

UN' INTRODUZIONE ALL'INFERENZA SU PROCESSI STOCASTICI

Il piu' semplice problema (statico) di inferenza statistica parametrica e' quello in cui la distribuzione di probabilita' comune a tutti i numeri aleatori (n.a.) osservabili $X(t)$, $t \geq 1$, dipende da un parametro non noto Θ di cui si cerca una valutazione numerica approssimata (stima) sulla base dei valori di un numero prefissato n di osservazioni (o misurazioni) $X(t) = x(t)$, $t = 1, 2, \dots, n$. In altri termini, per il parametro incognito Θ si cerca uno stimatore S dipendente dai n.a. $X(1), \dots, X(n)$ che goda di buone proprieta'. Ovviamente queste ultime devono garantire che la probabilita' di rilevanti errori di stima $\Theta - S$ sia opportunamente piccola. E' noto che tale obiettivo viene perseguito in vari modi ai quali corrispondono vari procedimenti o metodi di stima puntuale; i piu' usati sono il "metodo di massima verosimiglianza", il "metodo dei momenti", il "metodo di minima varianza",

Quando il parametro incognito Θ varia con t , cioe' quando esso costituisce un processo stocastico $\{\Theta(t), t \geq 1\}$, il problema inferenziale assume **un carattere dinamico**: per ogni t , si tratta di trovare uno stimatore $S(t) = \xi[X(1), X(2), \dots, X(t)]$ per il n.a. $\Theta(t)$ che abbia buone proprieta' nel senso anzidetto. La forma funzionale $\xi[., \dots, .]$ dello stimatore dipende, ovviamente, dal metodo di stima adottato. Nel problema dinamico di stima che abbiamo ora introdotto compaiono due processi stocastici: l'insieme di variabili osservabili, o **processo di osservazione**, $\{X(t); t \geq 1\}$ e l'insieme dei parametri incogniti, o **processo delle variabili di stato**, $\{\Theta(t); t \geq 1\}$. La maggiore o minore complessita' del problema di stima dipende naturalmente dalle caratteristiche dei due processi stocastici e dal tipo di relazione sussistente tra essi.

Per un semplice esempio di problema inferenziale dinamico si pensi che $X(t)$ rappresenti il numero aleatorio di chiamate telefoniche su una data linea nel periodo t -esimo, con riferimento ad una sequenza di periodi della stessa durata (per esempio oraria). La distribuzione di probabilita' comune a tutte le variabili osservabili $X(t)$ sia di tipo poissoniano $P\{X(t) = n\} = \frac{[\Theta(t)]^n}{n!} e^{-\Theta(t)}$, ove $\Theta(t)$ e' l'intensita' non nota delle chiamate nel t -esimo periodo, cioe' il numero medio di chiamate in quel periodo. Il problema di inferenza statistica in questo caso consiste nella stima delle intensita' $\Theta(t)$, per ogni t , sulla base delle osservazioni dei valori di $X(1), \dots, X(t)$, fissate che siano le caratteristiche dei due processi $\{X(t)\}$ e $\{\Theta(t)\}$.

Da quanto abbiamo detto finora si intuisce la centralita' della nozione di processo stocastico: presenteremo quindi sommariamente tale nozione e descriveremo alcune principali tipologie di processi.

La nozione di processo stocastico

Dal punto di vista matematico, fissato uno spazio di probabilita' (Ω, \mathcal{A}, P) , un processo stocastico deve essere considerato una **funzione di due variabili**, $X(\omega, t)$, $\omega \in \Omega$, $t \in T$, tale che:

1) per ogni fissato valore t' di t , in T , $X(., t')$ e' un numero aleatorio, cioe' una funzione reale, \mathcal{A} -misurabile, definita in Ω , con funzione di ripartizione $F_{t'}(x) = P\{X_{t'}^{-1}(-\infty, x]\}$;

2) per ogni fissato valore ω' di ω , in Ω , $X(\omega', .)$ e' una funzione deterministica definita in T e denominata "traiettoria" o "realizzazione" del processo $X(t)$.

Da qui nascono le due interpretazioni del processo stocastico quale "famiglia di numeri aleatori", secondo la 1), e quale "funzione aleatoria", secondo la 2).

Come per un numero aleatorio, anche per un processo stocastico si possono individuare piu' (in teoria infiniti) **livelli di conoscenza o di specificazione**: quello massimale comporta la conoscenza

della "legge temporale", cioè della totalità delle funzioni di ripartizione congiunte di dimensione finita $F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$, $n \geq 1$, $t_j \in T$ per ogni j .

Per i processi a parametro discreto in cui $T = Z^+$ (cioè per le sequenze illimitate $\{X_t; t \geq 1\}$ di numeri aleatori), tale famiglia di distribuzioni può essere sostituita dalla più semplice successione $F_1(x_1), F_{1,2}(x_1, x_2), F_{1,2,3}(x_1, x_2, x_3), \dots$ che soddisfi la seguente condizione di coerenza:

Teorema di A.N.Kolmogorov: condizione necessaria e sufficiente affinché tale successione di funzioni di ripartizione costituisca la legge temporale di un processo stocastico a parametro discreto è che ogni elemento della successione sia implicato dai successivi come loro distribuzione marginale e che implichi i precedenti come sue distribuzioni marginali.

Un livello di conoscenza inferiore al precedente e molto usato nelle applicazioni è quello che comporta la conoscenza dei soli momenti del primo e secondo ordine dei numeri aleatori X_t , e cioè delle speranze matematiche, delle varianze e delle covarianze. Si definiscono per il processo $\{X_t; t \geq 1\}$ le due funzioni: la **funzione valor medio** $\varphi(t) = E(X_t)$ e la **funzione di covarianza** $\psi(s, t) = \text{Cov}(X_s, X_t)$; quest'ultima soddisfa le seguenti proprietà generali:

- 1) $\psi(t, t) = \text{Var}(X_t) \geq 0, \forall t$;
- 2) $|\psi(s, t)| \leq [\psi(s, s) \cdot \psi(t, t)]^{1/2}$;
- 3) $\psi(s, t) = \psi(t, s)$;
- 4) $\sum_{s, t=1}^N a_s \cdot a_t \cdot \psi(s, t) \geq 0; \forall N, \mathbf{a} \in R^N$.

Le principali categorie di processi stocastici sono i processi **markoviani**, quelli **stazionari** ed i processi **martingala**; i fondamenti teorici di tali processi sono stati posti negli anni venti e trenta del secolo ventesimo. I contributi principali sono dovuti a P.Lévy, A.N.Kolmogorov, A.Y.Khintchine, W.Feller, B. de Finetti, H.Cramèr, H.Wold, J.L.Doob e altri.

La nostra esposizione riguarderà soprattutto i processi a parametro discreto $\{X_t; t \in Z^+\}$ anche se, occasionalmente, parleremo di alcuni processi a parametro continuo $\{X(t), t \in R\}$; a questo punto introduciamo sinteticamente le definizioni dei tre tipi di processi suddetti quando il parametro operativo t assume valori in Z^+ :

α) il processo $\{X_t; t \in Z^+\}$ è **markoviano** se per ogni scelta dell'intero n e della sequenza

$$t_1 < t_2 < \dots < t_n \text{ risulta } F(x_{t_n} / x_{t_1}, \dots, x_{t_{n-1}}) = F(x_{t_n} / x_{t_{n-1}});$$

β) il processo $\{X_t; t \in Z^+\}$ è **stazionario** se per ogni scelta dell'intero n e della sequenza

$$(t_1, t_2, \dots, t_n) \text{ risulta } F(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = F(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_n+h}), \text{ ove } h \text{ è un qualunque intero tale che ogni } t_j + h \in Z^+;$$

γ) il processo $\{X_t; t \in Z^+\}$ è una **martingala** se per ogni intero positivo $t \geq 1$ risulta $E[|X_t|] < \infty$ e $E(X_t / X_1, \dots, X_{t-1}) = X_{t-1}$.

