal{A} genera dal processo \underline{X} un altro processo

$$\underline{Y} = \mathcal{A} \underline{X} = A \underline{X}; \quad (2.4.16)$$

come è ovvio dalla (2.4.16) $\underline{Y} \in H_{\underline{X}}$. Di questo nuovo processo ci interessa ora capire come sia fatto l'operatore di covarianza e se la nuova successione $\{Y_n\}$ possa essere usata come una base, magari una base di Riesz, in $H_{\underline{X}}$.

Teorema 2.4.2: *(propagazione della covarianza). Sia \mathcal{A} un operatore lineare limitato in $H_{\underline{X}}$ e sia $A(A^+)$ la sua rappresentazione matriciale; l'operatore matriciale*

$$C_{\underline{Y} \underline{Y}} = A C A^+ \quad (2.4.17)$$

rappresenta la covarianza di \underline{Y} .

□ Basta verificare che

$$C_{Y_n Y_m} = E\{Y_n Y_m\} = E\{\underline{a}_n^+ \underline{X} \underline{X}^+ \underline{a}_m\} = \underline{a}_n^+ C \underline{a}_m ,$$

il che implica la (2.4.17). □

Teorema 2.4.3: *se \underline{X} è una base di Riesz in $H_{\underline{X}}$ e se $\underline{Y} = \mathcal{A}\underline{X}$ con \mathcal{A} lineare limitato ed \mathcal{A}^{-1} è pure limitato in $H_{\underline{X}}$ allora \underline{Y} è una base di Riesz in $H_{\underline{X}}$.*

□ Rammentiamo che $C \asymp D$ per ipotesi. Inoltre, essendo \mathcal{A} limitato

$$\| \underline{\lambda}^+ \underline{Y} \|^2 = \underline{\lambda}^+ C_{Y Y} \underline{\lambda} = \| \mathcal{A} \underline{\lambda}^+ \underline{X} \|^2 \leq k \| \underline{\lambda}^+ \underline{X} \|^2 = k \underline{\lambda}^+ C \underline{\lambda}$$

ed anche, essendo \mathcal{A}^{-1} limitato,

$$\| \underline{\lambda}^+ \underline{Y} \|^2 = \underline{\lambda}^+ C_{Y Y} \underline{\lambda} = \| \mathcal{A} \underline{\lambda}^+ \underline{X} \|^2 \geq h \| \underline{\lambda}^+ \underline{X} \|^2 = h \underline{\lambda}^+ C \underline{\lambda} .$$

Ma allora

$$C_{\underline{Y} \underline{Y}} \asymp C \asymp D$$

ed \underline{Y} è una base di Riesz per il Teorema 2.2.4 □

2.5 Operatori causali e processi con lo stesso ordine temporale

In questo paragrafo considereremo l'indice k di X_k come un tempo discreto. Sia \mathcal{A} un operatore lineare e limitato in $H_{\underline{X}}$ ed $\underline{Y} = \mathcal{A}\underline{X}$.

Definizione 2.5.1: *diciamo che \mathcal{A} genera una trasformazione causale di \underline{X} se vale*

$$\mathcal{A}H_n \subseteq H_n , \quad \forall n . \tag{2.5.1}$$

Notiamo che la (2.5.1) essenzialmente ci dice che conoscendo il processo \underline{X} fino al tempo n si conosce anche \underline{Y} fino allo stesso tempo, infatti $Y_k = \mathcal{A}X_k$, $k \leq n$, appartiene anche lui ad H_n e quindi è esprimibile come combinazione lineare degli X_i , $i \leq n$ che sono noti.

Lemma 2.5.1: *se A è il rappresentante matriciale di un operatore \mathcal{A} casuale, allora A è una matrice triangolare bassa, ovvero*

$$a_{nm} = 0 \quad \forall m > n . \quad (2.5.2)$$

□ Se

$$Y_n = \mathcal{A}X_n \in H_n , \quad (2.5.3)$$

allora nella rappresentazione matriciale

$$Y_n = \underline{a_n^+} \underline{X} = \sum_{m=-\infty}^{+\infty} a_{nm} X_m$$

deve anche essere

$$a_{nm} = 0 \quad \forall m > n$$

se no la (2.5.3) non può essere verificata. □

Esempio 2.5.1: sia \mathcal{B} l'operatore, detto backward, definito da

$$\mathcal{B}X_n = X_{n-1} ; \quad (2.5.4)$$

è chiaro che per \mathcal{B} vale

$$\mathcal{B}H_n = H_{n-1} \subset H_n \quad (2.5.5)$$

e quindi \mathcal{B} è causale.

Il rappresentante matriciale B di \mathcal{B} ha come elementi

$$b_{nk} = \delta_{n,k+1}$$

ovvero ha una diagonale di 1 sulla prima sottodiagonale principale

e perciò \mathcal{A}^{-1} è causale. □

In effetti ritornando ai due esempi precedenti appare chiaro che

$$\mathcal{B}H_n \equiv H_{n-1}$$

e quindi $\mathcal{B}^{-1} \equiv \mathcal{F}$ non è causale, mentre

$$\mathcal{A}^{-1} = \sum_{k=0}^{+\infty} a^k B^k : H_n \leftrightarrow H_n$$

ed \mathcal{A}^{-1} è causale.

Consideriamo ora un altro problema; partendo da un processo $\underline{X} \equiv \{X_n\}$ generiamo $\underline{Y} \equiv \{Y_n\}$ mediante una trasformazione lineare. La questione è: possiamo ritenere che il parametro n per \underline{Y} abbia lo stesso significato di tempo che per \underline{X} , ovvero l'ordine temporale di \underline{Y} è lo stesso di \underline{X} ?

Per rispondere a questa domanda possiamo pensare di usare gli operatori \mathcal{B} od \mathcal{F} che generano l'evoluzione temporale di \underline{X} . Notiamo che dalle definizioni

$$\mathcal{B}X_n = X_{n-1} , \quad \mathcal{F}X_n = X_{n+1} ,$$

appare chiaro che \mathcal{B}, \mathcal{F} sono definiti specificamente in relazione ad X così che si potrà anche scrivere $\mathcal{B} \equiv \mathcal{B}^X$, $\mathcal{F} \equiv \mathcal{F}^X$.

Il nuovo processo \underline{Y} avrà anch'esso i propri operatori di evoluzione temporale $\mathcal{B}^Y, \mathcal{F}^Y$; la questione è se essi coincidono con quelli di X , ovvero se

$$\mathcal{B} \equiv \mathcal{B}^X \equiv \mathcal{B}^Y \quad \text{e} \quad \mathcal{F} \equiv \mathcal{F}^X \equiv \mathcal{F}^Y \quad (2.5.13)$$

così che si possa affermare anche che

$$\mathcal{B}Y_n = Y_{n-1} , \quad \mathcal{F}Y_n = Y_{n+1} .$$

Definizione 2.5.2: diciamo che \underline{X} e \underline{Y} hanno lo stesso ordine (ovvero la stessa cadenza) temporale se i rispettivi operatori di traslazione temporale soddisfano la (2.5.13).

Di certo la (2.5.13) non è sempre vera come mostra il seguente controesempio.

Esempio 2.5.3: sia \underline{X} un processo con operatore di backward $\mathcal{B} \equiv \mathcal{B}^X$ e sia \mathcal{A} la trasformazione lineare generata dalla matrice

$$A = \begin{array}{c} \begin{array}{c} i \\ \vdots \\ i \\ 0 \\ 1 \end{array} \left| \begin{array}{cccc} & & & \\ & & & \\ \dots & 1 & & \\ & & 1 & \\ \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & 1 \end{array} \right. \begin{array}{c} 0 \\ \vdots \\ \vdots \\ 1 \\ 0 \end{array} \end{array} \begin{array}{c} 1 \\ \vdots \\ \vdots \\ 1 \\ 0 \end{array} \left| \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \dots \end{array} \right.$$

ovvero

$$\begin{aligned} Y_k &= X_k \quad k \neq 0, 1 \\ Y_0 &= X_1 \\ Y_1 &= X_0 ; \end{aligned}$$

Allora è chiaro da $\mathcal{B}^Y \neq \mathcal{B}$; infatti

$$\mathcal{B}^Y Y_0 = Y_{-1}$$

per definizione, mentre

$$\mathcal{B} Y_0 = \mathcal{B} X_1 = X_0 = Y_1 \neq Y_{-1} .$$

Alla domanda di quando \underline{X} e \underline{Y} abbiano lo stesso ordine temporale, ovvero sia verificata la (2.5.13), risponde il seguente teorema:

Teorema 2.5.1: sia \underline{X} una base di Riesz in $H_{\underline{X}}$ e sia

$$\underline{Y} = \mathcal{A}\underline{X} = A\underline{X} \tag{2.5.14}$$

\underline{Y} ha la stessa cadenza temporale di \underline{X} se e solo se A ha una struttura di Toeplitz, ovvero

$$A \equiv \{a_{n,k}\} \equiv \{a_{n-k}\} , \tag{2.5.15}$$

ovvero la (2.5.14) è un'operazione di convoluzione

$$Y_n = \sum_k a_{n-k} X_k \quad (2.5.16)$$

□ Se vale la (2.5.16), posto $\mathcal{B} \equiv \mathcal{B}^X$, si ha

$$\mathcal{B}Y_n = \sum_k a_{n-k} \mathcal{B}X_k = \sum_k a_{n-k} X_{k-1} = \sum_\ell a_{n-1-\ell} X_\ell \equiv Y_{n-1} ,$$

perciò è anche $\mathcal{B} \equiv \mathcal{B}^Y$.

Viceversa se

$$\begin{aligned} \mathcal{B}Y_n &= \mathcal{B} \sum_k a_{n,k} X_k = \sum_k a_{n,k} X_{k-1} = \\ &= \sum_\ell a_{n,\ell+1} X_\ell \equiv Y_{n-1} = \sum_\ell a_{n-1,\ell} X_\ell \end{aligned}$$

allora, essendo $\{X_n\}$ una base di Riesz, deve pure essere

$$a_{n,\ell+1} = a_{n-1,\ell} ,$$

cioè A è costante lungo le diagonali parallele alla diagonale principale, ovvero ha una struttura di Toeplitz. □

Corollario 2.5.1: *supponiamo che \mathcal{A} abbia un inverso limitato $\mathcal{D} \equiv \mathcal{A}^{-1}$; allora anche la matrice $D \equiv \mathcal{A}^{-1}$ ha necessariamente la struttura di Toeplitz.*

□ Infatti in tal caso è

$$\underline{X} = D\underline{Y}$$

ed è contemporaneamente $\mathcal{B}^Y \equiv \mathcal{B}^X$ così che D deve essere della forma $D \equiv \{d_{n-k}\}$ per il Teorema 2.5.1. □

Osservazione 2.5.3: se \underline{X} e \underline{Y} sono basi di Riesz allora $\{a_{n-k}\}, \{d_{k-j}\}$ devono essere vettori in ℓ^2 ovvero

$$\sum_k a_k^2 < +\infty, \sum_k d_k^2 < +\infty . \quad (2.5.17)$$

Quindi la condizione $A \cdot D \equiv I$, ovvero

$$\sum_k a_{n-k} d_{k-j} = \delta_{nj} \quad (2.5.18)$$

può essere studiata con un opportuno strumento come la trasformata di Fourier.

In effetti moltiplicando la (2.5.17) per $e^{-i2\pi jt}$ e sommando su j si trova ⁷

$$\begin{aligned} & \sum_k a_{n-k} \sum_j d_{k-j} e^{-i2\pi jt} \equiv \sum_k a_{n-k} e^{-i2\pi kt} . \\ & \cdot \sum_\ell d_\ell e^{i2\pi \ell t} \equiv e^{-i2\pi nt} \sum_m a_m e^{i2\pi mt} \cdot \sum_\ell d_\ell e^{i2\pi \ell t} \equiv \\ & \equiv \sum_j e^{-i2\pi jt} \delta_{nj} \equiv e^{-i2\pi nt} ; \end{aligned}$$

questa identità ci dice che le funzioni

$$a(t) = \sum_k a_k e^{i2\pi kt} , d(t) = \sum_k d_k e^{i2\pi kt} ,$$

che certamente sono in $L^2(0,1)$ in conseguenza della (2.5.17), devono essere tali che

$$a(t) \cdot d(t) \equiv 1 , t \in [0, 1] . \quad (2.5.19)$$

La (2.5.19) costituisce uno strumento sia teorico che pratico per lo studio di D . Infatti la (2.5.19) ci dice che \mathcal{A}^{-1} sarà un operatore limitato se

$$d(t) \equiv [a(t)]^{-1} \quad (2.5.20)$$

è una funzione in $L^2(0,1)$, ciò che si verifica ad esempio se

$$|a(t)| \geq \alpha > 0 ;$$

⁷Lo scambio della somma è reso lecito dalle condizioni $\{a_n\}, \{d_n\} \in \ell^2$.

inoltre quando tale condizione sia soddisfatta i coefficienti d_k che permettono di costruire D possono essere ricavati anche numericamente come coefficienti di Fourier della $a^{-1}(t)$ ovvero da

$$d_k = \int_0^1 e^{-i2\pi kt} [a(t)]^{-1} dt . \quad (2.5.21)$$

Osservazione 2.5.4: quando \mathcal{A} è contemporaneamente causale e tale da mantenere l'ordine temporale, devono valere contemporaneamente le relazioni

$$a_{n,k} = a_{n-k} \quad a_{n,k} = 0 , \quad k > n ,$$

così che si ha

$$Y_n = \sum_{k=-\infty}^n a_{n-k} X_k = X_n = \sum_{k=0}^{+\infty} a_k X_{n-k} . \quad (2.5.22)$$

2.6 Innovazione; processi regolari; decomposizione di Cholesky

Associato ad un processo $\underline{X} \equiv \{X_n\}$ ed al tempo n con il suo naturale senso di scorrimento definiamo un nuovo processo $\{\varepsilon_n\} \equiv \underline{\varepsilon}$ detto innovazione di \underline{X} .

Definizione 2.6.1: *supponiamo che*

$$X_n \notin H_{(n-1)} , \quad \forall n \quad (2.6.1)$$

e poniamo

$$\begin{aligned} \hat{X}_n &= P_{n-1} X_n && (\text{proiezione ortogonale su } H_{(n-1)}) \\ \mathcal{E}_n &= \| X_n - \hat{X}_n \| , \end{aligned}$$

osservando che risulta sempre $\mathcal{E}_n \neq 0$ per la (2.6.1); il processo $\{\varepsilon_n\}$ è allora definito da

$$\varepsilon_n = \frac{X_n - \hat{X}_n}{\mathcal{E}_n} . \quad (2.6.2)$$

Notiamo che in sostanza ε_n è il vettore che definisce il complemento ortogonale di $H_{(n-1)}$ in $H_{(n)}$; ne segue che $\{\varepsilon_n\}$ è una successione O.N. in quanto se $m \neq n$, ad esempio $m < n$, $\varepsilon_n \perp H_{(n-1)} \supset H_{(m)}$ mentre $\varepsilon_m \in H_{(m)}$.

Ci si può chiedere se $\{\varepsilon_n\}$ oltre ad essere O.N. in $H_{\underline{X}}$ sia anche completa; lo studio di questa questione ha portato alla definizione di processo regolare.

Definizione 2.6.2: *il processo $\{X_n\} \equiv \underline{X}$ è regolare se*

$$\bigcap_n H_n = \underset{\sim}{H} \equiv \emptyset . \quad (2.6.3)$$

A riguardo vale il seguente Lemma.

Lemma 2.6.1: *se \underline{X} non è regolare, allora $\underset{\sim}{H}$ è temporalmente invariante, ovvero*

$$\mathcal{B} \underset{\sim}{H} \subseteq \underset{\sim}{H} . \quad (2.6.4)$$

□ Se $X \in \underset{\sim}{H} = \bigcap_n H_n$ allora $\mathcal{B}X \in \bigcap_n \mathcal{B}H_n \equiv \bigcap_n H_{n-1} \equiv \underset{\sim}{H}$. □

Si noti anche che $\underset{\sim}{H}$, come intersezione di sottospazi chiusi è a sua volta un sottospazio chiuso di $H_{\underline{X}}$.

Possiamo ora provare il teorema fondamentale sul problema della completezza.

Teorema 2.6.1: *per $H_{\underline{X}}$ vale la decomposizione in spazi complementari ortogonali*

$$\begin{cases} H_{\underline{X}} = \underset{\sim}{H} \oplus [\text{Span}\{\varepsilon_n\}] \\ \underset{\sim}{H} \perp [\underset{\sim}{\text{Span}}\{\varepsilon_n\}] . \end{cases} \quad (2.6.5)$$

□ Poiché $H_{\underline{X}} = [U_N H_{(N)}]$ si può dimostrare il Teorema provando che $\forall N$

$$\begin{cases} H_{(N)} = \underset{\sim}{H} \oplus [\text{Span}\{\varepsilon_n, n \leq N\}] \\ \underset{\sim}{H} \perp [\text{Span}\{\varepsilon_n, n \leq N\}] . \end{cases}$$

A sua volta per vedere ciò basta provare che

$$\text{a) } h \in \underset{\sim}{H} \Rightarrow \langle h, \varepsilon_n \rangle = 0 \quad \forall n \leq N$$

$$\text{b) } h \in H_{(N)}, \langle h, \varepsilon_n \rangle = 0 \quad \forall n \leq N \Rightarrow h \in \underset{\sim}{H}.$$

$$\text{a) } h \in \underset{\sim}{H} \Rightarrow h \in H_{(m)}, \quad \forall m ; \text{ dunque fissato } m, h \in H_{(m-1)} \Rightarrow \\ \Rightarrow \langle h, \varepsilon_m \rangle = 0; \text{ perciò } h \perp [\text{Span}\{\varepsilon_n, n \leq N\}],$$

b) in primo luogo si noti che essendo $H_{(N)}$ una successione crescente vale $\bigcap_{n \leq N} H_{(N)} = \bigcap_n H_{(n)} = \underset{\sim}{H}$; ora se $h \in H_{(N)}$, $\langle h, \varepsilon_n \rangle = 0 \quad \forall n \leq N$, allora $\langle h, \varepsilon_N \rangle = 0 \Rightarrow h \in H_{(N-1)}$; ma è anche $\langle h, \varepsilon_{N-1} \rangle = 0 \Rightarrow h \in H_{(N-2)}$ e così via; perciò

$$h \in \bigcap_{n \leq N} H_{(n)} = \underset{\sim}{H} .$$

□

Corollario 2.6.1: *il processo $\underline{\varepsilon} \equiv \{\varepsilon_n\}$ costituisce una base ortonormale e completa in $H_{\underline{X}}$ se e solo se \underline{X} è regolare.*

Come sempre nei paragrafi precedenti, anche qui troviamo che se vale la condizione che $\{X_n\}$ sia una base di Riesz in $H_{\underline{X}}$, allora la situazione si semplifica.

Teorema 2.6.2: *se $\underline{X} \equiv \{X_n\}$ è una base di Riesz in $H_{\underline{X}}$, ovvero se $C \asymp D$ (oppure addirittura $C \asymp I$), allora $\underline{\varepsilon} \equiv \{\varepsilon_n\}$ è una successione ortonormale completa in $H_{\underline{X}}$ ed il processo \underline{X} è regolare.*

□ Ricordiamo che

$$H_{(-,n)} = H_{(n-1)} \oplus H^{(n+1)} ; H_{\underline{X}} = H_{(-,n)} \oplus [\lambda \xi_n]$$

ove $[\lambda \xi_n]$ è lo spazio monodimensionale generato da ξ_n che a sua volta è ortogonale ad $H_{(-,n)}$.

Ricordiamo anche che per il Teorema 2.2.4 se $C \asymp D$ $\{\xi_n\}$ è una base di Riesz, completa in $H_{\underline{X}}$. Allora

$$\{X \in H_{(-,n)}, \quad \forall n\} \equiv \{X \in \bigcap_n H_{(-,n)}\} \Rightarrow \langle X, \xi_n \rangle = 0, \quad \forall n, \Rightarrow X = 0 ,$$

ovvero è

$$\bigcap_n H_{(-,n)} = \emptyset .$$

Ma in questo caso risulta

$$\underset{\sim}{H} = \bigcap_n H_{(n-1)} \subset \bigcap_n H_{(-,n)} \equiv \emptyset$$

cioè \underline{X} è regolare ed $\{\underline{\varepsilon}\}$ è completa per il Corollario al Teorema 2.6.1. \square

Osservazione 2.6.1: se $C \asymp I$ anche il processo ottenuto da \underline{X} invertendo il tempo è regolare in quanto

$$\tilde{H} = \bigcap_n H^{(n)} \subset \bigcap_n H_{(-,n)} = \emptyset$$

e quindi anche la successione dell'innovazione "all'indietro" risulta ortonormale e completa.

Osservazione 2.6.2: per costruzione (cfr. (2.6.2)) appare ovvio che

$$Span\{\varepsilon_n, n \leq N\} \in H_{(N)}$$

così che, se $C \asymp D(C \asymp I)$, per il Teorema 2.6.2 è pure

$$[Span\{\varepsilon_n, n \in N\}] \equiv H_{(N)} . \quad (2.6.6)$$

Ma allora $\forall n$ si potrà porre

$$X_n = \sum_{k=-\infty}^n M_{nk} \varepsilon_k , \quad (2.6.7)$$

essendo $X_n \perp \{\varepsilon_{n+1}, \varepsilon_{n+2}, \dots\}$.