Processi stocastici del secondo ordine

Cominciando con i processi stocastici a parametro discreto $\{Y_t; t = 0, 1, 2, \dots\}$ diremo che $\{Y_t\}$ è una **sequenza aleatoria del secondo ordine** se i numeri aleatori che la costituiscono hanno momenti secondi finiti; formalmente: $E(Y_t^2) < \infty, \forall t$.

La totalità dei n.a. di questo tipo definiti in uno spazio di probabilità (Ω, A, P) si indica con $H = L_2(\Omega, A, P)$ e tale insieme costituisce uno spazio di Hilbert con funzione prodotto – interno $\langle X, Y \rangle = E(XY)$ e norma $\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{E(X^2)}$; si definiscono anche la distanza tra due elementi di H come $d(X, Y) = \|X - Y\|$ e l'ortogonalità, $X \perp Y$, tra due elementi di H che è verificata se e solo se $\langle X, Y \rangle = 0$. Spesso per semplicità si considera anziché H , lo spazio di Hilbert H_0 dei n.a. definiti in (Ω, A, P) , dotati di momento secondo finito e aventi valor medio nullo: in questo caso è $\langle X, Y \rangle = Cov(X, Y)$, $\|X\| = \sqrt{Var(X)}$, $d(X, Y) = \sqrt{Var(X - Y)}$ e infine $X \perp Y \Leftrightarrow Cov(X, Y) = 0$.

Supponiamo che i n.a. della sequenza $\{Y_t; t = 0, 1, 2, \dots\}$ siano elementi di H_0 : diremo che essi convergono in norma al n.a. $Y \in H_0$ se $d(Y_t, Y) = \|Y_t - Y\| \rightarrow 0$ per $t \rightarrow \infty$; poiché è $d(Y_t, Y) = \|Y_t - Y\| = \sqrt{E(Y_t - Y)^2}$ si ha che la convergenza in norma coincide con la convergenza in media quadratica $Y_t \xrightarrow{q.m.} Y$.

Le definizioni suddette si applicano allo stesso modo a processi stocastici a parametro continuo $\{Y(t); t \in T \subseteq R\}$ detti anche “funzioni aleatorie”. Per questi processi esiste un “calcolo differenziale in media quadratica” basato sulle nozioni di convergenza, continuità, derivabilità e integrabilità in media quadratica:

- convergenza ad Y per $t \rightarrow \tau$: $\lim E[Y(t) - Y]^2 = 0$;
- continuità per $t \rightarrow \tau$: $\lim E[Y(t) - Y(\tau)]^2 = 0$;
- derivabilità in t : $\exists Z = Y'(t), \lim E\left\{\frac{1}{h}[Y(t+h) - Y(t)] - Z\right\}^2 = 0$ per $h \rightarrow 0$;
- integrabilità su (a, b) : $\exists Z = \int_a^b Y(t)dt, \lim E\left\{\sum_{i=0}^{n-1} Y(t_i) \cdot [t_{i+1} - t_i] - Z\right\}^2 = 0$

per $\max_i [t_{i+1} - t_i] \rightarrow 0$ in corrispondenza ad una sequenza di partizioni P_n sempre più fini di (a, b) , ove $P_n = (a = t_0 < t_1 < \dots < t_n = b)$.

Ognuna delle suddette proprietà della funzione aleatoria $\{Y(t); t \in T \subseteq R\}$ è ricollegabile a opportune proprietà della funzione di covarianza $\psi_Y(s, t) = Cov(Y_s, Y_t)$ del processo; per esempio, si ha la convergenza ad Y per $t \rightarrow \tau$ se e solo se $\exists \lim \psi_Y(s, t)$ per $s, t \rightarrow \tau$; si ha la continuità per $t \rightarrow \tau$ se e solo se $\psi_Y(s, t)$ è continua in (τ, τ) e così via. Rinviamo il lettore interessato ad approfondire l'argomento, per esempio, al testo di J. Lamperti – Stochastic Processes (Springer – Verlag, 1977) oppure al manuale di M. Loeve – Probability Theory (Springer-Verlag, 1977).

Per qualche semplice esempio in proposito si considerino i seguenti casi ove A_1 e A_2 sono due numeri aleatori: se $Y(t) = A_1 t$ è $Y'(t) = A_1$; se $X(t) = Y^2(t) = A_1^2 t^2$ si ha $X'(t) = 2 A_1^2 t$; se $Z(t) = X'(t) = 2 A_1^2 t$ si ha $\int_0^t Z(s) ds = A_1^2 t^2$; se $W(t) = \sin A_1 t + \cos A_2 t$ si ha $W'(t) = A_1 \cos A_1 t - A_2 \sin A_2 t$. Invitiamo il lettore ad essere cauto sull'evidente analogia dei suddetti risultati con le regole ben note concernenti la derivazione e l'integrazione di funzioni deterministiche; ognuno dei risultati enunciati andrebbe dimostrato sulla base delle definizioni del calcolo in media quadratica.

Si è già detto che i tipi principali di processi stocastici sono quelli stazionari, markoviani e martingale: esistono versioni analoghe di queste condizioni per i processi del secondo ordine:

- si parla di **stazionarietà del secondo ordine** quando la funzione valor medio $\phi_Y(t)$ è costante e quando la funzione di covarianza $\psi_Y(s, t)$ dipende soltanto dalla differenza $s - t$ degli argomenti;
- si parla di **markovianità del secondo ordine** quando per ogni n e ogni n.a. X dipendente dai n.a. $\{Y_t; t = n, n+1, \dots\}$ si ha $E_2(X / Y_n, Y_{n-1}, \dots, Y_1) = E_2(X / Y_n)$, ove $E_2(X / Y)$ è l'approssimatore ottimale dei minimi quadrati per X in termini di Y ;
- si parla di **martingalità del secondo ordine** se sussistono per ogni n le condizioni $E_2(Y_n / Y_1, Y_2, \dots, Y_{n-1}) = Y_{n-1}$. Ovviamente $E_2(Y_n / Y_1, \dots, Y_{n-1})$ indica l'approssimatore ottimale dei minimi quadrati per Y_n in termini di Y_1, \dots, Y_{n-1} .

I processi stocastici del secondo ordine sono spesso impiegati nelle applicazioni in quanto la loro specificazione riguarda soltanto i momenti del primo e secondo ordine, e cioè valori medi, varianze e covarianze, senza chiamare in causa i momenti di ordine superiore al secondo o addirittura le distribuzioni di probabilità congiunte finite-dimensionali. Ad esempio, l'Analisi delle serie temporali (o storiche), disciplina statistica molto usata sia nelle Scienze naturali che in quelle sociali, usa modelli lineari nei quali sia l'input che l'output sono costituiti da processi (a parametro discreto) del secondo ordine. Più precisamente, i modelli matematici usati sono equazioni alle differenze finite lineari e stocastiche $Y_t - \sum_{i=1}^p a_i Y_{t-i} = Z_t + \sum_{j=1}^q b_j Z_{t-j}$ ove il processo input $\{Z_t; t \geq 0\}$ è specificato dalle funzioni valor medio $\phi_Z(t)$ e di covarianza $\psi_Z(s, t)$. Ovviamente la soluzione dell'equazione, e cioè il processo $\{Y_t; t \geq 0\}$, non può avere una specificazione più dettagliata di quella di $\{Z_t; t \geq 0\}$: risulteranno quindi determinate le sole due funzioni $\phi_Y(t)$ e $\psi_Y(s, t)$.