La (2.6.7) si scrive in forma matriciale

$$\underline{X} = M \underline{\varepsilon} \quad (2.6.8)$$

con M triangolare bassa e dunque rappresentante di un operatore causale \mathcal{M} , definito dalla relazione

$$\mathcal{M} \varepsilon_n = X_n \quad (2.6.9)$$

Teorema 2.6.3: *l'operatore causale \mathcal{M} con rappresentante M realizza la decomposizione di Cholesky della matrice di covarianza C del processo \underline{X} , ovvero*

$$C = MM^+ \quad (2.6.10)$$

con M triangolare bassa.

□ Basta osservare che $C_{\underline{\varepsilon}\underline{\varepsilon}} \equiv I$ ed applicare la propagazione della covarianza a (2.6.8). □

Esempio 2.6.1: sia $\{X_n, n \geq 1\}$ un processo con covarianza

$$C \equiv \begin{vmatrix} 1 & 1 & 1 & 1 & \dots \\ & 2 & 2 & 2 & \dots \\ & & 3 & 3 & \dots \\ & & & 4 & \dots \\ & & & & \ddots \end{vmatrix};$$

la decomposizione di C come in (2.6.10) è operata mediante la matrice M

$$C \equiv \begin{vmatrix} 1 & & & & \\ 1 & 1 & & & 0 \\ 1 & 1 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \end{vmatrix}$$

corrispondente a

$$X_n = \sum_{k=1}^n \varepsilon_k .$$

Si osservi che in questo caso $M^{-1} = A$ può essere trovata esplicitamente e risulta

$$A = M^{-1} = \begin{vmatrix} 1 & & & & \\ -1 & 1 & & & 0 \\ 0 & -1 & 1 & & \\ 0 & 0 & \ddots & \ddots & \\ & \ddots & \ddots & & \end{vmatrix};$$

con questa matrice si può invertire la (2.6.8) e trovare che

$$\underline{\varepsilon} = A\underline{X}, \quad \varepsilon_n = X_n - X_{n-1},$$

che costituisce la forma cosiddetta autoregressiva del processo \underline{X} .

Esempio 2.6.2: sia $\{X_n\}$ un processo con covarianza

$$C = \frac{1}{1-a^2} \begin{vmatrix} \ddots & \ddots & \ddots & & & & & & \\ & 1 & a & a^2 & \dots & & & & \\ & & 1 & a & a^2 & \dots & & & \\ & & & 1 & a & a^2 & \dots & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & & & & \end{vmatrix};$$

si noti che C ha struttura di Toeplitz.

È facile verificare direttamente che $C = MM^+$ con M data da

$$M = \begin{vmatrix} \ddots & 0 & & & & & & & \\ \ddots & 1 & 0 & & & & & & \\ \ddots & a & 1 & 0 & & & & & \\ \ddots & a^2 & a & 1 & 0 & & & & \\ & a^3 & a^2 & a & 1 & 0 & & & \\ & & \ddots & \ddots & \ddots & \ddots & & & \end{vmatrix} = \sum_{k=0}^{+\infty} a^k B^k$$

ovvero

$$X_n = \sum_{k=0}^{+\infty} a^k \varepsilon_{n-k} .$$

Anche in questo caso M può essere invertita dando luogo alla forma regressiva

$$\begin{aligned} \underline{\varepsilon} &= A\underline{X} \\ \varepsilon_n &= X_n - aX_{n-1} \end{aligned}$$

corrispondente a

$$A = M^{-1} = \begin{vmatrix} \ddots & & & & & & & & \\ \ddots & & & & & & & & 0 \\ & -a & 1 & & & & & & \\ & & -a & 1 & & & & & \\ 0 & & & -a & 1 & & & & \\ & & & & \ddots & \ddots & & & \end{vmatrix} = I - aB .$$

Esempio 2.6.3: sia $\{X_n\}$ un processo con covarianza

$$C = \begin{vmatrix} \ddots & \ddots & & & & & \\ \ddots & 1+a^2 & -a & & & & 0 \\ & -a & 1+a^2 & -a & & & \\ & & -a & 1+a^2 & -a & & \\ 0 & & & & \ddots & \ddots & \ddots \end{vmatrix};$$

è facile verificare che in questo caso

$$M = \begin{vmatrix} \ddots & & & & & & \\ \ddots & 1 & & & & & 0 \\ & -a & 1 & & & & \\ & & -a & 1 & & & \\ 0 & & & \ddots & \ddots & & \end{vmatrix} = I - aB$$

mentre risulta

$$A = M^{-1} = \begin{vmatrix} & & & & & & 0 \\ \ddots & & & & & & 1 & 0 \\ \ddots & a & & & & & 1 & 0 \\ \ddots & a^2 & a & & & & 1 & 0 \\ & a^3 & a^2 & a & & & 1 & 0 \\ & & \ddots & \ddots & \ddots & \ddots & & \end{vmatrix} = \sum_{k=0}^{+\infty} a^k B^k.$$

Si noti che essendo C una matrice con banda finita di ampiezza 1, altrettanto succede ad M , confermando così questa regola della decomposizione di Cholesky valida in dimensione finita.

2.7 Il problema della predizione lineare ottimale; caso infinito

Riprendiamo qui la trattazione del problema visto nel §2.3, generalizzato in questo caso all'ipotesi che si conoscano infiniti valori del processo \underline{X} , per tutti i tempi $m \leq n$, e che si voglia predire il processo stesso al tempo $n+k$.

Da un punto di vista astratto lo stimatore lineare è facilmente determinabile; in effetti se chiamiamo

$$(\widehat{X}_{n+k})_n = P_n X_{n+k} \quad (2.7.1)$$

con P_n proiettore ortogonale su H_n , il principio di Wiener-Kolmogorov di minimizzazione della distanza tra X_{n+k} , che dobbiamo predire, ed una funzione lineare delle osservabili, il predittore,

$$\widehat{X}_{n+k} = \sum_{m=-\infty}^n \lambda_m X_m, \quad (\widehat{X}_{n+k} \in H_n)$$

ha come soluzione precisamente

$$\widehat{X}_{n+k} = (\widehat{X}_{n+k})_n. \quad (2.7.2)$$

Se poi teniamo conto che è anche

$$H_n \equiv [\text{Span}\{X_m, m \leq n\}] \equiv [\text{Span}\{\varepsilon_m, m \leq n\}], \quad (2.7.3)$$

e che $\{\varepsilon_m\}$ è O.N.C., si vede che il risultato (2.7.2) può essere posto sotto la forma del Lemma seguente:

Lemma 2.7.1: *il miglior predittore lineare di X_{n+k} secondo il criterio di Wiener-Kolmogorov, basato sulla conoscenza di $\{X_m, m \leq n\}$ è dato da*

$$(\widehat{X}_{n+k})_n = \sum_{m=-\infty}^n M_{n+k,m} \varepsilon_m, \quad (2.7.4)$$

inoltre l'errore quadratico medio di stima $\mathcal{E}_{n+k,n}^2$ è dato da

$$\mathcal{E}_{n+k,n}^2 = \| X_{n+k} - (\widehat{X}_{n+k})_n \|^2 = \sum_{m=n+1}^{n+k} M_{n+k,m}^2. \quad (2.7.5)$$

□ In effetti è chiaro per quanto detto che

$$(\widehat{X}_{n+k})_n = \sum_{m=-\infty}^n \langle \varepsilon_m, X_{n+k} \rangle \varepsilon_m,$$

che congiuntamente alla relazione

$$X_{n+k} = \sum_{m=-\infty}^n M_{n+k,m} \varepsilon_m + \sum_{m=n+1}^{n+k} M_{n+k,m} m \varepsilon_m , \quad (2.7.6)$$

tenendo conto dell'ortonormalità degli ε_m prova tanto la (2.7.4) che la (2.7.5)
□

Naturalmente la (2.7.4) ci offre una soluzione sensata solo nella misura in cui si sia capaci di calcolare ε_m , $m \leq n$, dati X_m , $m \leq n$.

Allo scopo di risolvere il problema proviamo il seguente Lemma sull'operatore \mathcal{M} .

Lemma 2.7.2: *sia $C \asymp I$, allora sia \mathcal{M} che $\mathcal{M}^{-1} = \mathcal{A}$ sono operatori limitati in $H_{\underline{X}}$ e quindi ammettono rappresentanti matriciali $M, M^{-1} = A$; inoltre \mathcal{A} è causale, cioè A è triangolare bassa.*

□ Sia $X = \underline{\mu}^+ \underline{\varepsilon}$; risulta

$$\mathcal{M}X = \underline{\mu}^+ \mathcal{M} \underline{\varepsilon} = \underline{\mu}^+ \underline{X} = \underline{\mu}^+ M \underline{\varepsilon} = (M^+ \underline{\mu})^+ \underline{\varepsilon} ,$$

così che ricordando che $\underline{\varepsilon}$ è O.N.C.,

$$\| \mathcal{M}X \|_{H_{\underline{X}}}^2 = \underline{\mu}^+ (MM^+) \underline{\mu} = \underline{\mu}^+ C \underline{\mu} \asymp \underline{\mu}^+ \underline{\mu} = \| X \|_{H_{\underline{X}}}^2 ; \quad (2.7.7)$$

dove $a \asymp b$ significa che $(\frac{a}{b})$ è una quantità positiva limitata sia superiormente che inferiormente. La (2.7.7) dimostra che \mathcal{M} e \mathcal{M}^{-1} sono limitati. Inoltre, ricordando la (2.7.3), si vede che \mathcal{M} crea una corrispondenza biunivoca di H_n in sé stesso e quindi per il Lemma 2.5.2 \mathcal{M}^{-1} è causale, cioè M^{-1} è triangolare bassa. □

In conseguenza di questo lemma si può porre

$$\underline{\varepsilon} = M^{-1} \underline{X} = A \underline{X} \quad (2.7.8)$$

ovvero

$$\varepsilon_m = \sum_{j=-\infty}^m A_m ; X_j . \quad (2.7.9)$$

Usando questa soluzione nella (2.7.4) ed invertendo le sommatorie troviamo la forma esplicita dello stimatore ottimale

$$(\underline{X}_{n+k})_n = \sum_{j=-\infty}^n \left[\sum_{m=j}^n M_{n+k,m} A_{mj} \right] X_j . \quad (2.7.10)$$

Si noti che nella (2.7.10) risulta sempre $j \leq n$ così che i coefficienti in parentesi quadra sono ben diversi da $\delta_{n+k,j}$, che risulterebbe dal prodotto di una intera riga di M per una intera colonna di A .

In ogni caso con le (2.7.10), (2.7.5) il problema della predizione lineare è risolto a partire da X mediante il calcolo di M e di $A = M^{-1}$.

Riprendiamo a questo proposito gli esempi visti nel §2.6.

Esempio 2.7.1: sia \underline{X} come nell'Esempio 2.6.1 e si cerchi la predizione di X_3 dati X_1, X_2 . Dalla (2.7.10) risulta

$$(\widehat{X}_3) = (M_{31}A_{11} + M_{32}A_{21})X_1 + (M_{32}A_{22})X_2 \equiv X_2 ;$$

si osservi anche che

$$(\widehat{X}_m)_2 = X_2 \quad \text{per } m > 2 .$$

Lo stesso risultato poteva immediatamente essere ricavato dalla forma auto-regressiva

$$X_3 = X_2 + \varepsilon_3$$

da cui appare chiaro che

$$(\widehat{X}_3)_2 = X_2 , \quad \mathcal{E}_{3,2}^2 = 1 .$$

Esempio 2.7.2: supposto \underline{X} come nell'Esempio 2.6.2, cerchiamo $(\widehat{X}_{n+1})_n$.

Notiamo che

$$M_{n+1,m}A_{n,m} = a \cdot 1 = a$$

mentre

$$M_{n+1,n-1}A_{n-1,n-1} + M_{n+1,n}A_{n-1,n} = a^2 \cdot 1 + a(-a) = 0$$

e così via per tutti i $j < n - 1$.

Perciò

$$(\widehat{X}_{n+1})_n = aX_n$$

ed inoltre

$$\mathcal{E}_{n+1,n}^2 = 1$$

Di nuovo la stessa conclusione poteva essere derivata dall'equazione autoregressiva

$$X_{n+1} = aX_n + \varepsilon_n .$$

Esempio 2.7.3: supposto \underline{X} come nell'Esempio 2.6.3 cerchiamo $(\widehat{X}_{n+1})_n$, $(\widehat{X}_{n+2})_n$; risulta

$$\begin{aligned} (\widehat{X}_{n+1})_n &= -a \sum_{k=0}^{+\infty} a^k X_{n-k} \\ (\widehat{X}_{n+2})_n &= 0 . \end{aligned}$$

In questo caso il risultato si giustifica osservando che

$$X_{n+2} = \varepsilon_{n+2} - a\varepsilon_{n+1}$$

così che $X_{n+2} \perp H_n$, cioè $(\widehat{X}_{n+2})_n = 0$.

Inoltre dalla forma autoregressiva $\underline{\varepsilon} = A\underline{X}$ si vede che

$$X_{n+1} = \varepsilon_{n+1} - aX_n - a^2X_{n-1} \dots$$

ed essendo $\varepsilon_{n+1} \perp H_n$, la formula di predizione di $(\widehat{X}_{n+1})_n$ è corretta.

2.8 Un'applicazione al filtro di Kalman-Bucy

In estrema sintesi quanto sin qui fatto sul problema della predizione lineare ottimale può essere riassunto da due affermazioni:

- il predittore lineare ottimale di una variabile in $H_{\underline{X}}$, sulla base di osservazioni, è dato dalla proiezione ortogonale della variabile stessa sullo spazio lineare generato dalle osservabili

- tutte le informazioni utili sui prodotti scalari in $H_{\underline{X}}$ sono contenute nella matrice C ; quando poi sia $C \simeq I$ tutte le regole di calcolo lineare finito-dimensionale si riportano senza cambiamenti di forma anche in infinite dimensioni (propagazione della covarianza, Cholesky, ecc.)

Vediamo ora un'applicazione di questo principio di proiezione ortogonale ad un particolare schema, detto di Kalman-Bucy, che noi analizziamo solo in versione lineare ed a tempi discreti $n = 0, 1, 2, \dots$

L'importanza di questo "filtro" sta nella sua versatilità nella modellizzazione di sistemi lineari in evoluzione dinamica, nella estrema semplicità dei conti richiesti e nella possibilità, proprio per la semplicità dei calcoli, di implementare la predizione ottimale in modo ricorsivo in tempo reale.

Il modello che analizzeremo è costituito dalla equazione di evoluzione di un sistema descritto tramite un vettore (finito-dimensionale) detto vettore di stato \underline{X}_n , dipendente dal tempo n ; all'istante iniziale per semplicità supponiamo che il sistema parta da uno stato \underline{X}_0 noto. Parallelamente, a partire dal tempo $n = 1$, si compiono sul sistema delle osservazioni raccolte in un vettore \underline{Y}_n che sono esclusivamente funzioni lineari dello stato \underline{X}_n allo stesso tempo.

Il modello è dunque

$$\underline{X}_{n+1} = D_{n+1}\underline{X}_n + G_{n+1}\underline{\varepsilon}_{n+1} \quad (2.8.1)$$

$$\underline{Y}_{n+1} = A_{n+1}\underline{X}_{n+1} + M_{n+1}\underline{\nu}_{n+1} ; \quad (2.8.2)$$

$$\dim \underline{X} = p , \dim \underline{Y} = q$$

D_{n+1} matrice della dinamica del sistema

A_{n+1} matrice delle osservazioni

$\underline{\varepsilon}_n$ disturbo del sistema

$\underline{\nu}_n$ errore di misura;

su $\underline{\varepsilon}$ e $\underline{\nu}$ vengono fatte le seguenti ipotesi:

$$E\{\underline{\varepsilon}_n\} = 0 , E\{\underline{\varepsilon}_n \underline{\varepsilon}_m^+\} = \delta_{mn} E_n \quad (2.8.3)$$

$$E\{\underline{\nu}_n\} = 0 , E\{\underline{\nu}_n \underline{\nu}_m^+\} = \delta_{mn} N_n \quad (2.8.4)$$

$$E\{\underline{\varepsilon}_n \underline{\nu}_m^+\} = 0 . \quad (2.8.5)$$

Facciamo subito tre osservazioni.

Osservazione 2.8.1: la natura fisica di $\underline{\varepsilon}_n$, ovvero di disturbo di sistema che guida l'evoluzione di stato, è in genere diversa da quella di $\underline{\nu}_n$ che rappresenta solo un disturbo dell'osservazione del sistema.

Ad esempio se \underline{X} rappresenta una particella sospesa in un liquido descritta mediante 6 gradi di libertà (posizione, velocità) mentre \underline{Y} rappresenta la misura ottica della sua posizione, $\underline{\varepsilon}$ può essere la somma degli impulsi esercitati in 1 secondo sulla particella dalle molecole del liquido (moto Browniano) mentre $\underline{\nu}$ rappresenta l'errore della misura ottica.

Osservazione 2.8.2: quando la matrice d'osservazione A_n sia invertibile in tutti i tempi, lo schema (2.8.1), (2.8.2) può essere ridotto all'evoluzione della sola variabile \underline{Y}_n , tramite l'equazione

$$\begin{aligned}\underline{Y}_{n+1} = & A_{n+1}D_{n+1}A_n^{-1}\underline{Y}_n + A_{n+1}G_{n+1}\underline{\varepsilon}_{n+1} + \\ & + M_{n+1}\underline{\nu}_{n+1} - A_{n+1}D_{n+1}A_n^{-1}M_n\underline{\nu}_n .\end{aligned}$$

Questo però non è affatto un caso generale, come già l'esempio dell'Osservazione 2.8.1 mostra.

Osservazione 2.8.3: l'equazione dinamica (2.8.1) non contiene una successione di forze esterne \underline{F}_n , non stocastiche, come invece di solito si ha fisicamente. Tuttavia essendo la (2.8.1) lineare, non è necessario trattare la parte deterministica, osservando che se

$$\underline{X}_{n+1} = D_{n+1}\underline{X}_n + G_n\underline{\varepsilon}_n + \underline{F}_n ,$$

posto

$$\begin{aligned}\underline{X}_n &= \underline{X}_{on} + \delta\underline{X}_n \\ \underline{X}_{on+1} &= D_{n+1}\underline{X}_{on} + \underline{F}_n , \quad \underline{X}_{00} = 0\end{aligned}$$

la componente stocastica $\delta\underline{X}_n$ torna a soddisfare la stessa equazione (2.8.1) senza termine forzante.

Il problema che vogliamo risolvere è quello di dare una predizione lineare ottimale $(\hat{\underline{X}}_{n+1})_{n+1}$ di \underline{X}_{n+1} basata sulla conoscenza della sequenza di osservazioni $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_{n+1}$.