Alcuni esempi.

- 1) Processo di "rumore bianco" (white noise): si tratta di un processo a parametro discreto costituito da n.a. equi (con valor medio zero), aventi una medesima varianza σ_Y^2 e mutuamente non correlati; viene solitamente indicato con $Y_t \approx WN(0, \sigma_Y^2)$.
- 2) Processo di "passeggiata aleatoria" (random walk): si tratta di un processo definito dalla equazione alle differenze $Y_t = Y_{t-1} + \varepsilon_t$; $\varepsilon_t \approx WN(0, \sigma_\varepsilon^2)$; $Y_0 = 0$.

- 3) Processo “autoregressivo del primo ordine” o AR(1): $Y_t = a.Y_{t-1} + \varepsilon_t$; $\varepsilon_t \approx WN(0, \sigma_\varepsilon^2)$;
 $Cov(Y_0, \varepsilon_t) \equiv 0$; $Y_0 \approx (m_0, \sigma_0^2)$.

Cenni sui processi markoviani

La condizione di **dipendenza markoviana** è stata introdotta da A. Markov nel 1906, ma furono A.N.Kolmogorov nel 1931 e, più tardi, W.Feller ad impostare rigorosamente la teoria dei processi stocastici markoviani. Nel seguito distingueremo il caso di numeri aleatori con un insieme **discreto** di valori possibili (detti anche “stati”) da quello di n.a. con un insieme **continuo** di valori e i processi a tempo **discreto** da quelli a tempo **continuo**.

Fissati arbitrariamente n valori $t_1 < t_2 < \dots < t_n$ del parametro operativo la condizione di dipendenza markoviana è espressa dalla $F(x_{t_n} / x_{t_1}, \dots, x_{t_{n-1}}) = F(x_{t_n} / x_{t_{n-1}})$ o dalla corrispondente uguaglianza tra le densità di probabilità se queste esistono. Questa condizione ha la seguente notevole implicazione per le distribuzioni congiunte che supporremo dotate di densità:

$$\begin{aligned} f(x_{t_1}, \dots, x_{t_n}) &= f(x_{t_n} / x_{t_1}, \dots, x_{t_{n-1}}) \cdot f(x_{t_1}, \dots, x_{t_{n-1}}) = f(x_{t_n} / x_{t_{n-1}}) \cdot f(x_{t_1}, \dots, x_{t_{n-1}}) = \\ &= \dots = \\ &= f(x_{t_1}) \cdot \prod_{j=2}^n f(x_{t_j} / x_{t_{j-1}}), \end{aligned}$$

il che significa che le distribuzioni congiunte del processo markoviano sono determinate dalle distribuzioni univariate marginali e dalle distribuzioni univariate condizionali. Una seconda notevole implicazione è la seguente: se si suppone che sia $s < \tau < t$ si ha

$$f(x_t / x_s) = \int_R f(x_t, x_\tau / x_s) dx_\tau = \int_R f(x_t / x_\tau, x_s) \cdot f(x_\tau / x_s) dx_\tau = \int_R f(x_t / x_\tau) \cdot f(x_\tau / x_s) dx_\tau,$$

e tale relazione costituisce una condizione di coerenza per le distribuzioni univariate condizionali, nota in letteratura come condizione di Chapman – Kolmogorov. Le densità condizionate $f(x_t / x_s)$ sono dette “densità di transizione dallo stato x_s allo stato x_t ”.

a) Catene di Markov omogenee con un numero finito di stati.

Il tipo più semplice di processo markoviano è denominato “catena markoviana omogenea” ed è costituito da una successione di n.a. $\{X_n ; n \geq 0\}$, aventi ciascuno lo stesso insieme finito di N valori possibili, caratterizzata per ogni $n \geq 0$ dalla condizione di dipendenza markoviana omogenea:

$$\text{Prob}\{X_{n+1} = j / (X_0 = a) \wedge (X_1 = b) \wedge \dots \wedge (X_n = i)\} = \text{Prob}\{X_{n+1} = j / X_n = i\} = p_{ij}(n),$$

Se le probabilità subordinate $p_{ij}(n)$, dette “di transizione dallo stato i allo stato j”, non dipendono dall’indice n allora si parla di “dipendenza markoviana omogenea”. Si verifica facilmente che in tal caso la struttura probabilistica del processo è univocamente determinata dalla matrice $N \times N$ delle probabilità subordinate $P = [p_{ij}]$ e dalla distribuzione del n.a. X_0 , detta “distribuzione iniziale”. Ovviamente, la somma degli elementi di ogni riga di P è uguale a 1 (matrice “stocastica”).

Indicata con il vettore riga $\mathbf{a}(0)$ la distribuzione di X_0 , quella $\mathbf{a}(1)$ del n.a. X_1 è data dal prodotto $\mathbf{a}(1) = \mathbf{a}(0) \cdot P$ e, in generale, quella di X_n da $\mathbf{a}(n) = \mathbf{a}(0) \cdot P^n = \mathbf{a}(n-1) \cdot P$. L'elemento generico della matrice P^n , $p_{ij}^{(n)}$, è detto "probabilità subordinata di transizione da i a j in n passi" e soddisfa, come si prova facilmente, la condizione di Chapman – Kolmogorov:

$$p_{ij}^{(n)} = \sum_{h=1}^N p_{ih}^{(n-m)} \cdot p_{hj}^{(m)},$$

per ogni intero positivo $m < n$.

Una catena markoviana è detta "regolare" se esiste un intero n_0 tale che per ogni $n > n_0$ tutti gli elementi di P^n risultano positivi; per le catene regolari sussiste il seguente importante risultato:

Teorema di Markov: la matrice P^n converge, al divergere di n , ad una matrice U con elementi positivi, cioè risulta $p_{ij}^{(n)} \rightarrow u_j > 0$ per ogni i ; le righe di U quindi sono tutte uguali e la somma degli elementi di riga è unitaria. Indicata con \mathbf{u} la generica riga di U , è $\mathbf{u} \cdot P = \mathbf{u}$, per cui \mathbf{u} è detta distribuzione stazionaria.

Alcuni esempi.

Si consideri innanzitutto il seguente schema generale di estrazioni ripetute da un'urna contenente inizialmente b palline bianche ed r palline rosse: si sceglie a caso una pallina dall'urna e, assieme ad essa, si immettono nell'urna c palline dello stesso colore di quella estratta e d palline di colore opposto; dopo la prima estrazione nell'urna ci sono quindi $b+r+c+d$ palline. Si effettua una seconda estrazione casuale con la stessa procedura e così via: in generale la composizione dell'urna varia colpo per colpo. Alcuni modelli particolari sono i seguenti:

- a) $c = d = 0$ (estrazioni ripetute con reimbussolamento della pallina estratta);
- b) $c = -1, d = 0$ (estrazioni ripetute senza reimbussolamento);
- c) $c > 0, d = 0$ (urna di Pòlya – modelli di contagio positivo);
- d) $c = -1, d = 1$ (urna di Ehrenfest).

L'ultimo modello realizza una catena di Markov nel senso che le variabili aleatorie X_n che contano il numero di palline bianche nell'urna dopo le prime n estrazioni costituiscono una catena di Markov caratterizzata dalle probabilità subordinate di transizione $p_{i,i+1} = (b+r-i) / (b+r)$, $p_{i,i-1} = i / (b+r)$ e $p_{i,j} = 0$ se j è diverso da $i+1$ e $i-1$. E' facile costruire la corrispondente matrice P se si tiene presente che il numero degli stati è $N = b+r+1$.