L'idea è quella di risolvere il problema iterativamente aggiornando la miglior predizione $(\hat{\underline{X}}_{n+1})_n$ (cioè basata su $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_n$) con l'introduzione di \underline{Y}_{n+1} ; anzi se $(\hat{\underline{Y}}_{n+1})_n$ è la proiezione di \underline{Y}_{n+1} sullo spazio generato da $\underline{Y}_1, \dots, \underline{Y}_n$, si potrà cercare $(\hat{\underline{X}}_{n+1})_{n+1}$ come

$$(\hat{\underline{X}}_{n+1})_{n+1} = (\hat{\underline{X}}_{n+1})_n + \underline{\xi}_{n+1} \quad (2.8.6)$$

dove la componente $\underline{\xi}_{n+1}$ è funzione (lineare) solo dell'innovazione $\underline{Y}_{n+1} - (\hat{\underline{Y}}_{n+1})_n$,

$$\underline{\xi}_{n+1} = \Lambda_{n+1} \{ \underline{Y}_{n+1} - (\hat{\underline{Y}}_{n+1})_n \}; \quad (2.8.7)$$

la nostra incognita sarà precisamente Λ_{n+1} . Se chiamiamo

$$H_n = \text{Span}\{ \underline{Y}_1, \dots, \underline{Y}_n \}, \quad (2.8.8)$$

possiamo osservare che $\underline{\varepsilon}_m, \underline{\nu}_m$ sono ortogonali ad H_n , $\forall m > n$. Pertanto dalla (2.8.1) si ha

$$(\hat{\underline{X}}_{n+1})_n = D_{n+1}(\hat{\underline{X}}_n)_n \quad (2.8.9)$$

La stima $(\hat{\underline{X}}_n)_n$ ha come errore di predizione

$$\underline{e}_{n,n} = \underline{X}_n - (\hat{\underline{X}}_n)_n \quad (2.8.10)$$

e noi supponiamo per ora di conoscere la covarianza

$$Q_{n,n} = E\{ \underline{e}_{n,n} \underline{e}_{n,n}^+ \}. \quad (2.8.11)$$

Sottraendo dalla (2.8.1) la (2.8.9) si trova l'errore di predizione di $(\hat{\underline{X}}_{n+1})_n$, ovvero

$$\underline{e}_{n+1,n} = \underline{X}_{n+1} - (\hat{\underline{X}}_{n+1})_n = D_{n+1} \underline{e}_{n,n} + G_{n+1} \underline{\varepsilon}_{n+1}. \quad (2.8.12)$$

Inoltre utilizzando la (2.8.2) si vede che

$$(\hat{\underline{Y}}_{n+1})_n = A_{n+1}, (\hat{\underline{X}}_{n+1})_n \quad (2.8.13)$$

che sottratta alla (2.8.2) stessa dà

$$Y_{n+1} - (\widehat{X}_{n+1})_n = A_{n+1}\underline{e}_{n+1,n} + M_{n+1}\underline{v}_{n+1} . \quad (2.8.14)$$

Usando in sequenza le (2.8.14), (2.8.7), (2.8.6), (2.8.10) si arriva all'equazione

$$\underline{e}_{n+1,n+1} = (I - \Lambda_{n+1}A_{n+1})D_{n+1}\underline{e}_{n+1,n} - \Lambda_{n+1}M_{n+1}\underline{v}_{n+1} . \quad (2.8.15)$$

Osservando che $\underline{e}_{n,n}, \underline{\varepsilon}_{n+1}$ e \underline{v}_{n+1} sono tra loro incorrelati, ovvero ortogonali, possiamo propagare la covarianza di $\underline{e}_{n,n}(Q_{n,n})$ a quella di $\underline{e}_{n+1,n+1}(Q_{n+1,n+1})$ tramite le (2.8.15), (2.8.12) trovando

$$\begin{aligned} Q_{n+1,n+1} = & (I - \Lambda_{n+1}A_{n+1})K_n(I - A_{n+1}^+\Lambda_{n+1}^+) + \\ & + \Lambda_{n+1}M_{n+1}N_{n+1}M_{n+1}^+\Lambda_{n+1}^+ \end{aligned} \quad (2.8.16)$$

dove

$$K_n = E\{\underline{e}_{n+1,n}\underline{e}_{n+1,n}^+\} = D_{n+1}Q_{n,n}D_{n+1}^+ + G_{n+1}E_nG_{n+1}^+ \quad (2.8.17)$$

Ora notiamo che

$$TrQ_{n+1,n+1} = E\{|\underline{e}_{n+1,n+1}|^2\}$$

è precisamente l'errore quadratico medio di predizione che deve essere minimizzato secondo il principio di Wiener-Kolmogorov. Dalla condizione

$$\min_{\Lambda_{n+1}} TrQ_{n+1,n+1}$$

ricaviamo l'equazione

$$\Lambda_{n+1} = K_n A_{n+1}^+ (A_{n+1} K_n A_{n+1}^+ + M_{n+1} N_{n+1} M_{n+1}^+)^{-1} . \quad (2.8.18)$$

Da Λ_{n+1} inizia l'iterazione secondo il seguente schema: suppongo di conoscere $(\widehat{X}_n)_n$ e $Q_{n,n}$ e poi

- dalla (2.8.9) calcolo $(\widehat{X}_{n+1})_n$
- dalla (2.8.13) calcolo $(\widehat{Y}_{n+1})_n$

- dalla (2.8.17) calcolo K_n
- dalla (2.8.18) calcolo Λ_{n+1}
- dalla (2.8.7) e dalla (2.8.13) calcolo $\underline{\xi}_{n+1}$
- dalla (2.8.6) calcolo $(\widehat{X}_{n+1})_{n+1}$.
- dalla (2.8.16), dalla (2.8.17) e dalla (2.8.18) calcolo $Q_{n+1,n+1}$

Così un ciclo è completato.

Lo schema parte dalla posizione

$$\begin{cases} (\widehat{X}_0)_0 = \underline{X}_0 \\ Q_{00} = 0 \end{cases} . \quad (2.8.19)$$

Illustriamo questo procedimento con un esempio estremamente semplice, benché fisicamente significativo.

Esempio 2.8.1: si ha un sistema a 2 gradi di libertà, per esempi una particella su un asse, che si muove sotto la spinta di impulsi casuali non correlati nel tempo, partendo dall'origine in quiete. Discretizzando nel tempo l'equazione del moto

$$\ddot{x} = \varepsilon$$

si ha l'equazione

$$x_{n+1} = 2x_n - x_{n-1} + \varepsilon_{n+1} , \quad (2.8.20)$$

che, come si vede connette tre tempi diversi

Per ricondurci alla forma (2.8.1), ovvero ad un'equazione alle differenze prime, poniamo

$$\underline{X}_n = \begin{vmatrix} x_n \\ x_{n+1} - x_n \end{vmatrix} , \quad (2.8.21)$$

ovvero il vettore di stato costituito da posizione e velocità della particella. Con questa posizione si può riscrivere l'equazione (2.8.20) nella forma

$$\underline{X}_{n+1} = \begin{vmatrix} x_{n+1} \\ x_{n+2} - x_{n+1} \end{vmatrix} = \quad (2.8.22)$$

$$= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} x_n \\ x_{n+1} - x_n \end{vmatrix} + \begin{vmatrix} 0 \\ 1 \end{vmatrix} \varepsilon_{n+1} \quad (2.8.23)$$

$$= D_{n+1} \underline{X}_n + G_{n+1} \varepsilon_{n+1} \quad (2.8.24)$$

Dalla (2.8.24) si vede che

$$D_{n+1} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \quad G_{n+1} = \begin{vmatrix} 0 \\ 1 \end{vmatrix}$$

ed inoltre, supposto che ε_{n+1} sia un rumore bianco,

$$E_{n+1} = \sigma_\varepsilon^2 .$$

Ora supponiamo che si osservi la velocità della particella, il che è fisicamente possibile mediante l'effetto Doppler; l'equazione d'osservazione discretizzata si può allora scrivere

$$\begin{aligned} Y_{n+1} &= x_{n+1} - x_n + \nu_{n+1} \\ &= A_{n+1} \underline{X}_n + \nu_{n+1} , \end{aligned} \quad (2.8.25)$$

dove sarà ovviamente

$$A_{n+1} = [1 \quad 1]$$

e

$$M_{n+1} = 1 , \quad N_{n+1} = \sigma_\nu^2 .$$

Lo stato iniziale del sistema (particella) in quiete sarà poi

$$\underline{X}_0 = 0 \quad Q_{0,0} = 0 . \quad (2.8.26)$$

Applicando le operazioni descritte nel paragrafo, troviamo il filtro di Kalman definito iterativamente nel seguente modo: posto

$$Q_{n,n} = \begin{vmatrix} q_n & r_n \\ r_n & p_n \end{vmatrix}$$

e supposto di conoscere $(\hat{X}_n)_n$, ciò che può essere inizializzato con (2.8.26) tenendo conto che $(\hat{X}_0) = \underline{X}_0 = 0$, si trovano $(\hat{X}_{n+1})_{n+1}$ e $Q_{n+1,n+1}$ con le

posizioni

$$K_n = \begin{vmatrix} k_n & \ell_n \\ \ell_n & h_n \end{vmatrix} = \begin{vmatrix} q_n + p_n + 2r_n & r_n + p_n \\ r_n + p_n & p_n + \sigma_\varepsilon^2 \end{vmatrix}$$

$$\Delta_n = q_n + 4p_n + 4r_n + \sigma_\varepsilon^2$$

$$\lambda_{n+1} = \begin{vmatrix} Q_n \\ b_n \end{vmatrix} = \frac{1}{\Delta} \begin{vmatrix} K_n + \ell_n \\ h_n + \ell_n \end{vmatrix}$$

$$\Lambda_{n+1} = \begin{vmatrix} a_n \\ b_n \end{vmatrix} \left(Y_{n+1} - [1 \ 2](\widehat{X}_n)_n \right)$$

$$(\widehat{X}_{n+1})_{n+1} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \left((\widehat{X}_n)_n + \underline{\xi}_{n+1} \right)$$

$$\begin{aligned} Q_{n+1,n+1} &= \begin{vmatrix} q_{n+1} & r_{n+1} \\ r_{n+1} & p_{n+1} \end{vmatrix} = \\ &= \begin{vmatrix} 1 - a_n & -a_n \\ -b_n & 1 - b_n \end{vmatrix} \begin{vmatrix} k_n & \ell_n \\ \ell_n & h_n \end{vmatrix} \begin{vmatrix} 1 - a_n & -b_n \\ -a_n & 1 - b_n \end{vmatrix} + \\ &+ \sigma_\nu^2 \begin{vmatrix} a_n^2 & a_n b_n \\ a_n b_n & b_n^2 \end{vmatrix}. \end{aligned}$$

3 I PROCESSI STAZIONARI

3.1 Processi stazionari in senso stretto

In termini qualitativi un processo a tempi discreti è stazionario se esso è indistinguibile, quanto a distribuzione in \mathbb{R}^∞ ,⁸ da un qualsiasi altro processo ottenuto dal primo per traslazione temporale.

Definizione 3.1.1: sia \underline{X} un processo in \mathbb{R}^∞ con distribuzione di probabilità $P_{\underline{X}}$ su $\mathcal{B}(\mathbb{R}^\infty)$; posto

$$\underline{Y} = \mathcal{B}\underline{X} , \quad (3.1.1)$$

\underline{X} è detto stazionario in senso stretto se risulta

$$P_{\underline{Y}} = P_{\underline{X}} , \quad (3.1.2)$$

ovvero se $\forall A \in \mathcal{B}(\mathbb{R}^\infty)$

$$P_{\underline{Y}}(A) = P_{\underline{X}}(\mathcal{B}A) = P_{\underline{X}}(A) \quad (3.1.3)$$

Osservazione 3.1.1: se \underline{X} è stazionario in senso stretto allora tutti i processi ottenuti da \underline{X} per traslazione temporale, ovvero

$$\underline{Y} = \mathcal{B}^k \underline{X} \quad \forall k \in \mathcal{Z} , \quad (3.1.4)$$

hanno la stessa distribuzione.

□ Infatti $\forall A \in \mathcal{B}(\mathbb{R}^\infty)$, per $k > 0$

$$\begin{aligned} P_{\underline{Y}}(A) &= P_{\underline{X}}[\mathcal{B}^k(A)] = P_{\underline{X}}[\mathcal{B}(\mathcal{B}^{k-1}A)] = \\ &= P_{\underline{X}}[\mathcal{B}^{k-1}A] = P_{\underline{X}}[\mathcal{B}^{k-2}A] \dots = P_{\underline{X}}[A] \end{aligned} \quad (3.1.5)$$

per la (3.1.3). Se $k < 0$ invece, basta riscrivere la (3.1.3) come

$$P_{\underline{X}}(A) = P_{\underline{X}}(\mathcal{B}^{-1}A)$$

e ripetere lo stesso ragionamento □

⁸In tutto questo paragrafo prendiamo come \mathbb{R}^∞ lo spazio lineare delle successioni illimitate nei due sensi

$$\underline{x} \equiv \{x_n; n = 0, \pm 1, \pm 2 \dots\} .$$

Osservazione 3.1.2: poiché una distribuzione di probabilità su $\mathcal{B}(\mathbb{R}^\infty)$ è definita per il Teorema di Kolmogorov dalle probabilità su \mathcal{C} , ovvero $\forall C \in \mathcal{B}(R^N) \equiv \{\underline{x} = \dots 0, x_1, \dots x_N, 0 \dots; (x_1, \dots x_N) \in C\}$

$$P_{\underline{X}}(C) = P_N(C) , \quad (3.1.6)$$

condizione necessaria e sufficiente affinché valga la (3.1.3) $\forall A \in \mathcal{B}(\mathbb{R}^\infty)$ è che la stessa relazione valga $\forall C \in \mathcal{B}(R^N), \forall N$.

Cioè \underline{X} e \underline{Y} devono avere le stesse distribuzioni marginali, ovvero

$$(X_0), (X_{-1}, X_0, X_1), (X_{-2}, X_{-1}, X_0, X_1, X_2)$$

ecc. devono avere le stesse distribuzioni di

$$(X_{-1}), (X_{-2}, X_{-1}, X_0)(X_{-3}, X_{-2}, X_{-1}, X_0, X_1)$$

Usando poi l'invarianza nella forma generale (3.1.4) si vede che le seguenti successioni di variabili

$$\begin{aligned} & \dots (X_{-1}), (X_0), (X_1) \dots \\ & \dots (X_{-2}, X_{-1}), (X_{-1}, X_0), (X_0, X_1) \dots \\ & \dots (X_{-3}, X_{-2}, X_{-1}), (X_{-2}, X_{-1}, X_0), (X_{-1}, X_0, X_1) \dots \end{aligned}$$

devono avere le stesse distribuzioni di probabilità rispettivamente in $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$ eccetera.

È facile vedere che ciò comporta anche che per ogni n -pla $i_1, \dots i_n$ le variabili

$$(X_{i_1}, \dots X_{i_n}) \sim (X_{i_1-k}, \dots X_{i_n-k}) \quad (3.1.7)$$

hanno identica distribuzione $\forall k \in \mathcal{Z}$.

Osservazione 3.1.3: quando in particolare le distribuzioni marginali di \underline{X} siano esprimibili per mezzo di densità di probabilità, si avrà per la (3.1.7),

$$f_{X_0}(x) = f_{X_1}(x) = f_{X_{-1}}(x) = \dots \quad (3.1.8)$$

ed ancora

$$f_{X_0 X_k}(x, y) = f_{X_n X_{n+k}}(x, y) , \quad \forall n \in \mathcal{Z} . \quad (3.1.9)$$

In particolare dalla (3.1.8) discendono le relazioni

$$E\{X_n\} = 0, \quad E\{X_n^2\} = C_0 = \sigma^2, \quad \forall n \in \mathcal{Z} \quad (3.1.10)$$

mentre dalla (3.1.9) si ha che

$$E\{X_{n+k}X_n\} = C_{n+k,n} = C_k \quad \forall n, k \in \mathcal{Z}. \quad (3.1.11)$$

Come si vede la seconda delle (3.1.10) e le (3.1.11) dicono che la matrice di covarianza C di \underline{X} è costante sulle diagonali parallele alla principale, ed è quindi dimostrato il seguente Lemma:

Lemma 3.1.1: *l'operatore matriciale di covarianza di un processo \underline{X} stazionario in senso forte ha la struttura di Toeplitz.*

Notiamo ancora che la proprietà di stazionarietà in senso forte, riguardo l'invarianza temporale di un processo, è una proprietà assai restrittiva; in effetti per descrivere un processo stazionario è necessario per così dire un numero molto più piccolo di distribuzioni marginali in quanto queste devono soddisfare le relazioni (3.1.7).

Questa proprietà della condizione di stazionarietà forte è ben caratterizzata da un teorema sull'operatore di traslazione temporale \mathcal{B} esteso in questo caso a tutto lo spazio $\mathcal{L}^2(\underline{X})$.

Definizione 3.1.2: *sia*

$$Y = g(\underline{X}) \in \mathcal{L}^2(\underline{X}); \quad (3.1.12)$$

definiamo

$$\mathcal{B}Y = \mathcal{B}g(\underline{X}) = g(\mathcal{B}\underline{X}). \quad (3.1.13)$$

Notiamo che la definizione è coerente con quella data per le funzioni lineari $g(\underline{X}) = \underline{\lambda}^+ \underline{X}$ in $H_{\underline{X}}$. Vale allora il seguente Teorema:

Teorema 3.1.1: *l'operatore \mathcal{B} relativo al processo stazionario in senso forte \underline{X} , è unitario in $\mathcal{L}^2(\underline{X})$.*

□ Infatti risulta

$$\begin{aligned}
\| \mathcal{B}g(\underline{X}) \|_{\mathcal{L}^2}^2 &= E\{g^2(\mathcal{B}\underline{X})\} = \int_{\mathbb{R}^\infty} g^2(\mathcal{B}\underline{X}) dP_{\underline{X}}(\underline{X}) = \\
&= \int_{\mathbb{R}^\infty} g^2(\underline{Y}) dP_{\underline{Y}}(\mathcal{B}^{-1}\underline{Y}) = \int_{\mathbb{R}^\infty} g^2(\underline{Y}) dP_{\underline{X}}(\underline{Y}) = \\
&= E\{g^2(\underline{X})\} = \| g(\underline{X}) \|_{\mathcal{L}^2}^2 .
\end{aligned} \tag{3.1.14}$$

Questa stessa proprietà potrebbe essere enunciata in modo anche un pò più generale come proprietà dell'operazione di media $E_{\underline{X}}$ sulla distribuzione di \underline{X} , dicendo che $E_{\underline{X}}$ è invariante per traslazione temporale, ovvero che $\forall g(\underline{X}) \in \mathcal{L}^1(\underline{X})$

$$E\{\mathcal{B}g(\underline{X})\} = E\{g(\mathcal{B}\underline{X})\} = E\{g(\underline{X})\} . \tag{3.1.15}$$

Esempio 3.1.1: un processo \underline{X} con componenti indipendenti ed identicamente distribuite, cioè un rumore bianco, è stazionario in quanto fissata la n -pla $(i_1 \dots i_n)$ ed il cilindro C di base $\{x_{i_1} \in I_1, \dots x_{i_n} \in I_n\}$ con I_k intervalli $(a_k, b_k]$ in \mathbb{R} , se inoltre chiamiamo $P_0(I)$ la distribuzione di X_0 in \mathbb{R} ,

$$\begin{aligned}
P\{\underline{X} \in C\} &= \prod_{k=1}^n P_0(I_k) = \\
&= P\{\mathcal{B}\underline{X} \in C\} = P\{X_{i_1-1} \in I_1 \dots X_{i_n-1} \in I_n\} .
\end{aligned}$$

Esempio 3.1.2: (processo normale). Ricordando l'Esempio 1.6 del Cap. 2, consideriamo un processo normale \underline{X} a media nulla e con covarianza C del tipo Toeplitz. Presa ora una n -pla $(i_1, \dots i_n)$ ed il corrispondente vettore $(X_{i_1}, \dots X_{i_n})$, ne chiamiamo C_n la matrice di covarianza (n -dimensionale), con elementi

$$(C_n)_{\ell,k} = C_{i_\ell - i_k} .$$

Notiamo anche che il vettore $(X_{i_1-1}, \dots X_{i_n-1})$ ha esattamente la stessa matrice di covarianza C_n e quindi la stessa distribuzione su \mathbb{R}^n .

Ora, dato un borelliano A in \mathbb{R}^n , per un qualsiasi cilindro C con base $(x_{i_1}, \dots x_{i_n}) \in A$, risulta

$$P(\underline{X} \in C) = \int_A \frac{1}{2\pi^{n/2}(\text{def} C_n)^{1/2}} e^{-1/2\mathbf{x}^+ C_n^{-1} \mathbf{x}} d_n x ;$$

evidentemente $P(\mathcal{B}\underline{X} \in C)$ ha lo stesso identico valore di $P(\underline{X} \in C)$ perché la forma della densità di probabilità è la stessa. Dunque un processo normale a media nulla e con matrice di covarianza di Toeplitz è stazionario in senso forte.

3.2 Processi stazionari in senso debole

Come si è già visto nel Lemma 3.1.1 un processo \underline{X} stazionario, come sempre a media nulla, ha una matrice di covarianza con la struttura di Toeplitz o, equivalentemente, i prodotti scalari in $H_{\underline{X}}$

$$\langle X_{n+k}, X_n \rangle = C_k \quad \forall n, k \in \mathcal{Z} \quad (3.2.1)$$

dipendono solo da k .

Poiché in questo testo ci limitiamo a sviluppare una teoria cosiddetta del secondo ordine lavorando sostanzialmente nello spazio $H_{\underline{X}}$, la cui struttura è intimamente legata alla forma di C , viene naturale chiedersi quali proprietà siano garantite dalla sola condizione (3.2.1).

Definizione 3.2.1: *un processo \underline{X} a media nulla è detto stazionario in senso debole (ed in seguito lo chiameremo semplicemente stazionario) se la sua matrice di covarianza C soddisfa la condizione di Toeplitz (3.2.1). In tal caso la funzione $C_k = C(k)$, definita qui per gli argomenti interi k , è detta funzione di covarianza.*

Dalla Definizione 3.2.1 appare chiaro che ogni processo stazionario in senso forte lo è anche in senso debole, il viceversa tuttavia non è affatto detto come mostrato dal seguente controesempio.

Esempio 3.2.1: sia $\{X_n\}$ un processo a componenti indipendenti con le componenti pari con densità di probabilità

$$X_0, X_{\pm 2}, \dots \sim f_0(x) ,$$

mentre le dispari hanno densità di probabilità

$$X_{\pm 1}, X_{\pm 3} \dots \sim f_1(x) .$$

Supponiamo che f_0, f_1 siano diverse tra loro ma tali che per entrambe

$$E_0\{X\} = E_1\{X\} = 0, \quad E_0\{X^2\} = E_1\{X^2\} = 1.$$

Allora il processo \underline{X} ha ovviamente media nulla, covarianza $C = I$ e nello stesso tempo non è stazionario in senso forte perché le componenti pari hanno distribuzioni monodimensionali diverse dalle componenti dispari.

Lemma 3.2.1: *l'operatore matriciale C ha la struttura di Toeplitz se e solo se*

$$BCB^+ = C \Leftrightarrow BC = CB \quad (3.2.2)$$

□ Il Lemma può essere dimostrato per applicazione diretta del prodotto BCB^+ ,

$$C_{i+1,k+1} = (BCB^+)_{i+1,k+1} = C_{i,k}. \quad (3.2.3)$$

□

La relazione tra questo Lemma ed i processi stazionari sta nel fatto che \underline{X} è stazionario se e solo se la sua covarianza C è di Toeplitz ma è pure stazionario se e solo se posto $\underline{Y} = B\underline{X}$, \underline{Y} ha la stessa covarianza di \underline{X} , cioè

$$C_{\underline{Y}\underline{Y}} = BCB^+ = C,$$

che appunto coincide con la (3.2.2).

Osservazione 3.2.1: il fatto che la matrice di Toeplitz C commuti con B , come nella seconda delle (3.2.2), è parte di un risultato molto più generale che concerne l'algebra delle matrici di Toeplitz che, ricordiamo, corrispondono ad operazioni di convoluzione.

Tale risultato deriva tra l'altro dal fatto ovvio che una qualunque matrice di Toeplitz T può essere messa nella forma

$$T = \sum_{k=-\infty}^{+\infty} t_k B^k; \quad (3.2.4)$$

infatti B^k è una matrice con coefficienti nulli dappertutto fuorché sulla k -esima diagonale dove valgono 1, pertanto t_k è il valore costante dei coefficienti di T che stanno su tale diagonale.

Viceversa la (3.2.4) rappresenta sempre una matrice di Toeplitz. Poiché la conoscenza di T è ovviamente equivalente alla conoscenza del vettore $\underline{t} = \{t_k\}$, cioè di una sua riga, tra le matrici T si può introdurre una nozione di norma, mutuata da una norma per \underline{t} ; a noi in particolare interessa la famiglia \mathcal{T}_1 di matrici T del tipo (3.2.4) per cui

$$\|\underline{t}\|_1 = \sum_{k=-\infty}^{+\infty} |t_k| < +\infty . \quad (3.2.5)$$

Vale allora il seguente risultato:

Lemma 3.2.2: *la famiglia \mathcal{T}_1 è un'algebra commutativa, ovvero $\forall T, V \in \mathcal{T}_1$*

$$TV = VT \in \mathcal{T}_1 . \quad (3.2.6)$$

□ Usando la (3.2.4) e l'analogia rappresentazione di V si ha

$$W = TV = \sum_{p=-\infty}^{+\infty} \mathcal{B}^p \sum_{k=-\infty}^{+\infty} t_{p-k} v_k . \quad (3.2.7)$$

Poiché, posto

$$w_p = \sum_{k=-\infty}^{+\infty} t_{p-k} v_k \quad (3.2.8)$$

risulta

$$\begin{aligned} \|\underline{w}\|_1 &= \sum_{p=-\infty}^{+\infty} |w_p| \leq \sum_{p=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} |t_{p-k}| |v_k| = \\ &= \|\underline{t}\|_1 \cdot \|\underline{v}\|_1 , \end{aligned}$$

si ha che la serie (3.2.8) è convergente e quindi la (3.2.7) rappresenta veramente una matrice in \mathcal{T}_1 .

Inoltre dall'identità

$$w_p = \sum_{k=-\infty}^{+\infty} t_{p-k} v_k = \sum_{h=-\infty}^{+\infty} t_h v_{p-h}$$

appare chiaro che

$$W = TV = VT .$$

□

Usando il Lemma 3.2.1 è facile ora dimostrare il seguente Teorema:

Teorema 3.2.1: *se \underline{X} è stazionario \mathcal{B} è un operatore unitario in $H_{\underline{X}}$, ovvero*

$$\forall X \in H_{\underline{X}} , \quad \| \mathcal{B}X \| = \| X \| . \quad (3.2.9)$$

□ In effetti posto $X = \underline{\lambda}^+ \underline{X}$, risulta

$$\| \mathcal{B}X \|^2 = \| \underline{\lambda}^+ \mathcal{B} \underline{X} \|^2 = \| \underline{\lambda}^+ \mathcal{B} C B^+ \underline{\lambda} \|^2 = \| \underline{\lambda}^+ C \underline{\lambda} \|^2 = \| X \|^2 .$$

□

Si ricordi qui che a causa dell'identità

$$\langle X, Y \rangle = \frac{1}{4} (\| X + Y \|^2 - \| X - Y \|^2) , \quad (3.2.10)$$

per l'operatore unitario \mathcal{B} vale anche l'identità

$$\langle \mathcal{B}X, \mathcal{B}Y \rangle = \langle X, Y \rangle \quad (3.2.11)$$

ovvero ancora

$$\mathcal{B}\mathcal{B}^+ = I , \quad \mathcal{B}^+ = \mathcal{B}^{-1} = \mathcal{F} . \quad (3.2.12)$$

Si osservi che le (3.2.12) valgono naturalmente sempre per le matrici B, B^+ ed F , ma solo per i processi stazionari valgono anche per i corrispondenti operatori in $H_{\underline{X}}$.

Terminiamo questa prima caratterizzazione dei processi stazionari in senso debole con un teorema tecnico di notevole importanza.

Teorema 3.2.2: *se \underline{X} è stazionario, il corrispondente processo di innovazione $\underline{\varepsilon}$ ha la stessa cadenza temporale di \underline{X} , ovvero*

$$\mathcal{B}^{\underline{\varepsilon}} \equiv \mathcal{B}^{\underline{X}} \equiv \mathcal{B} . \quad (3.2.13)$$

□ Ricordando la Definizione 6.1 del Capitolo 2, si ha

$$\varepsilon_n = \lambda_n(X_n - \widehat{X}_n) \quad (3.2.14)$$

dove

$$\begin{aligned} \lambda_n &= \|X_n - \widehat{X}_n\|^{-1} \\ \widehat{X}_n &= \text{proiezione ortogonale di } X_n \text{ su } H_n \equiv [\text{Span}\{X_k, k \leq n\}] . \end{aligned}$$

Dunque, ricordando che $\mathcal{B} : H_n \rightarrow H_{n-1}$ e $B : H_{n-1} \rightarrow H_{n-2}$

$$\mathcal{B}\varepsilon_n = \lambda_n(X_{n-1} - \mathcal{B}\widehat{X}_n) \in H_{n-1} . \quad (3.2.15)$$

Inoltre, usando la (3.2.11) si vede che per $k \geq 2$

$$\langle \mathcal{B}\varepsilon_n , \mathcal{B}X_{n-k+1} \rangle = \langle \varepsilon_n , X_{n-k+1} \rangle = 0 \quad (3.2.16)$$

perché ε_n è ortogonale ad H_{n-1} ; d'altro canto la (3.2.16) può essere riscritta come

$$\langle \mathcal{B}\varepsilon_n , X_{n-k} \rangle = 0 , \quad \forall k \geq 2 \quad (3.2.17)$$

il che dimostra che

$$\mathcal{B}\varepsilon_n \perp H_{n-2} . \quad (3.2.18)$$

Infine, per la (3.2.9)

$$\|\mathcal{B}\varepsilon_n\| = \|\varepsilon_n\| = 1 . \quad (3.2.19)$$

Le (3.2.15), (3.2.18) e (3.2.19) insieme dicono che

$$\mathcal{B}\varepsilon_n = \varepsilon_{n-1} ; \quad (3.2.20)$$

questa dimostra esattamente che $\mathcal{B} \equiv \mathcal{B}^{\varepsilon}$. □

L'importanza di questo teorema non è tanto nella (3.2.16) in sé quanto nelle sue conseguenze, riassunte nel seguente Corollario.

Corollario 3.2.1: se \underline{X} è stazionario e $C \asymp D$, gli operatori matriciali M ed A (cfr. §2.6, §2.7) tali che

$$\underline{X} = M\underline{\varepsilon}, \quad \underline{\varepsilon} = A\underline{X}, \quad (3.2.21)$$

hanno la struttura di Toeplitz, così che si può scrivere

$$X_n = \sum_{k=-\infty}^n M_{n,k} \varepsilon_k = \sum_{k=-\infty}^n m_{n-k} \varepsilon_k = \sum_{k=0}^{+\infty} m_k \varepsilon_{n-k} \quad (3.2.22)$$

$$\varepsilon_n = \sum_{k=-\infty}^n A_{n,k} X_k = \sum_{k=-\infty}^n a_{n-k} X_k = \sum_{k=0}^{+\infty} a_k X_{n-k}. \quad (3.2.23)$$

□ Il Corollario deriva dall'applicazione del Teorema 2.5.1 alle (3.2.21), tenuto conto del Teorema 3.2.2. □

Si noti che per la propagazione della covarianza (cfr. Teorema 2.6.3) deve essere

$$C = MM^+; \quad (3.2.24)$$

poiché se M è una matrice di Toeplitz altrettanto lo è M^+ , questa relazione è coerente col fatto che pure C deve avere tale struttura. Esplicitando la (3.2.24) in componenti, si trova la relazione che lega il vettore $\underline{c} \equiv \{\dots c_2 c_1 c_0 c_1 c_2 \dots\}$ associato alla matrice (simmetrica C) al vettore $\underline{m} \equiv \{\dots m_2 m_1 m_0 0 \dots\}$ associato alla matrice (triangolare bassa) M , ovvero

$$C_n = \sum_{k=-\infty}^{+\infty} m_k m_{k-n} = \sum_{h=|n|}^{+\infty} m_h m_{h-|n|} = \sum_{k=0}^{+\infty} m_{|n|+k} m_k. \quad (3.2.25)$$

Nel Cap. 2 gli Esempi 1.1, 1.5, 3.2, 3.3, 3.4, 6.1, 6.2, 6.3 riguardano processi stazionari; altri esempi si vedranno nei prossimi paragrafi.

3.3 Funzione di covarianza: caratteristiche

Come si è visto nel §3.2 un processo stazionario \underline{X} ha una matrice di covarianza C con struttura di Toeplitz cosicché C è completamente caratterizzata

da una sua riga (colonna) \underline{c} ; le componenti di \underline{c} in funzione del parametro “tempo” k costituiscono la cosiddetta funzione di covarianza, che di volta in volta scriveremo,

$$C_k \equiv c_k \equiv C(k) . \quad (3.3.1)$$

Vogliamo caratterizzare $C(k)$ ovvero definire condizioni necessarie e sufficienti affinché $C(k)$ sia una funzione di covarianza. Ovviamente tali condizioni devono derivare dal fatto che

$$C \equiv \{C_{kj}\} \equiv \{C(k-j)\} \quad (3.3.2)$$

deve essere un operatore di covarianza, in \mathbb{R}^∞ . Cominciamo col vedere alcune condizioni necessarie:

1) $C(k)$ deve essere pari

$$C(-k) = C(k) \quad (3.3.3)$$

2)

$$C_0 \geq |C(k)| \quad (3.3.4)$$

3) $C(k)$ deve essere definita positiva, ovvero

$$\sum_{k=-N}^N \sum_{j=-N}^N \lambda_{jk} \lambda_j C(k-j) > 0 \quad (3.3.5)$$

$$\forall N, \forall \{\lambda_k, -N \leq k \leq N\} \neq 0 .$$

La 1) deriva dal fatto che C deve essere simmetrica ed ha la struttura di Toeplitz; la 2), usando la Definizione 3.2.1, dice che per un coefficiente di correlazione vale

$$|\rho(k)| = \frac{|C(k)|}{C_0} \leq 1 ;$$

la 3) esprime la definita positività di C . notiamo che se in più si ha a che fare con un processo \underline{X} che è base di Riesz di $H_{\underline{X}}$ la 3) deve essere sostituita dalla condizione più forte

$$3') \quad \alpha \sum_{k=-N}^N \lambda_k^2 \leq \sum_{k,j=-N}^N \lambda_k \lambda_j C(k-j) \leq \beta \sum_{k=-N}^N \lambda_k^2$$

con α, β indipendenti da N (cfr. Teorema 2.2.4).

Notiamo anche che la 2) discende dalla 3) (o dalla 3'), cosicché non stupisce che valga il seguente Lemma:

Lemma 3.3.1: *una funzione $C(k)$ che soddisfi 1) e 3) è funzione di covarianza di un processo stazionario; se inoltre vale la 3') allora \underline{X} è regolare ed è una base di Riesz in $H_{\underline{X}}$.*

Questo Lemma non ha bisogno di particolari dimostrazioni in quanto è già stato illustrato nell'esempio 2.1.6 sulla costruzione dei processi normali, cui occorre solo aggiungere che la condizione 3') garantisce che $C \asymp I$, cosicché vale anche l'ultima conclusione, per il Teorema 2.2.4.

Notiamo anche che nell'Esempio 2.1.16 l'indice k variava sugli interi non negativi, tuttavia la generalizzazione a $-\infty < k < +\infty$ è immediata.

Naturalmente una verifica diretta delle condizioni 3) o 3') è assai difficile, tuttavia vedremo nel prossimo paragrafo un comodo criterio per tale verifica. Qui ci accontentiamo di fornire una tabella (Tabella 3.1) di famiglie di funzioni di covarianza oltre che alcuni Lemmi che permettono da queste di costruire altri modelli.

$$\frac{A_0}{\left(1 + \frac{k^2}{\alpha^2}\right)^\ell}$$

$$A_0 = 1, \alpha = 2, \ell = 1$$

Tab. 3.1.a

$$A_0 \left(1 - \frac{k}{N}\right)$$
$$A_0 = 1, N = 5$$

Tab. 3.1.b

$$A_0 \cos \omega k$$
$$A_0 = 1, \omega = \frac{\pi}{10}$$

Tab. 3.1.c

$$A_0 e^{-\alpha|k|}$$

$$A_0 = 1, \alpha = 0.3$$

Tab. 3.1.d

$$A_0 e^{-\alpha k^2}$$

$$A_0 = 1, \alpha = 0.1$$

Tab. 3.1.e

$$A_0 \frac{\sin \omega k}{\omega k}$$

$$A_0 = 1$$

Tab. 3.1.f

$$A_0 J_0(\alpha k)$$

$$A_0 = 1$$

Tab. 3.1.g

Naturalmente si potrebbero citare molti altri modelli di covarianza, tuttavia diamo qui delle regole utili per costruire nuove funzioni di covarianza, combinandone altre già note.

Prima di procedere notiamo che la funzione $A_0 \cos \omega k$ non è una funzione di covarianza di un processo regolare, in quanto essa torna periodicamente al valore massimo A_0 (quando ω è uguale a π per un razionale) mostrando che tra variabili del tipo X_n, X_{n+k} , con $\omega k = r\ell\pi$, esiste un coefficiente di correlazione $\rho = 1$, cioè una perfetta dipendenza lineare. Questo modello però è utile in combinazione con altri.

Lemma 3.3.2: *la funzione*

$$\begin{aligned} C(k) &= \lambda C_1(k) + \mu C_2(k) \\ \lambda, \mu &> 0 \end{aligned} \tag{3.3.6}$$

è una funzione di covarianza se lo sono $C_1(k)$ e $C_2(k)$.

Lemma 3.3.3: *la funzione*

$$C(k) = C_1(k) \cdot C_2(k) \tag{3.3.7}$$

è una funzione di covarianza se lo sono C_1, C_2 .

- Infatti basta costruire un processo \underline{X} con covarianza $C_1(k)$ ed un processo \underline{Y} indipendente con covarianza $C_2(k)$. Allora il processo

$$\underline{V} = V_n, \quad V_n = X_n Y_n$$

è ovviamente a media nulla, stazionario e con covarianza $C(k)$ data da (3.3.7).
 \square

Lemma 3.3.4: *la funzione*

$$C(k) = \sum_{\ell=-\infty}^{+\infty} C_1(k-\ell)C_2(\ell) \quad (3.3.8)$$

è una funzione di covarianza se lo sono C_1 e C_2 .

\square Ricordando il Lemma 3.2.2, dalla (3.3.8) si deduce che

$$C = C_1 C_2 \quad (3.3.9)$$

con C_1, C_2 due operatori matriciali di covarianza. D'altro canto si potrà porre

$$C_1 = M_1 M_1^+$$

con M_1 matrice di Toeplitz, così che sempre per il Lemma 3.2.2, M_1 e C_2 commutano; ma allora si può scrivere,

$$C = M_1 C_2 M_1^+ ,$$

che dimostra che C è simmetrica e definita positiva, cioè una matrice di covarianza. \square

Osservazione 3.3.1: nel riconoscere l'appartenenza di una funzione di covarianza ed una certa famiglia parametrica, si possono tenere in conto alcuni elementi qualitativi che caratterizzano la funzione stessa. Tra questi citiamo

- a) la presenza o meno di zeri e l'eventuale cadenza degli zeri
- b) il comportamento nell'origine (con tangente nulla o negativa)
- c) la rapidità di decadimento della funzione se non vi sono zeri, ovvero il decadimento di massimi e minimi quando vi siano zeri.

Osservazione 3.3.2: se un processo stazionario \underline{X} con covarianza C e funzione di covarianza $C(k)$ è sottoposto ad una trasformazione $T \in \mathcal{T}_1$,

$$\underline{Y} = T\underline{X}, \quad (Y_n = \sum_m t_{n-m} X_m) \quad (3.3.10)$$

per la propagazione della covarianza \underline{Y} ha come operatore $C_{\underline{Y}}$

$$C_{\underline{Y}} = TCT^+ . \quad (3.3.11)$$

Mettendo la (3.3.11) in componenti si trova la relazione tra la funzione di covarianza \underline{X} , $C(k)$ e quella di \underline{Y} , $C_{\underline{Y}}(k)$, ovvero

$$C_{\underline{Y}}(n) = \sum_{k,j=-\infty}^{+\infty} t_{n-k} C(k+j) t_j ; \quad (3.3.12)$$

dunque $C_{\underline{Y}}(k)$ è ottenuta da $C(k)$ con una convoluzione ed una anti-convoluzione con t_k .

Chiudiamo il paragrafo notando che vi sono altri modelli di covarianza direttamente derivabili dalle leggi di evoluzione dei processi, che considereremo più avanti.

3.4 Lo spettro di potenza ed il calcolo spettrale

Allo scopo di semplificare la presentazione seguente, faremo d'ora in poi l'ipotesi fondamentale che $C(k) \in \mathcal{T}_1$ ovvero che

$$\sum_{k=-\infty}^{+\infty} |C(k)| < +\infty . \quad (3.4.1)$$

Questa ipotesi significa tra l'altro che due componenti del processo stazionario \underline{X} , X_{n+k} ed X_n assai lontane nel tempo hanno tra loro una correlazione sempre più piccola, talché la condizione (3.4.1) sia verificata. In altri termini il processo \underline{X} perde memoria dei valori assunti a grande distanza nel tempo passato.

Sotto l'ipotesi (3.4.1) la serie di Fourier con coefficienti $C(k)$ è assolutamente uniformemente convergente, così che ha significato la definizione che diamo qui di seguito.

Definizione 3.4.1: Si chiama densità spettrale la funzione

$$f(p) = \sum_{k=-\infty}^{+\infty} C(k)e^{i2\pi kp} . \quad (3.4.2)$$

È ovvia conseguenza di quanto detto il seguente Lemma.

Lemma 3.4.1: la densità spettrale, che per semplicità chiameremo anche spettro, è una funzione continua e periodica di periodo 1.

Osserviamo ancora che essendo $C(k)$ una funzione pari, la (3.4.2) può essere riscritta come

$$f(p) = C_0 + 2 \sum_{k=1}^N C(k) \cos 2\pi kp(k) \quad (3.4.3)$$

il che dimostra che $f(p)$ è una funzione reale pari.

Alcune proprietà fondamentali dello spettro, oltreché il suo nome, derivano da un Lemma che qui dimostriamo dopo aver dato una nuova Definizione.

Definizione 3.4.2: chiameremo *periodogramma del processo stazionario \underline{X}* , la funzione

$$I_N(p) = \frac{1}{2N+1} \left| \sum_{k=-N}^N X_k e^{i2\pi kp} \right|^2 . \quad (3.4.4)$$

Ovviamente $I_N(p)$ è una funzione non negativa di p , continua e periodica di periodo 1. Naturalmente $I_N(p)$ è anche una variabile casuale funzione di \underline{X} , $\forall p$.

Lemma 3.4.2: risulta

$$\lim_{N \rightarrow \infty} E\{I_N(p)\} = f(p) ; \quad (3.4.5)$$

pertanto è necessariamente

$$f(p) \geq 0 \quad \forall p . \quad (3.4.6)$$

In sostanza $I_N(p)$ è uno stimatore asintoticamente corretto dello spettro, che risulta quindi essere una funzione non negativa.

□ Si ha

$$\begin{aligned}
E\{I_N(p)\} &= \frac{1}{2N+1} E \left\{ \sum_{k,j=-N}^N X_k X_j e^{i2\pi(k-j)p} \right\} = \\
&= \frac{1}{2N+1} E \left\{ \sum_{\ell=-N}^N X_\ell^2 + 2 \cos 2\pi p \sum_{\ell=-N}^{N-1} X_\ell X_{\ell+1} + \right. \\
&\quad \left. + 2 \cos 4\pi p \sum_{\ell=-N}^{N-2} X_\ell X_{\ell+2} + \dots \right\} \\
&= \frac{1}{2N+1} \{ (2N+1)C_0 + 2 \cos 2\pi p \cdot 2NC_1 + 2 \cos 4\pi p \cdot \\
&\quad \cdot (2N-1)C_2 + \dots \} = \\
&= C_0 + 2 \sum_{k=1}^N \cos 2\pi kp \frac{2N+1-k}{2N+1} C_k \tag{3.4.7}
\end{aligned}$$

Poiché risulta

$$\left| \cos 2\pi kp \frac{2N+1-k}{2N+1} \right| \leq 1$$

ed inoltre è $\sum |C(k)| < +\infty$ per ipotesi, nella (3.4.7) si può passare al limite ottenendo

$$\lim_{N \rightarrow \infty} E\{I_N(p)\} = C_0 + 2 \sum_{k=1}^{+\infty} \cos 2\pi kp C_k = f(p)$$

per la (3.4.3). □

Riassumendo le caratteristiche della densità spettrale (condizioni necessarie), per una covarianza che soddisfi la (3.4.1), si ha

- 1) $f(p)$ è continua e periodica di periodo 1
- 1) $f(p)$ è reale e pari

1) $f(p) \geq 0$, $\forall p$.