Si verifica facilmente che le variabili X_n nel modello dell'urna di Pòlya costituiscono ancora una catena di Markov, mentre invece le variabili Y_n (indicatori di eventi) che assumono i valori 1 e 0 a seconda che dall'urna di Pòlya sia estratta, al colpo n -mo, una pallina bianca o rossa non verificano la condizione di dipendenza markoviana, ma costituiscono un processo stazionario scambiabile. Infine, le variabili Z_n che denotano le proporzioni di palline bianche dopo n estrazioni dall'urna di Pòlya costituiscono un processo martingala.

Un esempio di catena di Markov con un'infinità numerabile di stati (e precisamente l'insieme degli interi non negativi) è costituito dal processo di passeggiata aleatoria $Y_t = Y_{t-1} + |E_t|$, $Y_0 = 0$, ove gli eventi E_t , $t \geq 1$, sono assunti indipendenti e ugualmente probabili con $P(E_t) \equiv p$.

b) Processi a tempo continuo con un insieme discreto di stati $\{s_i\}$.

In questi processi le probabilità di transizione in un intervallo di tempo di ampiezza t dallo stato s_i allo stato s_j , $P[X(t+\tau) = s_j / X(\tau) = s_i]$, sono indicate con $p_{ij}(t)$, per ogni τ , e soddisfano le seguenti condizioni:

$$p_{ij}(t) \geq 0; \sum_j p_{ij}(t) = 1, \forall t; p_{ij}(t+\tau) = \sum_h p_{ih}(t) \cdot p_{hj}(\tau).$$

Se il numero di stati è finito, l'ultima condizione (le relazioni di Chapman – Kolmogorov) può essere espressa dalla $P(t+\tau) = P(t) \cdot P(\tau)$, ove $P(t) = [p_{ij}(t)]$, assumendo che sia $P(0) = I$. Il teorema di Markov stabilisce ora che l'ipotesi $p_{ij}(t) > 0$, per ogni i, j, t , implica l'esistenza di una matrice U tale che $U = \lim_{t \rightarrow \infty} P(t)$ e $U \cdot P(t) = U$. Come nel caso precedente e sempre nell'ipotesi che gli stati siano finiti, indicata con il vettore riga $a(0)$ la distribuzione di $X(0)$ è:

$$a(t) = a(0) \cdot P(t) = a(0) \cdot P(\tau) \cdot P(t-\tau) = a(\tau) \cdot P(t-\tau).$$

Esempio: il processo di Poisson.

L'insieme degli stati è ora l'insieme degli interi non negativi e le variabili del processo, $N(t)$, contano il numero di eventi (di un tipo fissato) che si verificano nell'intervallo di tempo $[0, t]$. Si assume $N(0) = 0$ e che gli incrementi del processo $N(t) - N(s)$ siano stocasticamente indipendenti e omogenei (o stazionari): ciò significa che se è $t_1 < t_2 \leq t_3 < t_4$ gli incrementi $N(t_4) - N(t_3)$ e $N(t_2) - N(t_1)$ sono indipendenti e che $N(t) - N(s)$ e $N(t+\tau) - N(s+\tau)$ sono ugualmente distribuiti qualunque sia τ . Sotto ipotesi non molto restrittive risulta che

$$P\{N(t) = n\} = e^{-\lambda t} \cdot (\lambda t)^n / n!$$

ove $\lambda > 0$ è l'unico parametro della distribuzione; esso è detto "intensità del processo degli arrivi" e rappresenta il numero medio di eventi nell'intervallo di tempo unitario. Si calcola facilmente che è $E[N(t)] = Var[N(t)] = \lambda t$.

Indicato con T_i il tempo di attesa tra l' $(i-1)$ -esimo e l' i -esimo evento, si prova che tali n.a. sono indipendenti e ugualmente distribuiti con densità di probabilità esponenziale negativa $f(t) = \lambda \cdot e^{-\lambda t}$. E' evidente che $\{N(t) = n\} = \{T_n \leq t\} \wedge \{T_{n+1} > t\}$.

Semplici generalizzazioni del processo di Poisson introducono ai **processi di puro ingresso** (facendo dipendere l'intensità λ dallo stato j raggiunto) e ai **processi di ingresso e uscita**, o di nascita e morte, (introducendo oltre alle intensità di ingresso λ_j anche intensità di uscita μ_j). Si veda in proposito, per esempio, il capitolo XVII del primo volume di "An introduction to probability theory and its applications" di W.Feller.

Se si generalizzano ulteriormente i processi markoviani, eliminando l'ipotesi di omogeneità nel tempo, le probabilità di transizione da i a j in un intervallo di tempo (s, t) non dipendono più dalla sola ampiezza dell'intervallo $t-s$, ma da entrambi gli estremi s e t e quindi dovranno essere indicate col simbolo $P_{ij}(s, t)$. Esse verificano le condizioni $P_{ij}(s, t) \geq 0$, $\sum_j P_{ij}(s, t) = 1$ e le relazioni di

Chapman – Kolmogorov: $P_{ij}(s, t) = \sum_h P_{ih}(s, \tau) \cdot P_{hj}(\tau, t)$, $s < \tau < t$.

c) Processi a tempo discreto e spazio degli stati continuo.

Assumeremo che lo spazio degli stati sia R , ma gli sviluppi formali che seguono non varierebbero se esso fosse un arbitrario spazio cartesiano, per esempio R^n .

Nel caso di un processo markoviano a parametro *discreto* le probabilità di transizione sono espresse da una funzione (stochastic kernel) $K(x,S) = \text{Prob} \{X_{n+1} \in S / X_n = x\}$ di due argomenti, $x \in R$ ed $S \subset R$; fissato x , $K(x, \cdot)$ è una probabilità sui sottoinsiemi boreliani di R , mentre fissato S , $K(\cdot, S)$ è una funzione di Baire in R .

Frequentemente nelle applicazioni è $K(x,S) = \int_S k(x,y) dy$ e la funzione $k(x,y)$ (stochastic density kernel) è definita in R^2 . Le relazioni di Chapman – Kolmogorov per le funzioni K e k sono rispettivamente: $K^{(m+n)}(x,S) = \int_R K^{(m)}(x,dy) \cdot K^{(n)}(y,S)$ e $k^{(m+n)}(x,y) = \int_R k^{(m)}(x,z) \cdot k^{(n)}(z,y) dz$.

Ponendo $m = 1$ e facendo crescere n si ottiene una definizione ricorsiva di $K^{(n)}(x,S)$ e $k^{(n)}(x,y)$; per esempio si hanno le $k^{(2)}(x,y) = \int_R k(x,z) \cdot k(z,y) dz$, $k^{(3)}(x,y) = \int_R k(x,z) \cdot k^{(2)}(z,y) dz$ e così via.

Un esempio: sia $k(x,y) = \lambda \cdot y^{\lambda-1} / x^\lambda$, per $0 < y < x$; si ha facilmente

$$k^{(n)}(x,y) = \frac{\lambda^n \cdot y^{\lambda-1} [\log(x/y)]^{n-1}}{\Gamma(n) \cdot x^\lambda}$$

ed inoltre, se $S = [a, b]$ è

$$K(x, [a,b]) = \int_a^b k(x,y) dy = (b^\lambda - a^\lambda) / x^\lambda.$$

d) Processi a tempo continuo e spazio degli stati continuo.

Per un processo markoviano omogeneo a parametro *continuo* denoteremo con $Q_t(x,S)$ le probabilità di transizione

$$Q_t(x,S) = \text{Prob} \{ X(t+\tau) \in S / X(\tau) = x \}, \quad \forall \tau.$$

Esse soddisfano le condizioni $Q_t(x,S) \geq 0$, $Q_t(x,R) = 1$ e le relazioni di Chapman – Kolmogorov :

$$Q_t(x,S) = \int_R Q_{t-\tau}(x,dy) \cdot Q_\tau(y,S), \quad 0 < \tau < t.$$

Un primo esempio molto importante è costituito dal **processo di Poisson composto**. Si tratta di un processo markoviano a parametro continuo molto usato nelle applicazioni la cui interpretazione è tipicamente la seguente: si pensi ad un n.a. $N(t)$ di eventi (o arrivi) osservabili nel periodo di tempo $[0, t]$ a ciascuno dei quali sia associato un “effetto” aleatorio Y_j e interessi conoscere la distribuzione dell’effetto complessivo $S(t) = \sum_{j=0}^{N(t)} Y_j$. Se indichiamo con $\{p_{nt}; n \geq 0\}$ la distribuzione di probabilità di $N(t)$ e con $F_Y(y)$ la funzione di ripartizione (f.d.r.) comune ai n.a. Y_j che assumiamo essere anche mutuamente indipendenti si ha che la distribuzione di $\sum_{j=0}^{N(t)} Y_j$ ha f.d.r. data alla:

$$\text{Prob} \{S(t) \leq x\} = \sum_{n \geq 0} P[N(t) = n] \cdot P[S(t) \leq x / N = n] = \sum_{n \geq 0} p_n \cdot F_Y^{n*}(x),$$

ove il simbolo $F_Y^{n*}(x)$ indica la f.d.r. della distribuzione della somma $\sum_{j=0}^n Y_j$ di n addendi (poiché per ipotesi è $Y_0 = 0$) i.i.d. Il nome di “processo di Poisson composto” è appropriato se la distribuzione $\{p_n; n \geq 0\}$ è poissoniana con parametro λt .

Il carattere markoviano di $\{S(t)\}$ si dimostra facilmente osservando che se $0 < t_1 < \dots < t_n < t$ e posto $N(t+\tau) - N(t) = N_\tau$ si ha:

$$\text{Prob} \{S(t+\tau) = k / [\bigcap_{i=1}^n S(t_i) = j_i] \cap [S(t) = j]\} = \text{Prob} \left\{ \sum_{h=1}^{N_t} Y_h = k - j \right\} = \text{Prob} \{S(t+\tau) = k / S(t) = j\}.$$

Può essere utile riportare l'espressione della funzione caratteristica del processo $S(t)$; si dimostra che è: $\Phi_{S(t)}(\xi) = \exp\{\lambda t [\Phi_Y(\xi) - 1]\}$, ove $\Phi_Y(\xi)$ indica la funzione caratteristica della distribuzione di probabilità dei n.a. Y_j .

Nel seguito, del suddetto processo a parametro continuo $\{S(t)\}$ considereremo gli incrementi relativi ad intervalli unitari di tempo $X_t = S(t) - S(t-1)$, $t = 1, 2, \dots$, definiti dalle $X_t = \sum_{j=0}^{N_t} Y_{tj}$, ove $N_t = N(t) - N(t-1)$ è l'incremento del processo poissoniano di arrivi $\{N(t); t \geq 0\}$ e ove Y_{tj} , che indica l'effetto associato al j -mo arrivo nell'intervallo unitario t -esimo di tempo, è l'elemento generico del “processo degli effetti” $\{Y_j; j \geq 1\}$. Solitamente si assume che i n.a. N_t siano indipendenti dagli Y_j e che la comune distribuzione degli Y_j sia di tipo Gamma; per semplicità noi assumeremo $Y_j \sim \text{Gamma}(1, \beta)$, cioè supporremo che ogni Y_j abbia una densità di probabilità di tipo esponenziale negativo, $f(y) = \beta \cdot \exp(-\beta \cdot y)$, per cui la convoluzione n -ma è una densità di tipo Gamma (n, β) . Si ha allora:

$$F_{X_t}(x) = \sum_{n \geq 0} p_n \cdot F_Y^{n*}(x) = \sum_{n \geq 0} \left(\frac{e^{-\lambda} \cdot \lambda^n}{n!} \right) \int_0^x \frac{\beta^n}{\Gamma(n)} z^{n-1} \cdot e^{-\beta \cdot z} dz, \text{ ed anche } f_{X_t}(x) = \sum_{n \geq 0} \left(\frac{e^{-\lambda} \cdot \lambda^n}{n!} \right) \left(\frac{\beta^n}{\Gamma(n)} x^{n-1} \cdot e^{-\beta \cdot x} \right).$$

Si ricavano, per i primi due momenti di X_t , i risultati:

$$E(X_t) = E(N_t) \cdot E(Y_j) = \lambda / \beta, \quad E(X_t^2) = E(N_t) \cdot \text{Var}(Y_j) + E(N_t^2) \cdot E^2(Y_j) = \lambda \cdot (\lambda + 2) / \beta^2,$$

dai quali si ottiene immediatamente $\text{Var}(X_t) = 2\lambda / \beta^2$.

Ci limitiamo ad accennare che il processo di Poisson composto ha anche una notevole importanza teorica, oltre che applicativa. Si dimostra, per esempio, che la più generale distribuzione “infinitamente divisibile” è rappresentabile come limite di un'appropriata sequenza di distribuzioni

di Poisson composto, $Q_t(x) = \sum_{n \geq 0} \frac{(\alpha \cdot t)^n \cdot \exp(-\alpha \cdot t)}{n!} F^{n*}(x)$. Ricordiamo che una distribuzione di

probabilità $F(x)$ è infinitamente divisibile se, indicata con $\varphi(\xi)$ la corrispondente funzione caratteristica, questa può essere rappresentata, per ogni $n \geq 1$, come potenza n -ma di una funzione caratteristica $\varphi_n(\xi)$, cioè se $\varphi(\xi) = [\varphi_n(\xi)]^n$.

Alcuni esempi di distribuzioni infinitamente divisibili sono costituiti dalle distribuzioni

- di Poisson con $\varphi(\xi) = \exp[\lambda(e^{i\xi} - 1)]$ e $\phi_n(\xi) = \exp[\lambda(e^{i\xi} - 1)/n]$,
- dalla distribuzione normale con $\varphi(\xi) = \exp(i\xi\mu - \sigma^2\xi^2/2)$ e $\phi_n(\xi) = \exp[(i\xi\mu - \sigma^2\xi^2/2)/n]$,
- dalla distribuzione Gamma con $\varphi(\xi) = \left(1 - \frac{i\xi}{\beta}\right)^{-\alpha}$ e $\phi_n(\xi) = \left(1 - \frac{i\xi}{\beta}\right)^{-\alpha/n}$,
- dalla distribuzione di Poisson composto con $\varphi_x(\xi) = \exp\{\lambda[\varphi_y(\xi) - 1]\}$ e $\phi_n(\xi) = \exp\{\lambda[\varphi_y(\xi) - 1]/n\}$.

Un altro esempio importante è costituito dal processo “**pseudo – poissoniano**” $\{X(t); t \geq 0\}$ caratterizzato dalle funzioni di transizione

$$Q_t(x,S) = \sum_{n=0}^{\infty} \left[\frac{e^{-\alpha t} (\alpha t)^n}{n!} \right] K^{(n)}(x,S),$$

ove $K(x,S)$ è una fissata probabilità di transizione di una catena markoviana $\{Z_n; n \geq 0\}$ ed $\{N(t)\}$ è un processo di Poisson con n.a. $N(t)$ indipendenti dai n.a. Z_n ; è allora $X(t) = Z_{N(t)}$.