Osservazione 3.4.1: se pensiamo a X_n come ad un fenomeno ondulatorio nel tempo n , possiamo prendere $I_N(p)$ come una misura del quadrato dell'ampiezza dell'armonica di frequenza p contenuta in X_n e quindi la sua media, per N grande, assume il significato di potenza media dell'onda a frequenza p , da cui il nome di spettro di potenza.

Osservazione 3.4.2: se ogni covarianza in \mathcal{T}_1 ammette uno spettro che soddisfa le condizioni 1), 2), 3), non è vero il viceversa.

In particolare una funzione reale, continua, positiva pari e di periodo 1, non è detto che abbia coefficienti di Fourier

$$C_k = \frac{1 + \delta_{k0}}{2} \int_0^1 f(p) \cos 2\pi kp dp \quad (3.4.8)$$

che soddisfano la (3.4.1).

L'identificazione precisa della classe di queste funzioni $\{f(p)\}$ non è semplice, tuttavia è facile dare una condizione sufficiente affinché la (3.4.1) sia verificata.

Lemma 3.4.3: se $f(p)$ soddisfa 1), 2), 3) e se per di più $f'(p) \in \mathcal{L}^2(0, 1)$ allora $f(p)$ è lo spettro di una covarianza $\{C_k\} \in \mathcal{T}_1$.

□ Se è

$$f(p) = C_0 + 2 \sum_{k=1}^{+\infty} \cos 2\pi kp C_k ,$$

con C_k dati dalla (3.4.8) e per di più

$$f'(p) = -4\pi \sum_{k=1}^{+\infty} \sin 2\pi kp \cdot k C_k \in L^2 ,$$

allora deve essere

$$\sum_{k=1}^{+\infty} k^2 C_k^2 < +\infty .$$

Ma in tal caso, applicando la disuguaglianza di Schwarz

$$\sum_{k=-\infty}^{+\infty} |C_k| = C_0 + \sum_{k \neq 0}^{+\infty} |C_k| \cdot k \cdot \frac{1}{k} \leq C_0 + 2 \left\{ \sum_{k=1}^{+\infty} C_k \cdot k^2 \cdot \sum_{k=1}^{+\infty} \frac{1}{k^2} \right\}^{1/2} < +\infty$$

□

Esempio 3.4.1: sia \underline{X} un rumore bianco con covarianza

$$C_k = \delta_{k0} \sigma^2 ; \quad (3.4.9)$$

la densità spettrale in questo caso è data da

$$f(p) = \sigma^2 ,$$

cioè la potenza è la stessa per tutte le frequenze, da cui il nome di rumore bianco.

Esempio 3.4.2: sia \underline{X} un processo stazionario con covarianza finita (cfr. ad esempio il modello 2) del §3.3), ovvero

$$\{C_k\} = \{C_0, C_2, \dots, C_L, 0, 0 \dots\} \text{ per } k \geq 0 ;$$

in questo caso

$$f(p) = C_0 + 2 \sum_{k=1}^L C_k \cos 2\pi k p ; \quad (3.4.10)$$

dunque lo spettro in questo caso è un polinomio trigonometrico di ordine finito L .

Esempio 3.4.3: sia \underline{X} un processo con covarianza (cfr. il modello 4) del §3.3)

$$C_k = A_0 e^{-\alpha|k|} = A_0 \rho^{|k|} , \quad (\rho = e^{-\alpha} < 1) ;$$

in questo caso, sommando la corrispondente serie di Fourier, si trova lo spettro

$$f(p) = \frac{C_0}{1 - \gamma \cos 2\pi p} \quad (3.4.11)$$

$$C_0 = A_0 \frac{1 - e^{-2\alpha}}{1 + e^{-2\alpha}} \quad \gamma = \frac{2e^{-\alpha}}{1 + e^{-2\alpha}} (< 1) .$$

Si noti come anche in questo caso, essendo $\gamma < 1$, risulta $f(p) > 0 \forall p$.

Passiamo ora ad un risultato centrale di questo paragrafo che è la caratterizzazione delle funzioni di covarianza strettamente definite positive, cioè quelle per cui $C \asymp I$, in termini di densità spettrale.

Per provare questo risultato ci verrà utile un Lemma, che premettiamo

Lemma 3.4.4: *sia $\gamma(p)$ una funzione continua e limitata su $(-1/2, 1/2)$; supponiamo che $\forall (2N + 1)$ -pla di numeri complessi $(\lambda_{-N}, \dots, \lambda_N)$ valga*

$$\int_{-1/2}^{1/2} \gamma(p) \left| \sum_{k=-N}^N \lambda_k e^{i2\pi kp} \right|^2 dp \geq 0, \quad (3.4.12)$$

allora deve necessariamente essere

$$\gamma(p) \geq 0 \quad \forall p \in (-1/2, 1/2). \quad (3.4.13)$$

□ Infatti sia $\varphi(p)$ una funzione liscia come in Fig. 3.4.1.

Essendo $\varphi(p)$ ad esempio continua e limitata e così pure $\varphi'(p)$, in tutto $(-1/2, 1/2)$ e potendo far sì che così sia per $\sqrt{\varphi(p)}$, si può porre per opportuni ρ_k ,

$$\sqrt{\varphi(p)} = \sum_{k=-\infty}^{+\infty} \rho_k e^{i2\pi kp}$$

con la serie uniformemente convergente, dovendo essere $\sum |\rho_k| < +\infty$ (cfr. Lemma 3.4.3).

Fig. 3.4.1

Dunque sarà, uniformemente in p ,

$$\varphi(p) = \lim_{N \rightarrow \infty} \left| \sum_{k=-N}^N \rho_k e^{i2\pi k p} \right|^2 . \quad (3.4.14)$$

Ora supponiamo per assurdo che

$$\gamma(p_0) < 0 , \quad p_0 \in (-1/2, 1/2) ;$$

ma allora dovrà pure essere, per ε opportuno,

$$\gamma(p) < 0 \quad \forall p \in [p_0 - \varepsilon, p_0 + \varepsilon] \subset (-1/2, 1/2) ;$$

allora fissato ε e di conseguenza la $\varphi(p)$ come in (3.4.14), consideriamo

$$\begin{aligned} \varphi(p - p_0) &= \lim_{N \rightarrow \infty} \left| \sum_{k=-N}^N \rho_k e^{i2\pi k (p - p_0)} \right|^2 = \\ &= \lim_{N \rightarrow \infty} \left| \sum_{k=-N}^N \lambda_k e^{i2\pi k p} \right|^2 \end{aligned} \quad (3.4.15)$$

con

$$\begin{aligned} \lambda_k &= \rho_k e^{-i2\pi k p_0} \\ |\lambda_k| &= |\rho_k| . \end{aligned}$$

Essendo $\sum |\lambda_k| < +\infty$, anche il limite in (3.4.15) sarà uniforme in p e quindi si trova

$$\begin{aligned} & \lim_{N \rightarrow \infty} \int_{-1/2}^{1/2} \gamma(p) \left| \sum_{k=-N}^N \lambda_k e^{i2\pi kp} \right|^2 dp = \\ & = \int_{-1/2}^{1/2} \gamma(p) \varphi(p - p_0) dp < 0 \end{aligned}$$

che è in contraddizione con la (3.4.12). □

Possiamo ora enunciare il Teorema principale.

Teorema 3.4.1: *sia $f(p)$ lo spettro di un processo con covarianza C ; la relazione*

$$C \asymp I \quad (\alpha I \leq C \leq \beta I) \quad (3.4.16)$$

vale se e solo se

$$0 < \alpha \leq f(p) \leq \beta \quad (3.4.17)$$

□ Dimostriamo che (3.4.16) \rightarrow (3.4.17). La (3.4.16) significa che valgono le relazioni

$$\beta I - C \geq 0 ; \quad C - \alpha I \geq 0 . \quad (3.4.18)$$

Prendiamo ad esempio la prima delle (3.4.18); questa implica che $\forall (\lambda_{-N}, \dots, \lambda_N)$, complessi,⁹

$$\sum_{k,j=-N}^N \lambda_k^* \lambda_j [\beta \delta_{kj} - C_{k-j}] \geq 0 ; \quad (3.4.19)$$

ma ricordando che

$$\begin{aligned} C_k &= \int_{-1/2}^{1/2} f(p) e^{-i2\pi kp} dp , \\ \delta_{ko} &= \int_{-1/2}^{1/2} e^{-i2\pi kp} dp \end{aligned}$$

⁹Risolvendo i complessi λ_k in parte reale ed immaginaria, $\lambda_k = a_k + ib_k$, è facile vedere che questa definizione di definita positività è perfettamente equivalente a quella puramente reale, essendo C una matrice reale simmetrica.

la (3.4.19) dà

$$\sum_{k,j=-N}^N \lambda_k^* \lambda_j \int_{-1/2}^{1/2} [\beta - f(p)] e^{i2\pi(j-k)p} dp = \int_{-1/2}^{1/2} [\beta - f(p)] \left| \sum_{j=-N}^N \lambda_j e^{i2\pi j p} \right|^2 dp \geq 0$$

e quindi per il Lemma 3.4.4 concludiamo che

$$\beta - f(p) \geq 0 .$$

In modo analogo si prova che

$$f(p) - \alpha \geq 0 .$$

Ora viceversa proviamo che (3.4.17) \rightarrow (3.4.16). Dalle identità

$$\underline{\lambda}^* C \underline{\lambda} = \sum_{k,j=-N}^N \lambda_k^* \lambda_j C_{k-j} = \int_{-1/2}^{1/2} f(p) \left| \sum_{j=-N}^N \lambda_j e^{i2\pi j p} \right|^2 dp \quad (3.4.20)$$

si vede immediatamente che

$$\begin{aligned} f(p) \leq \beta &\Rightarrow \underline{\lambda}^* C \underline{\lambda} \leq \beta \int_{-1/2}^{1/2} \left| \sum_{j=-N}^N \lambda_j e^{i2\pi j p} \right|^2 dp = \beta \sum_{j=-N}^N |\lambda_j|^2 \\ f(p) \geq \alpha &\Rightarrow \underline{\lambda}^* C \underline{\lambda} \geq \alpha \int_{-1/2}^{1/2} \left| \sum_{j=-N}^N \lambda_j e^{i2\pi j p} \right|^2 dp = \alpha \sum_{j=-N}^N |\lambda_j|^2 \end{aligned}$$

□

Dunque il Teorema 3.4.1 ci dice che tutte le proprietà di regolarità di un processo \underline{X} e della matrice di covarianza C già commentate nel Cap. 2 e che si riassumono nella condizione $C \asymp I$, possono essere verificate sullo spettro $f(p)$ del processo con la semplice condizione (3.4.17). Tra l'altro si noti che per un processo stazionario che sulla diagonale di C ha la costante C_0 , le condizioni $C \asymp D, C \asymp I$ sono pienamente equivalenti. In particolare essendo $f(p)$ continua per ipotesi, la condizione $f(p) \leq \beta$ deve essere automaticamente soddisfatta; perciò è la condizione $f(p) \geq \alpha > 0$ che garantisce che \underline{X} sia una base di Riesz in $H_{\underline{X}}$.

Osservazione 3.4.3: ci si può chiedere cosa avvenga invece se si assume che $f(p)$ si annulli in un intervallino, ad esempio in $(p_0 - \varepsilon, p_0 + \varepsilon) \subset (-1/2, 1/2)$. In questo caso in effetti è facile vedere, usando la stessa $\varphi(p)$ definita nel Lemma 3.4.4, che

$$\underline{\lambda}^* C \underline{\lambda} = \int_{-1/2}^{1/2} f(p) \varphi(p) dp = 0$$

per un vettore complesso $\underline{\lambda}$ non identicamente nullo; di conseguenza la covarianza C non è definita positiva ma solo semidefinita e una delle componenti di \underline{X} può essere espressa come combinazione lineare delle altre, ciò che abbiamo escluso fin dall'inizio del Cap. 2. Resterebbe da caratterizzare il caso in cui $f(p)$ abbia degli zeri isolati o una successione di zeri, ma tale analisi fuoriesce dallo scopo di questa trattazione.

Osservazione 3.4.4: si consideri ancora una trasformazione $T \in \mathcal{T}_1$ di \underline{X} ,

$$\underline{Y} = T \underline{X} ; \quad (3.4.21)$$

con la formula (3.3.12) abbiamo caratterizzato la trasformazione dalla covarianza $C(k)$ di \underline{X} alla covarianza $C_{\underline{Y}}(k)$ di Y . Ci si chiede ora come si trasformi lo spettro $f(p)$ di \underline{X} , in quello $f_{\underline{Y}}(p)$ del nuovo processo \underline{Y} .

A questo scopo introduciamo la seguente definizione:

Definizione 3.4.3: data una T come in (3.4.21) definiamo funzione di trasferimento la

$$T(p) = \sum_{k=-\infty}^{+\infty} t_k e^{i2\pi kp} ; \quad (3.4.22)$$

Notiamo che la condizione $\sum_k |t_k| < +\infty$ garantisce la convergenza uniforme della serie a $T(p)$ che perciò in generale risulta essere una funzione continua e periodica di periodo 1, ma non reale, né con qualche tipo di simmetria.

Moltiplicando la (3.3.12) per $e^{i2\pi np}$ e sommando su tutti gli $n \in \mathcal{Z}$, sfruttando i teoremi standard sulle convoluzioni, otteniamo la relazione

$$f_{\underline{Y}}(p) = f(p) |T(p)|^2 \quad (3.4.23)$$

che è l'equivalente della propagazione della covarianza. Si noti che, come già sappiamo, vi sono molte trasformazioni che danno luogo alla stessa trasformazione spettrale e quindi alla stessa covarianza; in effetti se $V \in \mathcal{T}_1$ è tale che

$$V(p) = T(p)e^{i\Phi(p)} \quad (3.4.24)$$

con $\Phi(p)$ reale è chiaro che posto

$$\underline{W} = V\underline{X} ,$$

si ha

$$f_{\underline{W}}(p) = f(p)|V(p)|^2 = f(p)|T(p)|^2 = f_{\underline{Y}}(p) . \quad (3.4.25)$$

Un esempio importante di questa molteplicità lo si ha nella ricerca di un “filtro” (cioè di una trasformazione T) sbiancante per un processo \underline{X} dato, con spettro $f(p)$ che soddisfa la (3.4.17). Si vuole cioè una trasformazione T , tale che

$$\underline{\eta} = T\underline{X}$$

sia un rumore bianco (in senso debole) ovvero abbia una covarianza $C = I$ o uno spettro $f_{\eta}(p) = 1$.

Usando la (3.4.23) appare chiaro che

$$T(p) = [f(p)]^{-1/2} \quad (3.4.26)$$

è una soluzione al nostro problema, in cui $T(p)$ è una funzione reale simmetrica, continua e limitata proprio perché $f(p) \geq \alpha > 0$. D'altro canto come già ricordato nel Corollario al Teorema 3.2.2, vale la relazione

$$\underline{\varepsilon} = A\underline{X} , \quad (\varepsilon_n = \sum_{k=0}^{+\infty} a_k X_{n-k}) \quad (3.4.27)$$

dove $\underline{\varepsilon}$, l'innovazione di \underline{X} , è un processo di rumore bianco con varianza 1; di nuovo applicando la (3.4.23) si vede che deve essere

$$f_{\underline{\varepsilon}}(p) = 1 = f(p)|A(p)|^2 .$$

Peraltro la $A(p)$ così definita è data da

$$A(p) = \sum_{k=0}^{+\infty} a_k e^{i2\pi kp} \quad (3.4.28)$$

che, come appare chiaro, non è una funzione reale, né è simmetrica rispetto a p e quindi è una funzione certamente diversa dalla (3.4.26). Torneremo nel prossimo paragrafo su questo problema.

3.5 La predizione ottimale; calcolo di M ed A dallo spettro

Il problema della predizione lineare ottimale definito come la ricerca della variabile $(\widehat{X}_{n+k})_n$, proiezione di X_{n+k} su $H_{(n)}$, e dell'errore quadratico medio $\mathcal{E}_{n+k,n}^2 = \| X_{n+k} - (\widehat{X}_{n+k})_n \|^2$ è stato affrontato e risolto nel §2.7 in termini generali.

Riportiamo qui quei risultati tenendo conto che per un processo stazionario \underline{X} , tanto la matrice di covarianza C che M ed A sono matrici di Toeplitz. In sostanza abbiamo

$$(\widehat{X}_{n+k})_n = \sum_{j=-\infty}^n m_{n+k-j} \varepsilon_j = \sum_{\ell=k}^{+\infty} m_{\ell} \varepsilon_{n+k-\ell} , \quad (3.5.1)$$

$$\mathcal{E}_k^2 = \sum_{j=n+1}^{n+k} m_{n+k-j}^2 = \sum_{j=0}^{k-1} m_j^2 , \quad (3.5.2)$$

$$\varepsilon_{n+k-\ell} = \sum_{j=0}^{+\infty} a_j X_{n+k-\ell-j} ; \quad (3.5.3)$$

notiamo che le (3.5.1) (3.5.3) possono essere unite per fornire la formula

$$(\widehat{X}_{n+k})_n = \sum_{j=-\infty}^n \Gamma_{k,n-j} X_j = \sum_{j=0}^{+\infty} \Gamma_{k,j} X_{n-j} \quad (3.5.4)$$

dove

$$\Gamma_{k,n-j} = \sum_{h=0}^{n-j} m_{n-j-h+k} a_h . \quad (3.5.5)$$

Come si vede il successo di tale formule dipende dalla conoscenza di M ed A che a loro volta sono legate alla covarianza dalle relazioni

$$C = MM^+ , A = M^{-1} . \quad (3.5.6)$$

Poiché la funzione covarianza di un processo stazionario è stimabile anche da una sola realizzazione, come vedremo nel prossimo paragrafo, occorre trovare un modo per calcolare M ed A dalle (3.5.6). Un'idea è quella di passare attraverso le rappresentazioni spettrali delle (3.5.6), ovvero

$$f(p) = |M(p)|^2 , \quad (3.5.7)$$

$$A(p) = \frac{1}{M(p)} , \quad (3.5.8)$$

dove

$$M(p) = \sum_{k=0}^{+\infty} m_k e^{ik2\pi p} \quad (3.5.9)$$

$$A(p) = \sum_{k=0}^{+\infty} a_k e^{ik2\pi p} . \quad (3.5.10)$$

La soluzione del problema posto è fornita dal seguente Teorema che ha carattere costruttivo.

Teorema 3.5.1: *le funzioni di trasferimento $M(p), A(p)$ corrispondenti agli operatori M ed A definiti in (3.5.6), sono le soluzioni uniche delle equazioni (3.5.7), (3.5.8), (3.5.9), (3.5.10) con m_k, a_k reali quando lo spettro $f(p)$ sia limitato sia sopra che sotto*

$$0 < \alpha \leq f(p) \leq \beta . \quad (3.5.11)$$

□ La linea della dimostrazione è la seguente: nella parte a) prendiamo le (3.5.7), (3.5.8), (3.5.9), (3.5.10) come equazioni in $M(p)$ ed $A(p)$ e dimostriamo costruttivamente che esiste una soluzione corrispondente a successioni reali $\{m_k, k \geq 0\}, \{a_k, k \geq 0\}$. Nella parte b) prendiamo la successione $\{a_k\}$ trovata e dimostriamo che posto

$$\eta_n = \sum_{k=0}^{+\infty} a_k X_{n-k} , \quad (\underline{\eta} = A\underline{X}) \quad (3.5.12)$$

η_n coincide con l'innovazione ε_n ; ne segue che la matrice A così trovata coincide necessariamente con l'operatore autoregressivo che stiamo cercando e quindi $M = A^{-1}$ coincide con quello di media mobile.

a) Dato $f(p)$ definiamo

$$u(p) = \frac{1}{2} \log f(p) ;$$

u è una funzione continua e limitata, per la (3.5.11), perciò si può porre

$$u(p) = \sum_{n=-\infty}^{+\infty} a_n e^{in2\pi p} . \quad (3.5.13)$$

Poiché $u(p)$ è una funzione reale pari, al posto della (3.5.13) si può scrivere

$$u(p) = u_0 + 2 \sum_{n=1}^{+\infty} u_n \cos n2\pi p . \quad (3.5.14)$$

Si osservi che se si suppone che $f(p)$ abbia una derivata in \mathcal{L}^2 , altrettanto vale per $u(p)$ e quindi si può in questo caso ritenere che $\sum |u_n| < +\infty$.

Notiamo che se si pone $\vartheta = 2\pi p$, $u(p)$ può essere pensata come traccia sulla circonferenza unitaria $\{r = 1\}$ della funzione

$$u(r, \vartheta) = u_0 + 2 \sum_{n=1}^{+\infty} u_n r^n \cos n\vartheta . \quad (3.5.15)$$

La (3.5.15) è notoriamente una funzione armonica in $\{r \leq 1\}$ e quindi raggiunge i suoi valori massimi e minimi sul contorno così che

$$\frac{1}{2} \log \alpha \leq u(r, \vartheta) \leq \frac{1}{2} \log \beta . \quad (3.5.16)$$

Ora poniamo

$$\psi(z) = u_0 + 2 \sum_{n=1}^{+\infty} u_n r^n e^{in\vartheta} = u_0 + 2 \sum_{n=1}^{+\infty} u_n z^n \quad (3.5.17)$$

ove

$$z = r e^{i\vartheta} = r e^{i2\pi p} .$$

È chiaro che $\psi(z)$ è analitica in $\{r \leq 1\}$ se la serie (3.5.17) è convergente e ciò è vero se, come supponiamo, $\sum |u_n| < \infty$. Notiamo che $u(z) = \operatorname{Re}\psi(z)$. Finalmente poniamo

$$M(z) = e^{\psi(z)} ; \quad (3.5.18)$$

poiché $\psi(z)$ è una funzione analitica in $\{r \leq 1\}$ altrettanto è $M(z)$ così che vale la rappresentazione di Taylor

$$M(z) = \sum_{k=0}^{+\infty} m_k z^k \quad (3.5.19)$$

Inoltre da

$$|M(z)| = e^{u(z)} \quad (3.5.20)$$

e dalla (3.5.16) si vede che

$$|M(z)| > \sqrt{\alpha} , \quad (3.5.21)$$

così che $M(z)$ non è mai nulla nel cerchio $\{r \leq 1\}$. Infine risulta pure

$$|M(z)|_{r=1}^2 = e^{2u(p)} = e^{\log f(p)} = f(p) ,$$

cioè la funzione

$$M(p) = \sum_{k=0}^{+\infty} m_k e^{in2\pi p}$$

è proprio tale che

$$|M(p)|^2 = f(p) . \quad (3.5.22)$$

Dunque la $M(p)$ così costruita è soluzione di (3.5.7), (3.5.9). Inoltre per la (3.5.21) si può anche porre

$$A(z) = \frac{1}{M(z)} \quad (3.5.23)$$

ed affermare che pure $A(z)$ è analitica regolare nel cerchio $\{r \leq 1\}$, così che si avrà

$$A(z) = \sum_{k=0}^{+\infty} a_k z^k .$$

Ora preso $z = e^{i2\pi p}$ si trova

$$A(p) = \frac{1}{M(p)} = \sum_{k=0}^{+\infty} a_k e^{ik2\pi p} . \quad (3.5.24)$$

Si può notare che mantenendo le ipotesi di regolarità fin qui riportate, si può anche affermare che $\sum_{k=0}^{+\infty} |a_k| < +\infty$, cioè $A(p) \in \mathcal{T}_1$;

b) notiamo che $\sum_{k=0}^{+\infty} |a_k| < +\infty \Rightarrow \sum_{k=0}^{+\infty} a_k^2 < +\infty$ così che la serie (3.5.12) è convergente, in quanto la condizione (3.5.11) garantisce che $C \asymp I$. Vogliamo dimostrare che $\eta_n = \varepsilon_n$.

In primo luogo notiamo che

$$\eta_n = \sum_{k=0}^{+\infty} a_k X_{n-k} \Rightarrow \eta_n \in H_n .$$

Inoltre

$$f_\eta(p) = |A(p)|^2 f(p) = 1$$

il che ci dice che $\underline{\eta}$ è un rumore bianco in senso debole, ovvero una successione ortonormale, così che è anche

$$\|\eta_n\| = 1 .$$

Infine osserviamo che

$$\begin{aligned} \langle \eta_n, X_{n-j} \rangle &= \sum_{k=0}^{+\infty} a_k C_{j-k} = \\ &= \sum_{k=0}^{+\infty} a_k \sum_{i=0}^{+\infty} m_i m_{i+j-k} = \end{aligned}$$

$$= \sum_{i=0}^{+\infty} m_i \sum_{k=0}^{+\infty} a_k m_{i+j-k} . \quad (3.5.25)$$

D'altro canto dall'identità

$$\begin{aligned} A(p)M(p) &= \sum_{k,h=0}^{+\infty} a_k m_h e^{i(k+h)2\pi p} = \\ &= \sum_{n=0}^{+\infty} e^{in2\pi p} \left(\sum_{k=0}^{+\infty} a_k m_{n-k} \right) \equiv 1 \end{aligned}$$

discende

$$\sum_{k=0}^{+\infty} a_k m_{n-k} = \sum_{k=0}^n a_k m_{n-k} = \delta_{n,0} ,$$

così che la (3.5.25) dà

$$\langle \eta_n, X_{n-j} \rangle = \sum_{\ell=0}^{+\infty} m_\ell \delta_{\ell+j,0} \equiv 0, \quad \forall j \geq 1$$

perché $\ell + j \geq 1$ così che $\delta_{\ell+j,0} \equiv 0$.

Ma allora η_n è un vettore in H_n di norma 1 e ortogonale ad H_{n-1} e perciò

$$\eta_n \equiv \varepsilon_n , \quad \forall n$$

ovvero la matrice A coincide con quella per cui

$$\underline{\varepsilon} = A\underline{X} .$$

Tale matrice infatti è unica perché \underline{X} è una base di Riesz. \square

Il risultato di questo paragrafo sta nell'aver dimostrato, anche in maniera costruttiva, come calcolare la predizione lineare ottimale di un processo, noto fino al tempo n , quando se ne conoscano la covarianza $C(k)$, ovvero lo spettro $f(p)$. Nel prossimo paragrafo ci occuperemo della stima di queste grandezze a partire da una realizzazione di \underline{X} .

3.6 Problemi di stima della funzione di covarianza e dello spettro

Supponiamo che di un processo stazionario \underline{X} sia nota una realizzazione, per un numero limitato di tempi N , che chiameremo

$$\underline{x}_N^+ = \{x_1, x_2 \dots x_N\} ; \quad (3.6.1)$$

spesso ci verrà comodo pensare a questo vettore prolungato con valori nulli nel tempo in entrambe le direzioni, ponendo

$$\underline{x}_0^+ \equiv \{\dots 0, x_1, \dots x_N, 0, \dots\} ; \quad (3.6.2)$$

ci poniamo il problema di stimare $C(k)$ e/o $f(p)$ a partire da \underline{x}_N , tenendo presente che la proprietà per noi più importante degli stimatori sarà la consistenza, in quanto è noto che una stima di una covarianza da un campione non numeroso può contenere errori rilevanti.

a) Stima della funzione di covarianza

Uno stimatore ovviamente corretto di C_k è dato da

$$\hat{C}_k = \frac{1}{N-k} \sum_{j=1}^{N-k} X_j X_{j+k}, \quad (k < N) ; \quad (3.6.3)$$

questo stimatore sfrutta a fondo il concetto di stazionarietà utilizzando una replica traslata di k tempi del vettore \underline{x}_N come una nuova realizzazione del processo.

I valori empirici \hat{C}_k ottenuti dalla (3.6.3) tuttavia in genere non possono essere direttamente presi come se appartenessero ad una funzione di covarianza come mostra il seguente controesempio.

Esempio 3.6.1: sia $N = 3$, $\underline{x}_3 = \{1, 0, 1\}$; mediante la (3.6.3) si possono stimare i tre valori

$$\hat{C}_0 = \frac{2}{3}, \quad \hat{C}_1 = 0, \quad \hat{C}_2 = 1 ,$$

che non possono appartenere ad una funzione di covarianza in quanto $\hat{C}_2 > \hat{C}_0$.

Tuttavia i valori empirici \widehat{C}_k , che costituiscono la cosiddetta funzione di covarianza empirica, possono essere interpolati con un qualche modello appartenente ad una delle famiglie parametriche illustrate nel §3.3 o ad una loro combinazione.

Osservazione 3.6.1: i valori empirici \widehat{C}_k di solito non vengono stimati per valori di k che vanno oltre il 20% di N poiché all'aumentare di k diminuisce il numero di coppie disponibili per formare \widehat{C}_k e la stima diviene più erratica; inoltre ad un intervallo di $0, 2N$ la funzione di covarianza deve già aver mostrato tutti i suoi caratteri salienti, se no vuol dire che il numero di dati disponibili non è sufficiente per svolgere il tipo di stima richiesto.

Esempio 3.6.2: si prenda la funzione di covarianza empirica in Fig. 3.6.1. Essa è stata generata da un campione di numerosità $N = 100$ di un processo normale con covarianza

$$C(k)e^{-0,1|k|} .$$

Come si vede la funzione empirica tende a zero senza mostrare oscillazioni significative e con un marcato approccio a cuspide nell'origine. Scelto perciò il modello $C_k = C_0e^{-\alpha|k|}$ si sono stimati i parametri C_0 ed α ottenendo l'interpolazione mostrata pure in Fig. 3.6.1

Fig. 3.6.1 Interpolazione della funzione di covarianza empirica con il modello $C_k = C_0e^{-\alpha|k|}$; $C_0 = 1, \alpha = 0.1$.

In Fig. 3.6.2 e Fig. 3.6.3 mostriamo esperimenti analoghi per i modelli

$$C(k) = C_0e^{-\alpha k^2} (C_0 = 1 \alpha = 0.1)$$

e

$$C(k) = C_0e^{-\alpha|k|} \cos \omega k, (C_0 = 1 \alpha = 0.1 \omega = \frac{\pi}{10}) .$$

Fig. 3.6.2 Interpolazione della funzione di covarianza empirica con il modello $C_k = C_0 e^{-\alpha k^2}$; $C_0 = 1, \alpha = 0.1$.

Fig. 3.6.3 Interpolazione della funzione di covarianza empirica con il modello $C_k = C_0 e^{-\alpha|k|} \cos \omega k$; $C_0 = 1, \alpha = 0.1$.

Nel primo caso la scelta della funzione interpolante della famiglia corretta è suggerita dalla velocità di annullamento, dall'assenza di oscillazioni significative e della forma sostanzialmente piatta nell'origine. Nel secondo caso la scelta di una esponenziale coseno è dettata dalla cadenza di zeri, massimi e minimi e del comportamento nell'origine.

Se anziché usare la procedura sopra descritta, si cercasse al posto di (3.6.3) uno stimatore empirico \tilde{C}_k che possa essere usato direttamente come funzione di covarianza, si potrebbe porre

$$\tilde{C}_k = \frac{1}{N} \sum_{j=1}^{N-k} X_j X_{j+k} . \quad (3.6.4)$$

Lemma 3.6.1: *lo stimatore \tilde{C}_k dato della (3.6.4) dà luogo ad una funzione definita positiva.*

□ In effetti sia \underline{x}_0 come in (3.6.2). Notiamo che

$$\begin{aligned} N\tilde{C}_{-\ell} &= \underline{x}_0^+ B^\ell \underline{x}_0 = \sum_{k=\ell+1}^N x_k x_{k-\ell} = \\ &= \underline{x}_0^+ B^{-\ell} \underline{x}_0 = \sum_{k=1}^{N-\ell} x_k x_{k-\ell} = N\tilde{C}_\ell ; \end{aligned}$$

pertanto, ricordando che $B^{-1} = B^+ = F$,

$$\begin{aligned} N \sum_{i,k=1}^N \lambda_i \lambda_k \tilde{C}_{i-k} &= N \sum_{i,k=1}^N \lambda_i \lambda_k \underline{x}_0^+ B^{i-k} \underline{x}_0 = \\ &= N \underline{x}_0^+ \left[\sum_{k=1}^N \lambda_k B^k \right]^+ \left[\sum_{i=1}^N \lambda_i B^i \right] \underline{x}_0 \geq 0 . \end{aligned}$$

□

Si può anche notare che \tilde{C}_k , benché non sia uno stimatore corretto di C_k , lo è almeno asintoticamente in quanto

$$E\{\tilde{C}_k\} = \frac{N-k}{N} C_k \xrightarrow{(N \rightarrow \infty)} C_k . \quad (3.6.5)$$

D'altro canto se \tilde{C}_k è uno stimatore consistente di C_k , altrettanto lo è $\tilde{C}_k = \frac{N-k}{N} \hat{C}_k$; pertanto \tilde{C}_k è uno stimatore accettabile di C_k , a condizione che sia consistente. A questa questione risponde un teorema dovuto a Bartlett.

Teorema 3.6.1: *supponiamo che \underline{X} sia un processo stazionario del tipo*

$$\underline{X} = A\xi \quad (3.6.6)$$

con ξ rumore bianco in senso forte (ovvero le componenti $\xi_k, \xi_j, k \neq j$ sono tra loro stocasticamente indipendenti) e con $A \in \mathcal{T}_1$, ovvero

$$X_n = \sum_{k=-\infty}^{+\infty} a_{n-k} \xi_k , \quad \sum_{k=-\infty}^{+\infty} |a_k| < +\infty ; \quad (3.6.7)$$

supponiamo inoltre che ξ_j abbia momento quarto finito

$$E\{\xi_j^4\} < +\infty ; \quad (3.6.8)$$

allora vale la formula asintotica¹⁰

$$\sigma(\tilde{C}_\ell, \tilde{C}_{\ell+h}) = \frac{1}{N} \left\{ \sum_{k=-\infty}^{+\infty} (C_k C_{k+h} + C_{k+\ell+h} C_{k-\ell}) \right\}$$

¹⁰In effetti se $A \in \mathcal{T}_1$ anche $C = AA^+ \in \mathcal{T}_1$, come si è visto nel Lemma 3.2.2; pertanto

$$\sum |C_\ell| < +\infty \Rightarrow \sum C_\ell^2 < +\infty$$

così che le serie in (3.6.9) sono convergenti.

$$+ (M_4 - 3)C_\ell C_{\ell+h} \} + o\left(\frac{1}{N}\right) . \quad (3.6.9)$$

□ Notiamo che la (3.6.9) per $h = 0$ ci dice che

$$\sigma(\tilde{C}_\ell) = o\left(\frac{1}{N}\right) ; \quad (3.6.10)$$

questa unita al fatto che

$$\lim_{N \rightarrow \infty} E\{\tilde{C}_\ell\} = C_\ell \quad (3.6.11)$$

ci permette di affermare che $\tilde{C}_\ell \rightarrow C_\ell$ in $\mathcal{L}^2(\Omega)$ e dunque anche in probabilità, così che \tilde{C}_ℓ è uno stimatore consistente di C_ℓ .

Ci limitiamo ad accennare la dimostrazione del Teorema solo nel caso $X_n = \xi_n, (A = I)$. In questo caso, quando $\ell \neq 0$,

$$E\{\tilde{C}_\ell^2\} = \frac{1}{N^2} \sum_{i,k=1}^N E\{\xi_i \xi_{i+\ell} \xi_k \xi_{k+\ell}\} .$$

Notiamo che se $i \neq k$ il secondo termine è $\equiv 0$, mentre per $i = k (\ell \neq 0)$ risulta

$$E\{\xi_i \xi_{i+\ell} \xi_k \xi_{i+\ell}\} = E\{\xi_i^2\} E\{\xi_{i+\ell}^2\} = 1 .$$

Pertanto, ricordando che $C_\ell = 0$ perché $X_n = \xi_n$,

$$E\{\tilde{C}_\ell^2\} = \sigma^2(\tilde{C}_\ell) = \frac{N - \ell}{N^2} = \frac{1}{N} + o\left(\frac{1}{N}\right)$$

come risulta anche nella (3.6.9) quando $C_\ell = \delta_{\ell,0}$.

Se ora invece prendiamo

$$\begin{aligned} E\{\tilde{C}_0^2\} &= \frac{1}{N^2} \sum_{i,k=1}^N E\{\xi_i^2 \xi_k^2\} = \\ &= \frac{1}{N^2} \sum_{i,k=01}^N [(M_4 - 1)\delta_{ik} + 1] = 1 + \frac{M_4 - 1}{N} , \end{aligned}$$

ricordando che $C_0 = 1$ (essendo $X_n = \xi_n$) si vede che tale relazione coincide con la (3.6.9), essendo

$$\sigma^2(\tilde{C}_0) = \frac{M_4 - 1}{N} .$$

□

b) Stima dello spettro

Naturalmente ogni stima di C_k che sia anche definita positiva, come \tilde{C}_k , dà luogo ad una stima della densità spettrale $f(p)$ con la formula

$$\tilde{f}(p) = \tilde{C}_0 + 2 \sum_{k=-1}^{N-1} \tilde{C}_k \cos k2\pi p \equiv \sum_{k=-N+1}^{N-1} \tilde{C}_k e^{ik2\pi p} \quad (3.6.12)$$

Vale il seguente Lemma.

Lemma 3.6.2: *risulta*

$$\tilde{f}(p) = I_N(p) . \quad (3.6.13)$$

Il periodogramma $I_N(p)$ è definito in (3.4.4).

□ Questa identità è già stata dimostrata nel Lemma 3.4.2. □

Ricordando la proprietà (3.4.5) di $I_N(p)$ si vede che $\tilde{f}(p)$ è uno stimatore asintoticamente corretto. Tuttavia sfortunatamente $\tilde{f}(p)$ non è uno stimatore consistente di $f(p)$ puntualmente in p , come è dimostrato dal seguente controesempio.

Esempio 3.6.3: sia $\{\xi_n\}$ un rumore bianco gaussiano. Posto

$$\begin{aligned} I_N(p) &= \frac{1}{N} \left| \sum_{k=1}^N \xi_k e^{ik2\pi p} \right|^2 = \\ &= \frac{1}{N} \left[\left(\sum_{k=1}^N \xi_k \cos k2\pi p \right)^2 + \left(\sum_{k=1}^N \xi_k \sin k2\pi p \right)^2 \right] , \end{aligned}$$

notiamo che le due variabili

$$U_N = \sum_{k=1}^N \xi_k \cos k2\pi p , \quad V_N = \sum_{k=1}^N \xi_k \sin k2\pi p$$

sono congiuntamente normali, con media nulla ed inoltre, preso $p = \frac{n}{N} = \frac{r}{s}$ con $k < s$ interi e successivamente $N = Ms$, $n = Mr$ in modo che p resti costante $\forall M$, si ha

$$E \left\{ U_N^2 \left(\frac{n}{N} \right) \right\} = \sum_{k=1}^N \cos^2 \frac{k2\pi n}{N} = \frac{N}{2}$$

$$E \left\{ V_N^2 \left(\frac{n}{N} \right) \right\} = \sum_{k=1}^N \sin^2 \frac{k2\pi n}{N} = \frac{N}{2} .$$

Ancora, vale

$$E \{ U_N V_N \} = \sum_{k=1}^N \cos \frac{k2\pi n}{N} \sin \frac{k2\pi n}{N} = 0$$

di conseguenza si vede che

$$\begin{aligned} \tilde{f} \left(\frac{r}{s} \right) &= I_N \left(\frac{r}{s} \right) = \frac{1}{2} \left[\frac{2}{N} U_N^2 + \frac{2}{N} V_N^2 \right] = \\ &= \frac{1}{2} \left[\frac{U_N^2}{\sigma^2(U_N)} + \frac{V_N^2}{\sigma^2(V_N)} \right] = \frac{1}{2} \chi_2^2 \end{aligned}$$

indipendentemente da $N = Ms$. Ne segue che $\tilde{f} \left(\frac{r}{s} \right)$ non può convergere ad una costante, $f \left(\frac{r}{s} \right)$, per $N = Ms$.

Sebbene non valga una convergenza in probabilità di $\tilde{f}(p)$ ad $f(p)$ per ogni p , tuttavia si può dimostrare che si ha sempre una convergenza in media quadratica di funzionali più regolari di $\tilde{f}(p)$ agli stessi funzionali di $f(p)$. Più precisamente vale il seguente Lemma che enunciamo senza dimostrazione:

Lemma 3.6.3: $\forall \varphi(p)$ misurabile limitata si ha

$$\lim_{N \rightarrow \infty} E \left\{ \int_0^1 \tilde{f}(p) \varphi(p) dp \right\} = \int_0^1 f(p) \varphi(p) dp, \quad (3.6.14)$$

$$\lim_{N \rightarrow \infty} \sigma^2 \left\{ \int_0^1 \tilde{f}(p) \varphi(p) dp \right\} = 0 . \quad (3.6.15)$$

Il Lemma 3.6.3 afferma essenzialmente che una versione lisciata di $\tilde{f}(p)$ tende in media quadratica ad un'analogo versione lisciata di $f(p)$.

Sulla scorta del Lemma 3.6.3, si è costruita una teoria per gli stimatori "lisciati" dello spettro.

Definizione 3.6.1: *un lisciatore o finestra è una funzione $w(p)$ che per semplicità supponiamo pari, continua con tutte le derivate richieste, tale che*

$$w(p) = 0 \quad |p| > \delta \quad (3.6.16)$$

e tale che

$$\int_{-1/2}^{1/2} w(p) dp = 1 . \quad (3.6.17)$$

Un esempio tipico è dato in Fig. 3.6.4

Dato $w(p)$ si può generare da questa una famiglia parametrica, traslando il picco in \bar{p} e cambiando l'ampiezza di $w(p)$ pur mantenendo la condizione (3.6.17);

$$w_\alpha(p - \bar{p}) = \alpha w[\alpha(p - \bar{p})] . \quad (3.6.18)$$

Notiamo che fissato \bar{p} e $\forall \alpha$ fisso $w_\alpha(p - \bar{p})$ è una funzione continua e limitata, così che in base al Lemma 3.6.3 risulta che

$$\bar{f}_{N,\alpha}(p) = \int_{-1/2}^{1/2} w_\alpha(p - \bar{p}) \tilde{f}(\bar{p}) d\bar{p} \quad (3.6.19)$$

è uno stimatore consistente di

$$\bar{f}_\alpha(p) = \int_{-1/2}^{1/2} w_\alpha(p - \bar{p}) f(\bar{p}) d\bar{p} . \quad (3.6.20)$$

D'altro canto, essendo $f(p)$ continua, appare chiaro che

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \bar{f}_\alpha(p) &= \lim_{\alpha \rightarrow \infty} \int_{1/2}^{1/2} w_\alpha(p - \bar{p}) f(\bar{p}) d\bar{p} = \\ &= \lim_{\alpha \rightarrow \infty} \int_{-\infty}^{+\infty} w(q) f\left(p - \frac{q}{\alpha}\right) dq = f(p) , \end{aligned}$$

così che si è pensato che agendo contemporaneamente su α e su N si potesse ottenere una successione \bar{S}_{N,α_N} che tenda in media quadratica ad $f(p)$. In effetti vale il seguente Lemma, che non dimostriamo.