Se $K(x,S) = \int_S \lambda \cdot y^{\lambda-1} \cdot x^{-\lambda} dy$ si prova che la corrispondente probabilità di transizione $Q_t(x,S)$ ha

densità espressa dalla $q_t(x,y) = \frac{e^{-\alpha t} \cdot \sqrt{\alpha \lambda t} \cdot y^{\lambda-1}}{x^{\lambda} \cdot \sqrt{\log(x/y)}} \cdot I_1(2\sqrt{\alpha \lambda t \cdot \log(x/y)})$, ove $I_1(z) = \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(k+2)} \cdot (z/2)^{2k+1}$ è

la funzione di Bessel di ordine 1, per $0 < y < x$ ed una massa concentrata pari a $e^{-\lambda t}$ in $y = 0$.

Chiaramente, un caso particolare del processo pseudo-poissoniano è costituito dal processo di Poisson composto se le variabili Z_n della catena markoviana sono somme di n.a. Y_j indipendenti e identicamente distribuiti.

Cenni sui processi con incrementi stazionari e indipendenti

I più semplici processi stocastici di questo tipo, a parametro **discreto**, sono quelli denominati

“processi di passeggiata aleatoria” e definiti dalle $S_n = \sum_{k=1}^n X_k$, $n \geq 1$, ove i n.a. X_k sono indipendenti

e ugualmente distribuiti (brevemente i.i.d.); per completezza si pone $S_0 = 0$. Equivalentemente, il processo $\{S_n\}$ può essere definito dalle $S_0 = 0$, $S_n = S_{n-1} + X_n$, $n \geq 1$. Si tratta, ovviamente, di processi markoviani e, più precisamente, di catene markoviane.

Per un primo esempio si assuma $X_k = |E_k|$, con eventi E_k indipendenti e ugualmente probabili (processo bernoulliano); in tal caso S_n è la frequenza di successo su n eventi associata a $\{X_k\}$ e per il processo $\{S_k\}$ si hanno i ben noti risultati

a) Teorema di Bernoulli : $p - \lim [S_n/n] = P(E_1)$;

b) Teorema di de Moivre – Laplace : $\frac{S_n - n \cdot P(E_1)}{\sqrt{n \cdot P(E_1) \cdot [1 - P(E_1)]}} \xrightarrow{d} N(0;1)$.

Se, più in generale, i n.a. X_k , non necessariamente indicatori di eventi, hanno valor medio $E(X_k) = \mu$ e varianza $\text{Var}(X_k) = \sigma^2$ finiti sussistono i seguenti teoremi che generalizzano i risultati precedenti:

c) Legge debole dei grandi numeri: $p - \lim [S_n/n] = \mu$;

d) Teorema centrale limite: $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0;1)$.

Per quanto concerne i momenti fino al secondo ordine del processo $\{S_k\}$ si ha che la funzione valor medio $\varphi_S(n) = E(S_n) = n\mu$ è crescente con n e che la funzione di covarianza è data dalla $\psi_S(m,n) = \text{Cov}(S_m, S_n) = \min(m\sigma^2, n\sigma^2)$.

Gli incrementi del processo $\{S_k\}$ sono dati dai n.a. $S_n - S_{n-1} = X_n$ per $n > 1$ e $S_1 - S_0 = S_1 = X_1$: come si è già assunto, essi sono i.i.d.

Il processo in cui $X_k = |E_k|$ con punto iniziale l'origine e $P(E_k) = p, \forall k$, può descriversi come una catena di Markov con un'infinità numerabile di stati (i numeri interi, positivi nulli e negativi) e matrice di transizione con elementi tutti nulli salvo per $p_{j,j+1} = p$ e $p_{j,j-1} = 1-p, \forall j$.

Un processo a parametro **continuo** $\{S(t); t \in \mathbb{R}\}$ ha incrementi indipendenti e stazionari se, considerata un'arbitraria partizione dell'intervallo $[s, s+t]$, $s = t_0 < t_1 < \dots < t_n = s+t$, gli incrementi $S(t_k) - S(t_{k-1}), k = 1, \dots, n$, sono mutuamente indipendenti e la loro distribuzione dipende soltanto dalle differenze $t_k - t_{k-1}$.

Si supponga ora che gli intervalli $[t_k - t_{k-1}]$ abbiano la medesima lunghezza di modo che gli incrementi $\Delta_k = S(t_k) - S(t_{k-1})$ siano ugualmente distribuiti: dalla $S(s+t) - S(s) = \sum_{k=1}^n \Delta_k$ discende che la distribuzione di $S(s+t) - S(s)$ è uguale alla convoluzione n -ma della comune distribuzione dei Δ_k . Poiché il numero n degli intervalli $[t_k - t_{k-1}]$ è arbitrario, l'incremento $S(s+t) - S(s)$ ha necessariamente una distribuzione "infinitamente divisibile".

Riprendendo quanto già detto, un n.a. Z , la sua funzione di ripartizione $F_Z(z)$ e la sua funzione caratteristica $\Phi(u) = E[\exp(iuZ)] = \int_{\mathbb{R}} e^{iuz} dF_Z(z)$ sono **infinitamente divisibili** se per ogni $n \geq 1$ esistono n.a. $\eta_1, \eta_2, \dots, \eta_n$ i.i.d. tali che $F_Z(z) = F_{\eta_1}^{n*}(z)$, oppure $\Phi_Z(u) = [\Phi_{\eta_1}(u)]^n$. Alcune distribuzioni infinitamente divisibili (inf. div.) sono la normale, la gamma e la distribuzione di Cauchy tra quelle **continue** e la binomiale negativa (e quindi la geometrica), la distribuzione di Poisson e quella di Poisson composto tra quelle **discrete**.

La caratterizzazione delle distribuzioni inf. div. è stata ottenuta nei primi anni 30 attraverso il lavoro di B. de Finetti, A.N. Kolmogorov, P. Lévy e, più tardi, di W. Feller e A.Y. Khintchine.

Di fondamentale importanza è il seguente risultato

Teorema: sia $\{\Phi_n(u); n \geq 1\}$ una successione di funzioni caratteristiche; condizione necessaria e sufficiente affinché esista una funzione limite continua $\Phi(u) = \lim [\Phi_n(u)]^n$ è che esista, continua, la funzione limite $\lim n.[\Phi_n(u) - 1] = \chi(u)$ e in tal caso si ha $\Phi(u) = \exp\{-\chi(u)\}$.

Ricordando che $\exp\{n.[\Phi_n(u) - 1]\}$ è la funzione caratteristica della distribuzione di Poisson composto con intensità n e funzione caratteristica della distribuzione degli effetti $\Phi_n(u)$, il teorema afferma che ogni distribuzione inf. div. è rappresentabile come limite di una opportuna successione di distribuzioni di Poisson composto. Inoltre esso afferma che ogni limite continuo di una successione di funzioni caratteristiche inf. div. è esso stesso inf. div..

Tra le distribuzioni inf. div. sono importanti le **distribuzioni stabili**; tra quelle sopra indicate sono stabili soltanto la normale e quella di Cauchy. Esse sono definite al modo seguente: una

distribuzione F e la sua funzione caratteristica Φ sono stabili se, per ogni $n \geq 1$ esistono costanti $a_n > 0$ e b_n tali che $[\Phi_F(u)]^n = \Phi_{a_n F + b_n}(u) = \exp\{i b_n u\} \cdot \Phi_F(a_n u)$ cioè, a parole, se la distribuzione della somma di n numeri aleatori X_i indipendenti con distribuzione F (la cui funzione caratteristica è $[\Phi_F(u)]^n$) coincide con quella del n.a. $a_n X_1 + b_n$ (con funzione caratteristica $\exp\{i b_n u\} \cdot \Phi_F(a_n u)$).

Si dimostrano le due seguenti proposizioni concernenti le distribuzioni stabili:

- le costanti a_n hanno la forma $a_n = n^{1/\alpha}$, con $0 < \alpha \leq 2$: per la distribuzione normale è $\alpha = 2$ mentre per quella di Cauchy è $\alpha = 1$;
- le distribuzioni stabili **simmetriche** hanno funzione caratteristica avente la forma $\Phi(u) = \exp\{-\gamma_\alpha |u|^\alpha\}$ con γ_α reale positivo: per la distribuzione $N(0; 1)$ è $\Phi(u) = \exp\{-\frac{1}{2}u^2\}$ mentre per quella di Cauchy è $\Phi(u) = \exp\{-\theta |u|\}$, ove è $\theta > 0$.

Presentiamo ora alcuni esempi di processi $\{S(t)\}$ con incrementi indipendenti e stazionari:

- processi di Poisson composto, ove $\text{Prob}\{S(t) \leq s\} = \sum_k \frac{e^{-\lambda t} (\lambda t)^k}{k!} F^{k*}(s)$;
- processi di Poisson: si ricavano dai precedenti specificando che la distribuzione F degli effetti è concentrata in 1; in tal caso è $\Phi_{S(t)}(u) = \exp\{\lambda t [e^{iu} - 1]\}$ in quanto e^{iu} è la funzione caratteristica della distribuzione concentrata in a ;
- processi di Wiener – Lévy (o di moto browniano) caratterizzati dall'ipotesi $S(0) = 0$ e dal fatto che gli incrementi $S(s+t) - S(s) = S(t) - S(0) = S(t)$ hanno distribuzione $N(0; \sigma^2 t)$.

Cenni sui processi stocastici martingala

Un processo stocastico $\{X_t\}$ è detto essere una martingala se per ogni intero positivo $t \geq 1$ si ha $E[|X_t|] < \infty$ ed è verificata (con probabilità 1) la condizione:

$$E(X_t / X_1, \dots, X_{t-1}) = X_{t-1} .$$

E' facile verificare che sussiste, per ogni t , l'uguaglianza $E[X_t] = E[X_{t-1}]$ e che, per s intero positivo, è $\text{Cov}(X_t, X_{t-s}) = \text{Var}(X_{t-s})$. Si ha inoltre $E(X_{t+h} / X_1, \dots, X_{t-1}) = X_{t-1}$ per ogni h intero non negativo, cosicchè ogni sottosequenza $X_{t_1}, X_{t_2}, \dots, X_{t_n}, \dots$ di un processo martingala è una martingala.

Una definizione più flessibile di martingala è la seguente: **il processo $\{X_t\}$ è una martingala rispetto al processo $\{Y_t\}$** se $E[|X_t|] < \infty$ e se $E[X_{t+1} / Y_1, \dots, Y_t] = X_t$ per ogni t . Una definizione ancora più generale di martingala è quella rispetto ad una sequenza crescente di σ -algebre di eventi anzichè rispetto a sequenze di vettori (Y_1, \dots, Y_t) . Noi ci atterremo peraltro alla prima, più semplice, definizione.

Teorema di rappresentazione: il processo $\{X_t\}$ è una martingala se e solo se $X_t = Y_1 + Y_2 + \dots + Y_t + c$ ove c è una costante e i n.a. Y_t sono **assolutamente equi** [$E(Y_1) = 0$ e $E(Y_t / Y_1, \dots, Y_{t-1}) = 0, \forall t$]. In forza di questo risultato, n.a. Y_t assolutamente equi sono detti "differenze, o

incrementi, di martingala". Si dimostra che le differenze di martingala sono n.a. equi e mutuamente non correlati.

Teorema di convergenza: se i momenti secondi dei n.a. X_t del processo martingala sono equilimitati, cioè se esiste un numero reale m per cui è $E[X_t^2] \leq m < \infty$, allora esiste un n.a. X verso cui la martingala converge, con probabilità 1, al divergere di t . Inoltre è $E(X) = E(X_t)$ per ogni t .

Alcuni esempi :

- 1) Le somme successive di n.a. Z_t equi e stocasticamente indipendenti costituiscono una martingala (è lo schema di un gioco equo): infatti i n.a. Z_t sono anche assolutamente equi per cui, ponendo $c = 0$ e $X_t = Z_1 + Z_2 + \dots + Z_t$ si ha che X_t è una martingala per il teorema di rappresentazione.
- 2) Le proporzioni X_t di palline bianche dopo ogni estrazione casuale da un'urna di Polya (contenente inizialmente b palline bianche ed r palline rosse, essendo c le nuove palline immesse nell'urna, dello stesso colore di quella appena estratta) costituiscono una martingala.
- 3) La successione delle funzioni di regressione $X_t = E[Z / Y_1, \dots, Y_t]$ di un n.a. Z rispetto ai n.a. Y_1, \dots, Y_t, \dots è un processo martingala. Occorre, in questo esempio, adottare per i n.a. X_t la definizione di martingala rispetto alla sequenza $Y_1, Y_2, \dots, Y_t, \dots$.

Ritornando all'esempio 1), e considerando che X_t sia una martingala rispetto a $\{Z_t\}$, è opportuno tener conto della possibilità che un giocatore decida colpo per colpo se giocare o no all'epoca n : si può far questo introducendo una "funzione di decisione" $\{\varepsilon_n\}$, ove $\varepsilon_n = 0$ oppure $\varepsilon_n = 1$ a seconda che il giocatore decida di non giocare, o giocare, la partita n -ma. Indicando con W_n il guadagno totale nelle prime n partite si ha $W_n = W_{n-1} + \varepsilon_n \cdot Z_n = W_{n-1} + \varepsilon_n \cdot (X_n - X_{n-1})$.

Osserviamo che in generale $W_{n-1} \neq X_{n-1} = \sum_{t=1}^{n-1} Z_t$ perché il giocatore può decidere di non giocare

qualcuna delle prime $n-1$ partite; se, per esempio, è uguale a 1 solo la prima variabile decisionale ε_1 mentre tutte le successive $n-2$ sono nulle si ha $W_{n-1} = Z_1 = X_1$.

Poiché, a partire dall'equazione sopra scritta, è:

$$E(W_n / Z_1, \dots, Z_{n-1}) = W_{n-1} + \varepsilon_n \cdot [E(X_n / Z_1, \dots, Z_{n-1}) - X_{n-1}],$$

e ricordando che $\{X_n\}$ è una martingala rispetto a $\{Z_n\}$ risulta $E(W_n / Z_1, \dots, Z_{n-1}) = W_{n-1}$ cosicchè anche $\{W_n\}$ è una martingala rispetto a $\{Z_n\}$. Tale risultato è noto come

Teorema di impossibilità (di modificare la struttura di martingala): qualunque funzione di decisione $\{\varepsilon_n\}$ trasforma una martingala $\{X_n\}$ in una nuova martingala $\{W_n\}$.

Una particolare funzione di decisione $\{\varepsilon_t\}$ è quella in cui esiste un intero n tale che $\varepsilon_t = 1$ per ogni $t \leq n$ e $\varepsilon_t = 0$ se $t > n$. Tipicamente, l'intero n è considerato aleatorio (e denotato N) di modo che si ha $\varepsilon_t = 1$ se $N > t-1$ mentre $\varepsilon_t = 0$ se $N \leq t-1$.

Ovviamente, neanche l'introduzione di un **tempo aleatorio di arresto** (stopping time) modifica la proprietà di martingala. Aggiungiamo alcune altre nozioni collegate a quella di martingala.

Un processo $\{X_t\}$ è detto costituire una **sub-martingala** (super-martingala) se soddisfa le condizioni $E(X_t / X_1, \dots, X_{t-1}) \geq X_{t-1}$ ($\leq X_{t-1}$). Sussiste l'importante teorema: "se $u(\cdot)$ è una funzione convessa ed $\{X_t\}$ una martingala allora il processo $\{u(X_t)\}$ è una sub-martingala a condizione che esista finita la $E[u(X_t)]$ per ogni t ".