Lemma 3.6.4: sia $f(p)$ lo spettro corrispondente ad una covarianza $C \in \mathcal{T}_1$ e sia α_N tale che

$$\alpha_N \rightarrow +\infty, \quad \frac{\alpha_N}{N} \rightarrow 0; \quad (3.6.21)$$

allora, $\forall p$

$$\begin{aligned} \lim_{N \rightarrow \infty} E\{\bar{f}_{N,\alpha_N}\} &= f(p) \\ \lim_{N \rightarrow \infty} \sigma^2(\bar{f}_{N,\alpha_N}) &= 0. \end{aligned}$$

Osservazione 3.6.2: l'effetto dell'uso di $w_\alpha(p - \bar{p})$ per costruire $\bar{f}_{N,\alpha}(p)$ può essere anche visto in termini di covarianza. In effetti ci si può chiedere quale sia la covarianza $\bar{C}_\alpha(k)$ corrispondente a $\bar{f}_{N,\alpha}(p)$. Poiché $w_\alpha(p)$ è una funzione pari si potrà porre

$$w_\alpha(p) = \sum_{n=0}^{+\infty} w_{n,\alpha} \cos 2\pi np;$$

sostituendo questa relazione nella (3.6.19) vediamo che

$$\bar{f}_{N,\alpha}(p) = \tilde{C}_0 w_{0,\alpha} + 2 \sum_{n=1}^{N-1} \tilde{C}_n w_{n,\alpha} \cos n2\pi p.$$

Questa equazione ci dice che è semplicemente

$$\bar{C}_\alpha(k) = \tilde{C}_k w_{k,\alpha}. \quad (3.6.22)$$

Esempio 3.6.4: esempi di alcune tra le finestre più usate per le stime spettrali sono associate ai coefficienti di Fourier:

$$\begin{aligned} \text{Finestra di Tuckey} \quad w_k = w_{-k} &= \begin{cases} \frac{1}{2} \left(1 + \cos \frac{\pi k}{M}\right) & 0 \leq k \leq M \\ 0 & k > M \end{cases} \\ \text{Finestra di Parzen} \quad w_k = w_{-k} &= \begin{cases} 1 - 6 \left(\frac{k}{M}\right)^2 + 6 \left(\frac{k}{M}\right)^3 & 0 \leq k \leq \frac{M}{2} \\ 2 \left(1 - \frac{k}{M}\right)^3 & \frac{M}{2} < k \leq M \\ 0 & k > M \end{cases} \\ \text{Finestra di Daniell} \quad w_k = w_{-k} &= \begin{cases} 1 & k = 0 \\ \frac{\sin \frac{Mk}{N} 2\pi}{M \sin \frac{k}{N} 2\pi} & 0 < k \leq N - 1 \end{cases} \end{aligned}$$

In tutti questi casi M prende il posto del parametro α ; una regola empirica, ma ben funzionante per la scelta di M è

$$M \cong 2\sqrt{N} .$$

In tutti i casi come si vede, per k fisso

$$\lim_{M \rightarrow \infty} w_k(M) = 1 .$$

3.7 Processi semplici

Definiamo semplici quei processi \underline{X} la cui funzione di trasferimento tra innovazione $\underline{\varepsilon}$ ed \underline{X} ,

$$M(p) = M(z)|_{z=e^{i2\pi p}} \quad (3.7.1)$$

ha la forma di una funzione trigonometrica razionale, ovvero

$$M(z) = \frac{P_L(z)}{Q_h(z)} , \quad (3.7.2)$$

con P_L, Q_h polinomi rispettivamente di grado L ed h .

Ricordando la discussione del §3.5, imponiamo ad $M(z)$ di essere una funzione analitica in tutto il cerchio $\{|z| \leq 1\}$ ed anche ivi $|M(z)| \neq 0$ oltre a soddisfare la (3.5.7), ovvero

$$|M(e^{i2\pi p})| = M(e^{i2\pi p})M(e^{-i2\pi p}) = f(p) . \quad (3.7.3)$$

Ne deduciamo in generale che P_L, Q_h devono necessariamente soddisfare le condizioni

$$P_L(z) \neq 0 , Q_h(z) \neq 0 \text{ in } \{|z| \leq 1\} ; \quad (3.7.4)$$

inoltre è logico usare per $M(z)$ una rappresentazione irriducibile così che supporremo che P_L e Q_h non abbiano zeri in comune; infine notando che P_L e Q_h possono essere moltiplicati per una costante senza modificare $M(z)$, possiamo scegliere arbitrariamente un loro coefficiente (purché $\neq 0$) ed in specifico assumeremo

$$q_0 \equiv 1 . \quad (3.7.5)$$

Osservazione 3.7.1: notiamo che se in $M(z)$ si sostituisce z con B si ottiene proprio l'operatore M matriciale per cui

$$\underline{X} = M\underline{\varepsilon} = M(B)\underline{\varepsilon}$$

ovvero l'operatore in $H_{\underline{X}}$ per cui

$$X_n = \mathcal{M}\varepsilon_n = M(\mathcal{B})\varepsilon_n = \frac{P_L(\mathcal{B})}{Q_h(\mathcal{B})}\varepsilon_n . \quad (3.7.6)$$

La (3.7.6) moltiplicata per $Q_h(\mathcal{B})$ dà

$$Q_h(\mathcal{B})X_n = P_L(\mathcal{B})\varepsilon_n \quad (3.7.7)$$

ovvero

$$X_n + q_1X_{n-1} + \dots + q_hX_{n-h} = p_0\varepsilon_n + p_1\varepsilon_{n-1} + \dots + p_L\varepsilon_{n-L} ; \quad (3.7.8)$$

un processo che segue una legge del tipo (3.7.8) viene anche detto autoregressivo e di media mobile (ARMA dall'inglese autoregressive-moving average). La parte $Q_h(B)$ è detta la parte autoregressiva dell'operatore mentre $P_L(B)$ è detta la parte di media mobile.

I processi per cui $Q_h \equiv 1$ sono detti di pura media mobile, o sinteticamente MA, mentre quelli per cui $P_L \equiv p_0$ sono detti puramente autoregressivi, ovvero AR. Per sfruttare meglio le loro specificità, dividiamo la trattazione dei MA degli AR.

a) Processi a media mobile

In questo caso (cfr. (3.4.3), (3.2.25))

$$M(z) \equiv P_L(z) = p_0 + p_1z + \dots + p_Lz^L , \quad m_k = p_k \quad (3.7.9)$$

$$f(p) = \left| \sum_{k=0}^L m_k e^{i2\pi kp} \right|^2 = C_0 + 2 \sum_{k=1}^L C_k \cos 2\pi kp , \quad (3.7.10)$$

$$C_k = \sum_{\ell=k}^L m_\ell m_{\ell-k} \quad (0 \leq k \leq L) . \quad (3.7.11)$$

Si noti che la caratteristica di questo processo è che

$$C_k \equiv 0, \quad k > L \quad (3.7.12)$$

cioè il processo perde completamente memoria del passato dopo un intervallo di tempo maggiore di L , cosa questa che aiuta anche ad identificare questi processi partendo da una covarianza empirica.

L'equazione di evoluzione del processo è

$$X_n = \sum_{k=0}^L m_k \varepsilon_{n-k}, \quad (3.7.13)$$

cioè appunto quella di una media mobile monolaterale (solo verso il passato). Il problema della predizione ottimale per \underline{X} , dati $(X_1, X_2 \dots X_N)$ è invece più complesso di quanto sembri; infatti se da un lato scrivendo la (3.7.13) nella forma

$$X_{N+j} = \sum_{k=0}^{j-1} m_k \varepsilon_{N+j-k} + \sum_{k=j}^L m_k \varepsilon_{N+j-k} \quad (j \leq L) \quad (3.7.14)$$

si rende subito evidente che il termine

$$\sum_{k=j}^{j-1} m_k \varepsilon_{N+j-k} \perp H_N,$$

così che esso entrerà per forza a far parte dell'errore di stima, non si deve commettere l'errore di credere che $\varepsilon_N, \varepsilon_{N-1} \dots \in H_{1,N} = \text{Span}\{X_1, X_2 \dots X_N\}$. Infatti, ad esempio, l'innovazione ε_N dipende in generale anche da $X_0, X_{-1} \dots$ così che la sua proiezione ortogonale su $H_{1,N}$ non è banale da ottenersi. Tuttavia, come mostreremo esplicitamente negli esempi, quando N è grande rispetto ad L è possibile ottenere una soddisfacente approssimazione degli ε necessari nella (3.7.14) in termini di $X_1, \dots X_N$. Notiamo anche che chiaramente per $j > L$

$$P_{1,N} X_{N+j} = 0 \quad (j > L) \quad (3.7.15)$$

$P_{1,N} =$ proiettore ortogonale su $H_{1,N}$

così che

$$\mathcal{E}_{j;1,N}^2 = m_0^2 + \dots + m_L^2 \quad (j > L) \quad (3.7.16)$$

Esempio 3.7.1: (MA1). In questo caso

$$M(z) = M_0 + m_1 z$$

e l'equazione di evoluzione è

$$X_n = m_0 \varepsilon_n + m_1 \varepsilon_{n-1}$$

Così le (3.7.11) danno

$$\begin{cases} C_0 = m_0^2 + m_1^2 \\ C_1 = m_0 m_1 \\ C_k = 0 \end{cases} \quad (k > 1) \quad (3.7.17)$$

La condizione (3.7.4) limita le eventuali soluzioni della (3.7.17) a quelle che soddisfano la condizione

$$|m_1| < |m_0| ; \quad (3.7.18)$$

notiamo che in particolare si può sempre assumere

$$m_0 > 0 , \quad (3.7.19)$$

poiché tutt'al più si può cambiare segno ad m_0, m_1 e ad $\{\varepsilon_n\}$, il che naturalmente non cambia la natura di rumore bianco di tale successione.

Notiamo anche che della (3.7.17) si può scrivere

$$\begin{cases} C_0 + 2C_1 = (m_0 + m_1)^2 \\ C_0 - 2C_1 = (m_0 - m_1)^2 \end{cases} \quad (3.7.20)$$

il che prova anche che deve essere

$$|C_1| < \frac{1}{2} C_0 ;$$

questa condizione peraltro è anche necessaria perché lo spettro

$$f(p) = C_0 + 2C_1 \cos 2\pi p$$

sia strettamente positivo.

Dalle (3.7.20) possiamo ricavare m_0, m_1 in funzione della covarianza C_0, C_1 ; la sola soluzione che soddisfi le condizioni (3.7.18), (3.7.19) è

$$\begin{cases} m_0 = \frac{1}{2} \left(\sqrt{C_0 + 2C_1} + \sqrt{C_0 - 2C_1} \right) \\ m_1 = \frac{1}{2} \left(\sqrt{C_0 + 2C_1} - \sqrt{C_0 - 2C_1} \right) . \end{cases} \quad (3.7.21)$$

Infine, osservato che dati X_1, \dots, X_N la predizione di X_{N+2} è $\widehat{X}_{N+2} = 0$ con e.q.m

$$\mathcal{E}^2 = m_0^2 + m_1^2 = C_0 .$$

Cerchiamo invece \widehat{X}_{N+1} ; poiché

$$X_{N+1} = m_0 \varepsilon_{N+1} + m_1 \varepsilon_N$$

è chiaro che

$$\widehat{X}_{N+1} = m_1 \widehat{\varepsilon}_N \quad (3.7.22)$$

con $\widehat{\varepsilon}_N$ la proiezione di ε_N su $H_{1,N}$.

Ora notiamo che

$$\varepsilon_N = A(\mathcal{B})X_N$$

dove

$$A(z) = \frac{1}{M(z)} = \frac{1}{m_0 \left(1 + \frac{m_1}{m_0} z \right)} .$$

Poiché, posto $\gamma = \frac{m_1}{m_0}$, deve essere $|\gamma| < 1$, risulta anche

$$A(z) = \frac{1}{m_0} \sum_{k=0}^{+\infty} (-\gamma)^k z^k ,$$

cioè

$$\varepsilon_N = \frac{1}{m_0} \left[\sum_{k=0}^{N-1} (-\gamma)^k X_{N-k} + \sum_{k=N}^{+\infty} (-\gamma)^k X_{N-k} \right] . \quad (3.7.23)$$

Consideriamo il secondo termine in parentesi quadra nella (3.7.23); si ha, ricordando che $\|X_n\|^2 = C_0$,

$$\left\| \sum_{k=N}^{+\infty} (-\gamma)^k X_{N-k} \right\| \leq \sum_{k=N}^{+\infty} |\gamma|^k \|X_{N-k}\| = \sqrt{C_0} \frac{|\gamma|^N}{1-|\gamma|},$$

così che se $|\gamma|$ non è troppo vicino ad 1 ed N è abbastanza grande, tale termine è trascurabile e possiamo porre

$$\varepsilon_N \cong \frac{1}{m_0} \sum_{k=0}^{N-1} (-\gamma)^k X_{N-k} \cong \widehat{\varepsilon}_N \in H_{1,N}. \quad (3.7.24)$$

La (3.7.24) usata nella (3.7.22) fornisce il predittore ricercato \widehat{X}_{N+1} ; il suo e.q.m. di stima inoltre sarà

$$\mathcal{E}_1^2 = m_0^2. \quad (3.7.25)$$

Osservazione 3.7.2: si noti che, partendo da un rumore bianco $\{\eta_n\}$ è sempre possibile definire un processo $\{X_n\}$ mediante un operatore di media mobile

$$X_n = m_0 \eta_n + m_1 \eta_{n-1};$$

qualsiasi siano i valori di m_0, m_1 , $\{X_n\}$ risulta stazionario, con una covarianza che soddisfa la (3.7.17). In particolare ciò è vero anche quando $|m_1| > |m_0|$ così che il polinomio caratteristico $m_0 + m_1 z$ ammette una radice interna al cerchio unitario. Ciò sembrerebbe in contraddizione con la teoria svolta nel §3.5 in cui si garantiva l'esistenza e l'unicità di $M(z)$, che per di più non si annullasse nel cerchio unitario. Tale contraddizione però non esiste in quanto effettivamente esiste un solo $M(z)$ con quelle caratteristiche; in specifico se $C_0 = m_0^2 + m_1^2$, $C_1 = m_0 m_1$, con $|m_1| > |m_0|$ è chiaro che $m'_0 = m_1, m'_1 = m_0$ sono soluzioni delle stesse equazioni e per di più soddisfano correttamente la condizione (3.7.18). Questa osservazione ha carattere generale per tutti i processi stazionari di media mobile.

Esempio 3.7.2: (MA2). In questo caso

$$M(z) = m_0 + m_1 z + m_2 z^2$$

e l'equazione di evoluzione del processo è

$$X_n = m_0 \varepsilon_n + m_1 \varepsilon_{n-1} + m_2 \varepsilon_{n-2} .$$

Per le (3.7.11) le relazioni tra m_0, m_1, m_2 e C_0, C_1, C_2 sono

$$\begin{cases} C_0 = m_0^2 + m_1^2 + m_2^2 \\ C_1 = m_0 m_1 + m_1 m_2 \\ C_2 = m_0 m_2 \end{cases} \quad (3.7.26)$$

mentre lo spettro è dato da

$$f(p) = C_0 + 2C_1 \cos 2\pi p + 2C_2 \cos 4\pi p .$$

In primo luogo cerchiamo di risolvere la (3.7.26); conviene porre

$$\begin{vmatrix} m_0 \\ m_1 \\ m_2 \end{vmatrix} = \sqrt{C_0} \begin{vmatrix} \cos \vartheta \cos \lambda \\ \sin \vartheta \\ \cos \vartheta \sin \lambda \end{vmatrix}$$

così che la prima delle (3.7.26) è automaticamente soddisfatta mentre dalle altre due, posto $R_1 = C_1/C_0$, $R_2 = C_2/C_0$, si ricava

$$\begin{cases} \cos \vartheta \sin \vartheta (\cos \lambda + \sin \lambda) = R_1 \\ \cos^2 \vartheta \cos \lambda \sin \lambda = R_2 \end{cases}$$

Quadrando la prima e sostituendo $\cos^2 \vartheta$ della seconda, posto $x = \cos \lambda \sin \lambda$ si riduce il sistema all'equazione

$$(R_1^2 - 2R_2)x^2 - R_2(1 - 2R_2)x + R_2^2 = 0 . \quad (3.7.27)$$

Notando che $2x = \sin 2\lambda$, l'equazione (3.7.27) va risolta col vincolo

$$|x| \leq 1/2 ;$$

sostituendo all'indietro si trova

$$\cos^2 \vartheta = \frac{R_2}{x} \quad (3.7.28)$$

il che dimostra che deve anche essere

$$0 < \frac{R_2}{x} \leq 1 ;$$

se tale condizione è soddisfatta possiamo ricavare ϑ dalla (3.7.28). Ulteriori condizioni sulle soluzioni (m_0, m_1, m_2) sono imposte dalla richiesta che il polinomio $P_2(z) = m_0 + m_1z + m_2z^2$ non abbia radici nel cerchio unitario, ciò che implica necessariamente (ma la condizione non è sufficiente) che

$$|m_0| > |m_2| \quad |m_1| > 2|m_2| .$$

Quanto al problema della predizione, come già nel caso del processo MA1, notiamo che hanno senso solo $(\widehat{X}_{N+1})_{1,N}, (\widehat{X}_{N+2})_{1,N}$ in quanto

$$(\widehat{X}_{N+j})_{1,N} = 0 \quad j > 2 .$$

In particolare si ha

$$\begin{cases} (\widehat{X}_{N+1})_{1,N} = m_1(\widehat{\varepsilon}_N)_{1,N} + m_2(\widehat{\varepsilon}_{N-1})_{1,N} \\ (\widehat{X}_{N+2})_{1,N} = m_2(\widehat{\varepsilon}_N)_{1,N} . \end{cases} \quad (3.7.29)$$

Vediamo che, come già nel caso MA1, si può spesso porre

$$(\widehat{\varepsilon}_{N-1})_{1,N} \cong \varepsilon_{N-1} \quad (\widehat{\varepsilon}_N)_{1,N} \cong \varepsilon_N \quad (3.7.30)$$

il che ci dà come valore approssimato dell'e.q.m. di stima per le (3.7.29) le espressioni

$$\begin{aligned} \mathcal{E}_1^2 &\cong m_0^2 \\ \mathcal{E}_2^2 &\cong m_0^2 + m_1^2 , \end{aligned}$$

mentre $\forall j > 2$ risulta

$$\mathcal{E}_j^2 = m_0^2 + m_1^2 + m_2^2 = C_0 .$$

Per verificare le (3.7.30) esprimiamo ε_N in funzione di $\{X_n, n \leq N\}$. Osservando che

$$\varepsilon_N = M(B)^{-1}X_N = \frac{1}{m_0}(I - \gamma_1 B)^{-1}(I - \gamma_2 B)^{-1}X_N ,$$

dove γ_1, γ_2 sono gli inversi delle radici di $M(z)$, così che

$$|\gamma_1| \leq |\gamma_2| \leq \rho < 1, \quad (3.7.31)$$

usiamo la consueta serie di Neuman

$$(I - \gamma B)^{-1} = \sum_{k=0}^{+\infty} \gamma^k B^k$$

per ottenere

$$\begin{aligned} \varepsilon_N &= \frac{1}{m_0} \sum_{k,h=0}^{+\infty} \gamma_1^k \gamma_2^h B^{k+h} X_N = \\ &= \frac{1}{m_0} \sum_{\ell=0}^{+\infty} \left(\sum_{k=0}^{\ell} \gamma_1^k \gamma_2^{\ell-k} \right) X_{N-\ell}. \end{aligned}$$

Per la condizione (3.7.31) vale la maggiorazione

$$\left| \sum_{k=0}^{\ell} \gamma_1^k \gamma_2^{\ell-k} \right| \leq \ell \rho^{\ell},$$

così che

$$\left\| \sum_{\ell=N}^{+\infty} \left(\sum_{k=0}^{\ell} \gamma_1^k \gamma_2^{\ell-k} \right) X_{N-\ell} \right\| \leq \sqrt{C_0} \cdot \frac{\rho^N}{(1-\rho)^2} [N(1-\rho) + \rho], \quad (3.7.32)$$

che è una quantità certamente piccola se ρ non è troppo vicino ad 1 ed N è abbastanza grande. Quando la (3.7.32) sia trascurabile, vale la (3.7.30) e si potrà porre

$$\begin{aligned} (\widehat{\varepsilon}_{N-1})_{1,N} &\cong \varepsilon_{N-1} \cong \frac{1}{m_0} \sum_{\ell=0}^{N-2} \left(\sum_{k=0}^{\ell} \gamma_1^k \gamma_2^{\ell-k} \right) X_{N-1-\ell} \\ (\widehat{\varepsilon}_N)_{1,N} &\cong \varepsilon_N \cong \frac{1}{m_0} \sum_{\ell=0}^{N-1} \left(\sum_{k=0}^{\ell} \gamma_1^k \gamma_2^{\ell-k} \right) X_{N-\ell}. \end{aligned}$$

b) Processi autoregressivi

In questo caso si ha per definizione

$$M(z) = \frac{p_0}{Q_h(z)} \quad (q_0 = 1) \quad (3.7.33)$$

$$f(p) = \frac{p_0^2}{\left| \sum_{k=0}^h q_k e^{i2\pi k p} \right|^2} ; \quad (3.7.34)$$

la relazione tra m_k e q_k in questo non è affatto semplice e così la relazione tra m_k e C_k .

Considerando invece che

$$A(z) = M^{-1}(z) ,$$

è elementare la relazione tra a_k e q_k , cioè

$$a_k = \frac{q_k}{p_0} , \quad k = 0, 1, \dots, h .$$

Poiché il processo \underline{X} soddisfa l'equazione

$$A(B)\underline{X} = \varepsilon ,$$

si ha

$$a_0 X_n + a_1 X_{n-1} \dots + a_h X_{n-h} = \varepsilon_n , \quad (3.7.35)$$

da cui il nome per \underline{X} di processo autoregressivo di ordine h .

Cerchiamo la relazione generale che lega gli a_k (ovvero p_0 e i q_k) alla covarianza C_k . In primo luogo osserviamo che essendo

$$A \cdot M = I ,$$

con A e M entrambe triangolari basse, si deve avere pure

$$a_0 m_0 = 1 , \quad m_0 = \frac{1}{a_0} = p_0 ; \quad (3.7.36)$$

Queste equazioni possono essere usate in duplice modo: o per ricavare la funzione di covarianza $\{C_k\}$, da $\underline{a}^+ = [a_0, a_1 \dots a_h]$ o viceversa per ricavare \underline{a} da $\{C_k\}$. Nel primo caso si risolve (3.7.39) rispetto a $[C_0, C_1 \dots C_h]$ e poi si imbecca la relazione ricorsiva

$$C_k = -\frac{a_1}{a_0}C_{k-1} \dots - \frac{a_h}{a_0}C_{k-h}, \quad (k > h),$$

che pure deriva dalla (3.7.37).

Nel secondo caso, chiamando ancora con C_h la matrice di Toeplitz e con \underline{e}_0 il vettore

$$C_h = \begin{vmatrix} C_0 & C_1 & \dots & C_h \\ C_1 & \ddots & \ddots & C_1 \\ C_h & C_1 & \ddots & C_0 \end{vmatrix}, \quad \underline{e}_0 = \begin{vmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{vmatrix}$$

poniamo

$$\tilde{\underline{a}} = C_h^{-1} \underline{e}_0$$

da cui si vede che il vettore \underline{a} soluzione della (3.7.39) è dato da

$$\underline{a} = p_0 \tilde{\underline{a}}. \quad (3.7.40)$$

Ora notiamo che

$$\tilde{a}_0 = \underline{e}_0^+ \tilde{\underline{a}} = \underline{e}_0^+ C_h^{-1} \underline{e}_0 > 0$$

poiché C_h è definita positiva, così che dalla (3.7.40), ricordando la (3.7.38), troviamo

$$a_0 = \underline{e}_0^+ \underline{a} = p_0 \tilde{a}_0$$

cioè

$$a_0 = \sqrt{\tilde{a}_0}, \quad p_0 = 1/\sqrt{\tilde{a}_0}.$$

Con ciò dalla (3.7.40) si trova la soluzione completa.

Occupiamoci ora del problema della predizione che è conveniente risolvere a passi.

A passo 1 avremo

$$X_{N+1} = -p_0 a_1 X_N \dots - p_0 a_h X_{N+1-h} + p_0 \varepsilon_{N+1} ;$$

se $h \leq N$ è chiaro che $X_N, \dots, X_{N+1-h} \in H_{1,N}$ mentre $\varepsilon_{N+1} \perp H_{1,N}$ così che si ha banalmente

$$\begin{cases} (\widehat{X}_{N+1})_{1,N} = -p_0 a_1 X_N \dots - p_0 a_h X_{N+1-h} \\ \mathcal{E}_1^2 = p_0^2 , \end{cases} \quad (3.7.41)$$

che risolve il nostro problema.

Vediamo come agire a passo 2; in primo luogo abbiamo

$$X_{N+2} = -p_0 a_1 X_{N+1} - p_0 a_2 X_N \dots - p_0 a_h X_{N+2-h} + p_0 \varepsilon_{N+2}$$

da cui immediatamente ricaviamo

$$(\widehat{X}_{N+2})_{1,N} = -p_0 a_1 (\widehat{X}_{N+1})_{1,N} - p_0 a_2 X_N \dots - p_0 a_h X_{N+2-h} \quad (3.7.42)$$

dove $(\widehat{X}_{N+1})_{1,N}$ è dato dalle (3.7.41).

In questo caso l'errore di predizione è

$$\begin{aligned} X_{N+2} - (\widehat{X}_{N+2}) &= -p_0 a_1 [X_{N+1} - (\widehat{X}_{N+1})_{1,N}] + p_0 \varepsilon_{N+2} = \\ &= -p_0 a_1 [p_0 \varepsilon_{N+1}] + p_0 \varepsilon_{N+2} \end{aligned}$$

così che l'e.q.m. di predizione a passo 2 è

$$\mathcal{E}_2^2 = p_0^4 a_1^2 + p_0^2 . \quad (3.7.43)$$

Procedendo in modo analogo si ricavano iterativamente predizioni ed e.q.m. di predizione per ogni passo.

Esempio 3.7.3: (AR1). In questo caso abbiamo

$$Q(z) = 1 + q_1 z ,$$

così che l'equazione evolutiva del processo è

$$a_0 X_n + a_1 X_{n-1} = \varepsilon_n , \quad (a_0 = p_0^{-1}, a_1 = p_0^{-1} q_1) ,$$

ovvero

$$X_n = q_1 X_{n-1} + p_0 \varepsilon_n$$

Notiamo che se deve essere $|Q(z)| > 0$ in $\{|z| \leq 1\}$ si deve avere

$$|q_1| = \left| \frac{a_1}{a_0} \right| < 1 .$$

Le (3.7.39) per un processo AR1 sono

$$\begin{cases} a_0 C_0 + a_1 C_1 = p_0 \\ a_0 C_1 + a_1 C_0 = 0 \end{cases} \quad (3.7.44)$$

e, per $k > 1$,

$$a_0 C_k + a_1 C_{k-1} = 0 . \quad (3.7.45)$$

Supposto di conoscere a_0, a_1 dalle (3.7.44) troviamo

$$\begin{cases} C_0 = \frac{p_0^2}{1 - q_1^2} \\ C_1 = (-q_1) C_0 \end{cases}$$

e così dalle (3.7.45) si ricava

$$C_k = (-q_1)^k C_0 , \quad (3.7.46)$$

che fornisce la forma generale della covarianza di un AR1.

Si può notare che se $q_1 > 0$, C_k è oscillante mentre se $q_1 < 0$, C_k è monotona; in ogni caso essendo $|C_k| = |q_1|^k C_0$ si vede che la covarianza deve decadere esponenzialmente a zero. Inoltre lo spettro del processo è dato da

$$f(p) = \frac{1}{|a_0 + a_1 e^{o21pip}|^2} = \frac{1}{1 + q_1^2 + 2q_1 \cos 2\pi p} ,$$

che come si vede è sempre positivo essendo $|q_1| < 1$.

Viceversa le (3.7.44) possono essere risolte per a_0, a_1

$$\begin{aligned} a_0 &= \sqrt{\frac{C_0}{C_0^2 - C_1^2}}, \quad (p_0 = a_0^{-1}), \\ a_1 &= -\frac{C_1}{C_0} a_0. \end{aligned}$$

Infine il procedimento di predizione visto alle (3.7.41), (3.7.42), (3.7.43), per un AR1 dà

$$\begin{cases} (\widehat{X}_{N+1})_{1,N} = -q_1 X_N \\ \mathcal{E}_1^2 = p_0^2 \end{cases}$$

$$\begin{cases} (\widehat{X}_{N+2})_{1,N} = q_1^2 X_N \\ \mathcal{E}_2^2 = p_0^2 + p_0^4 a^2 - 1 = p_0^2(1 + q_1^2) \end{cases}$$

ed è facile vedere che più in generale

$$\begin{cases} (\widehat{X}_{N+j})_{1,N} = (-q_1)^j X_N \\ \mathcal{E}_j^2 = p_0^2[1 + q_1^2 + \dots + q_1^{2(j-1)}]. \end{cases}$$

Esempio 3.7.4: (AR2). In questo caso si ha

$$Q(z) = 1 + q_1 z + q_2 z^2$$

e l'equazione d'evoluzione del processo è

$$a_0 X_n + a_1 X_{n-1} + a_2 X_{n-2} = \varepsilon_n$$

ovvero

$$X_n = -q_1 X_{n-1} - q_2 X_{n-2} + p_0 \varepsilon_n.$$

Le equazioni di Yule-Walker sono

$$\begin{cases} a_0 C_0 + a_1 C_1 + a_2 C_2 = p_0 \\ a_0 C_1 + a_1 C_0 + a_2 C_1 = 0 \\ a_0 C_2 + a_1 C_1 + a_2 C_0 = 0, \end{cases}$$

a cui si può aggiungere l'equazione ricorsiva

$$a_0 C_k + a_1 C_{k-1} + a_2 C_{k-2} = 0 \quad (k > 2) \quad (3.7.47)$$

ovvero

$$C_k = -q_1 C_{k-1} - q_2 C_{k-2} \quad (k > 2)$$

La soluzione delle equazioni di Yule-Walker rispetto a C_k è data da

$$\left\{ \begin{array}{l} C_0 = \frac{1}{D} \frac{a_0 + a_2}{a_0 - a_2} \\ C_1 = -\frac{1}{D} \frac{a_1}{a_0 - a_2} \\ C_2 = \frac{1}{D} \frac{a_1^2 - a_2^2 - a_0 a_2}{a_0(a_0 - a_2)} \\ (D = (a_0 + a_2)^2 - a_1^2) . \end{array} \right. \quad (3.7.48)$$

Si osservi che le (3.7.48) impongono dei limiti sui valori di C_0, C_1, C_2 , in quanto deve necessariamente essere $a_0 \neq a_2$ e $|a_0 + a_2| \neq |a_1|$. La forma generale della funzione di covarianza C_k va ricavata risolvendo la (3.7.47), ciò che richiede una certa discussione.

Intanto consideriamo la decomposizione del polinomio caratteristico

$$Q(z) = (1 - \gamma_1 z)(1 - \gamma_2 z) ;$$

come si vede γ_1, γ_2 sono gli inversi degli zeri del polinomio e quindi per ipotesi devono soddisfare la disuguaglianza $|\gamma_1|, |\gamma_2| < 1$.

Notiamo intanto che se $\gamma_1 \neq \gamma_2$, le funzioni di $k, \gamma_1^k, \gamma_2^k$ sono soluzioni particolari della (3.7.47) tra loro linearmente indipendenti; pertanto la funzione

$$C_k = \lambda_1 \gamma_1^k + \lambda_2 \gamma_2^k, \quad (k > 1) \quad (3.7.49)$$

fornisce una soluzione generale della (3.7.47) dipendente da due costanti arbitrarie che possono a loro volta essere calcolate imponendo che la (3.7.49)

coincida con C_0, C_1 date dalla (3.7.48)

$$\begin{cases} \lambda_1 = \frac{\gamma_2 C_0 - C_1}{\gamma_2 - \gamma_1} \\ \lambda_2 = \frac{\gamma_1 C_0 - C_1}{\gamma_1 - \gamma_2} . \end{cases}$$

Si osservi che quando γ_1, γ_2 sono complesse devono anche essere coniugate tra loro $\gamma_2 = \gamma_1^*$, così che $\gamma_2 - \gamma_1$ è immaginaria pura; pertanto in tal caso risulta pure $\lambda_2 = \lambda_1^*$ che garantisce la realtà di C_k . Quando invece sia $\gamma_1, \gamma_2 = \gamma$, che è quindi per forza un numero reale, tale che $|\gamma| < 1$, si ha

$$C_k = \lambda_1 \gamma^k + \lambda_2 k \gamma_k \quad (k > 1) \quad (3.7.50)$$

con $\lambda_1 \lambda_2$ date da

$$\begin{cases} \lambda_1 = C_0 \\ \lambda_2 = C_1 - \frac{1}{\gamma} C_0 . \end{cases}$$

Per completare la discussione notiamo infine che se γ_1, γ_2 sono reali la (3.7.49) dà una combinazione di due esponenziali, possibilmente alternanti con k quando γ_1 o γ_2 siano negativi,

$$\gamma_1^k = (\pm)^k e^{-\alpha_1 k}, \gamma_2^k = (\pm)^k e^{-\alpha_2 k} \quad (\alpha_1, \alpha_2 > 0) .$$

Quando invece sia $\gamma_2 = \gamma_1^* = e^{\alpha \pm i\beta}$, si può porre la (3.7.49) nella forma

$$C_k = Re^{-\alpha n} \cos(\beta n + a) ,$$

cioè di una esponenziale-coseno. Come sempre le equazioni di Yule-Walker possono essere risolte rispetto ad \underline{a} e ricordiamo che si può porre (cfr. (3.7.40) e seguenti)

$$\underline{a} = p_0 \tilde{\underline{a}}, \quad p_0 = \frac{1}{\sqrt{\tilde{a}_0}} ;$$

nel nostro caso $\tilde{\underline{a}}$ può essere esplicitamente scritto come

$$\begin{vmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \tilde{a}_2 \end{vmatrix} = D \begin{vmatrix} 2(C_0^2 - C_1^2) \\ C_1(3C_0^2 - C_1) \\ 2(C_0 C_2 - C_1^2) \end{vmatrix}$$

$$D = \frac{1}{2(C_0 - C_2)(C_0^2 + C_0C_2 - 2C_1^2)} .$$

La densità spettrale del processo AR2 può essere posta nella forma

$$f(P) = \frac{p_0^2}{|Q(e^{i2\pi p})|^2} = \frac{p_0^2}{(1 + a_1^2 + a_2^2) + 2(q_1 + q_2) \cos 2\pi p + 2q_2 \cos 4\pi p} .$$

Quanto alla predizione del processo, usando le (3.7.41), (3.7.42), (3.7.43) si ha

$$\begin{cases} (\hat{X}_{N+1})_{1,N} = -q_1 X_N - q_2 X_{N-1} \\ \mathcal{E}_1^2 = p_0^2 , \\ (\hat{X}_{N+2})_{1,N} = (q_1^2 - q_2) X_N + q_1 q_2 X_{N-1} \\ \mathcal{E}_2^2 = p_0^2 (1 + q_1^2) . \end{cases}$$

In modo analogo si ricavano le predizioni per i passi successivi.

c) I processi autoregressivi a media mobile

Aniché fare una trattazione generale dei processi ARMA ci limitiamo qui a studiare il processo del tipo ARMA(1,1), corrispondente ad $L = h = 1$, essendo peraltro tale esempio facilmente generalizzabile.

Esempio 3.7.5: (ARMA (1,1)). In questo caso si ha

$$M(z) = \frac{P_1(z)}{Q_1(z)} = \frac{p_0 + p_1 z}{1 + q_1 z} . \quad (3.7.51)$$

Ricordiamo subito che dovendo essere $P_1(z) \neq 0, Q_1(z) \neq 0$ nel cerchio $|z| \leq 1$ si hanno le limitazioni

$$p_0 > 0 , |p_0| > |p_1| , |q_1| < 1 . \quad (3.7.52)$$

Potendosi scrivere

$$\underline{X} = M(B)\underline{\varepsilon} = Q_1^{-1}(B)P_1(B)\underline{\varepsilon}$$

ovvero

$$Q_1(B)\underline{X} = P_1(B)\underline{\varepsilon} ,$$

$$Q_1(B)C = \begin{matrix} & & & & j \\ & & & & \vdots \\ & & \ddots & & \vdots \\ j & \left| \begin{array}{ccc} \dots & C_2 + q_1 C_1 & C_1 + q_1 C_0 & C_0 + q_1 C_1 \\ & \ddots & \ddots & \ddots \end{array} \right. \end{matrix}$$

Da queste relazioni scrivendo la (3.7.54) per le componenti (k, j) con $k \leq j$ si ha

$$\begin{cases} C_0 + q_1 C_1 = p_0 m_0 + p_1 m_1 = p_0^2 + p_1^2 - q_1 p_1 p_0 \\ C_1 + q_1 C_0 = p_1 m_0 = p_0 p_1 \\ C_k + q_1 C_{k-1} = 0, \quad (k \geq 2) \end{cases} \quad (3.7.55)$$

Usando le (3.7.55) per ricavare C_k da p_0, p_1, q_1 , si vede intanto che

$$C_k = (-q_1)^{(k-1)} C_1, \quad (k \geq 2)$$

cioè dal passo 2, C_k si comporta come la covarianza di un AR1 decadendo esponenzialmente; inoltre le prime due delle (3.7.55) danno

$$\begin{vmatrix} C_0 \\ C_1 \end{vmatrix} = \frac{1}{1 - q_1^2} \begin{vmatrix} p_0^2 + p_1^2 - 2q_1 p_0 p_1 \\ -q_1(p_0^2 + p_1^2) - (1 - q_1^2)p_0 p_1 \end{vmatrix}.$$

Viceversa volendo usare le (3.7.55) per ricavare p_0, p_1, q_1 da C_0, C_1, C_2 , si può intanto scrivere immediatamente

$$q_1 = -\frac{C_2}{C_1};$$

inoltre usando la seconda nella prima delle (3.7.55) queste equazioni si portano nella forma

$$\begin{cases} p_0^2 + p_1^2 = (1 - q_1^2)C_0 + 2q_1 C_1 \\ p_0 p_1 = C - 1 + q_1 C_0, \end{cases}$$

la cui unica soluzione soddisfacente (3.7.52) è

$$\begin{cases} p_0 = \frac{1}{2} \left[\sqrt{(1 + 2q_1 - q_1^2)C_0 + 2(1 + q_1)C_1} + \sqrt{(1 + 2q_1 - q_1^2)C_0 - 2(1 - q_1)C_1} \right] \\ p_1 = \frac{1}{2} \left[\sqrt{(1 + 2q_1 - q_1^2)C_0 + 2(1 + q_1)C_1} - \sqrt{(1 + 2q_1 - q_1^2)C_0 - 2(1 - q_1)C_1} \right]. \end{cases}$$

La densità spettrale di un processo ARMA 1,1 è ricavata come al solito da

$$f(p) = \frac{|P_1(e^{i2\pi p})|^2}{|Q_1(e^{i2\pi p})|^2} = \frac{p_0^2 + p_1^2 + 2p_0p_1 \cos 2\pi p}{1 + q_1^2 + 2q_1 \cos 2\pi p} ,$$

che, stanti le (3.7.52), è una funzione sempre positiva.

Infine occupiamoci del problema della predizione almeno per due passi a partire da N .

Poiché si ha una parte di media mobile in $M(B)$, è chiaro che anche qui si porrà il problema di stimare $(\widehat{\varepsilon}_N)_{1,N}$. A questo scopo notiamo che posto

$$Y_n = X_n + q_1 X_{n-1} ,$$

risulta $Y_n \in H_{1,N}$, $n = 2, \dots, N$; allora, riscritta la (3.7.53) come

$$\begin{aligned} \varepsilon_n &= a_0 Y_n - \gamma \varepsilon_{n-1} \\ a_0 &= p_0^{-1} \quad \gamma = \frac{p_1}{p_0} \quad (|\gamma| < 1) , \end{aligned}$$

si può porre in via approssimata

$$\varepsilon_N \cong a_0 \sum_{k=0}^{N-2} (-\gamma)^k Y_{N-k} \cong (\widehat{\varepsilon}_N)_{1,N} \in H_{1,N} \quad (3.7.56)$$

considerando così $(\widehat{\varepsilon}_N)_{1,N}$ come nota con errore trascurabile.

Pertanto da

$$X_{N+1} = -q_1 X_N + p_0 \varepsilon_{N+1} + p_1 \varepsilon_N$$

si vede immediatamente che, usando la (3.7.56),

$$\begin{cases} (\widehat{X}_{N+1})_{1,N} = -q_1 X_N + p_1 (\widehat{\varepsilon}_N)_{1,N} \\ \mathcal{E}_1^2 \cong p_0^2 . \end{cases}$$

Inoltre da

$$\begin{aligned} X_{N+2} &= -q_1 X_{N+1} + p_0 \varepsilon_{N+2} + p_1 \varepsilon_{N+1} \cong \\ &\cong -q_1 (\widehat{X}_{N+1})_{1,N} + p_0 \varepsilon_{N+2} + (p_1 - q_1 p_0) \varepsilon_{N+1} , \end{aligned}$$

si vede che

$$\begin{cases} (\widehat{X}_{N+2})_{1,N} = -q_1(\widehat{X}_{N+1})_{1,N} \\ \mathcal{E}_2^2 = p_0^2 + (p_1 - q_1 p_0)^2 ; \end{cases}$$

il procedimento può poi essere continuato per le predizioni ai passi successivi.