Un processo $\{X_t\}$ è una **martingala in senso lato** se $E_2[X_t / X_1, \dots, X_{t-1}] = X_{t-1}$ ove il primo membro indica l'approssimazione lineare dei minimi quadrati di X_t rispetto ai n.a. X_1, \dots, X_{t-1} . Si dimostra che un processo $\{X_t\}$ è una martingala in senso lato se e solo se $X_t = Y_1 + \dots + Y_t + c$ ove i n.a. Y_t sono equi e mutuamente ortogonali.

Il processo martingala presentato nel terzo esempio è costituito da **approssimatori dei minimi quadrati** $X_t = E[Z / Y_1, \dots, Y_t]$ del n.a. Z in termini di funzioni arbitrarie dei n.a. Y_t supposti osservabili. La condizione $E(Z^2) < \infty$ implica la $E(X_t^2) \leq m < \infty$ per ogni t e dunque la convergenza in media quadratica di $\{X_t\}$ ad un n.a. X : quest'ultimo dovrebbe potersi interpretare come approssimatore ottimale in quanto utilizza la totalità dei n.a. osservabili Y_t . Inoltre, è naturale pensare che al crescere della numerosità della sequenza Y_1, \dots, Y_t migliori progressivamente l'approssimazione di Z perché si utilizza una quantità di informazione crescente: è infatti quanto viene stabilito dal seguente teorema.

Teorema: posto $D_t = [E(Z - X_t)^2]^{1/2}$, al crescere di t la sequenza $\{D_t\}$ decresce e tende ad un limite D , mentre la sequenza $\{E(X_t^2)\}$ cresce tendendo al limite $E(Z^2) - D^2$. La dimostrazione è semplice se si ricorda che $E[Z / Y_1, \dots, Y_t]$ è interpretabile come proiezione ortogonale di Z sul sottospazio lineare chiuso dell'insieme dei n.a. con momento secondo finito costituito dalle funzioni $\psi(Y_1, \dots, Y_t)$ tali che sia $E[\psi(Y_1, \dots, Y_t)]^2 < \infty$. Intanto risulta $0 \leq D_t \leq [E(Z^2)]^{1/2}$ e poiché la sequenza $\{D_t\}$ decresce al crescere di t esiste un limite D . Inoltre, per la proprietà di ortogonalità di $Z - E[Z / Y_1, \dots, Y_t]$ nei confronti di tutte le funzioni $\psi(Y_1, \dots, Y_t)$ si ha $Z - X_t \perp X_t$ e dunque $E(Z^2) = D_t^2 + E(X_t^2)$. Risulta quindi che $E(X_t^2) \uparrow E(Z^2) - D^2$.

Cenni sui processi stazionari

Un processo stocastico a parametro discreto $\{X_t; t \geq 1\}$ è **stazionario in senso stretto** se le sue distribuzioni congiunte finite-dimensionali soddisfano la seguente condizione di invarianza

$$F(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = F(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_n+h})$$

per ogni intero n , ogni sequenza di interi distinti (t_1, t_2, \dots, t_n) e ogni vettore di numeri reali $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$. Ne discende che i numeri aleatori della sequenza $\{X_t; t \geq 1\}$ sono ugualmente distribuiti, le coppie di n.a. (X_s, X_t) aventi uguale differenza degli indici $s - t$ sono ugualmente distribuite,

Un processo stocastico a parametro discreto $\{X_t; t \geq 1\}$ è **stazionario in senso lato** se i n.a. X_t hanno momenti secondi finiti e se

- 1) la funzione valor medio è costante, cioè se $\varphi_X(t) = E(X_t) \equiv E(X_1)$,
- 2) la funzione di covarianza $\Psi_X(s, t) = Cov(X_s, X_t)$ dipende da s e t solo attraverso $s - t$.

Da quanto detto si ricava che la stazionarietà in senso stretto implica quella in senso lato soltanto se i n.a. X_t hanno tutti momenti secondi finiti; l'implicazione inversa sussiste solo quando le distribuzioni congiunte dipendono dai soli momenti del primo e secondo ordine, come accade per esempio alle distribuzioni di tipo Gaussiano e Student - t .

Un primo esempio di processo stazionario in entrambi i sensi è costituito da una sequenza di n.a. indipendenti e ugualmente distribuiti con varianza finita σ_X^2 . La funzione di covarianza $\Psi_X(s, t)$ ha due valori, $\Psi_X(s, t) = 0$ se $s \neq t$ e $\Psi_X(s, t) \equiv \sigma_X^2$ se $s = t$. I casi più significativi di stazionarietà sono però quelli in cui i n.a. X_t sono mutuamente dipendenti: l'esempio più prossimo al precedente è dato da una sequenza di n.a. ugualmente distribuiti con una funzione di covarianza con due valori, $\Psi_X(s, t) = \gamma \neq 0$ se $s \neq t$ e $\Psi_X(s, t) \equiv \sigma_X^2$ se $s = t$ (nel seguito useremo il termine di “scambiabilità del secondo ordine” per indicare un modello di questo tipo).

a) Funzione di covarianza e funzione spettrale

Lo studio dei processi stazionari (del secondo ordine) può avvenire da due diversi punti di vista: l'approccio **temporale** che utilizza la funzione di covarianza quale strumento principale e l'approccio **frequenziale** (o spettrale) che utilizza quale strumento principale la funzione spettrale, definita come la trasformata di Fourier della funzione di covarianza. Si ha in proposito il seguente

Teorema: se $\{X_t; t \geq 1\}$ è un processo stazionario in senso lato, a valori reali, con funzione di covarianza $\Psi_X(h)$, $h \in \mathbb{Z}$, si ha :

- a) $\Psi_X(h)$ è semi-definita positiva, cioè soddisfa, per ogni intero positivo N ed ogni sequenza di numeri reali (a_1, \dots, a_N) , la condizione $\sum_{i,j=1}^N a_i \cdot a_j \cdot \Psi_X(i-j) \geq 0$ e, viceversa, ogni funzione definita in \mathbb{Z} e semi-definita positiva è la funzione di covarianza di un processo stazionario;
- b) esiste una funzione (spettrale) $F(\lambda)$, a valori reali, monotona non decrescente e limitata, tale che $\Psi_X(h)$ ha la rappresentazione $\Psi_X(h) = \int_{-\pi}^{\pi} \cos(\lambda h) dF(\lambda)$; la corrispondenza tra l'insieme delle $\Psi_X(h)$ e quello delle funzioni spettrali $F(\lambda)$ può essere resa biunivoca imponendo, per esempio, l'ulteriore condizione $F(-\pi) = 0$ e assumendo $F(\lambda)$ continua a destra in ogni eventuale punto di discontinuità;
- c) se $\Psi_X(h)$ è assolutamente sommabile, cioè se verifica la condizione $\sum_{h=-\infty}^{+\infty} |\Psi_X(h)| < +\infty$, allora esiste $f(\lambda) = F'(\lambda)$, detta densità spettrale, continua in $(-\pi, \pi)$, simmetrica rispetto a $\lambda = 0$, tale che $\Psi_X(h) = \int_{-\pi}^{\pi} \cos(\lambda h) \cdot f(\lambda) d\lambda$ e che $f(\lambda) = (2\pi)^{-1} \cdot \sum_{h=-\infty}^{+\infty} \cos(\lambda h) \cdot \Psi_X(h)$.

Per la dimostrazione di questo teorema e degli altri che seguiranno rinviamo il lettore per esempio al testo di P.J.Brockwell e R.A.Davis “Time Series: Theory and Methods” della Springer. Vediamo ora alcuni esempi di processi stazionari e di funzioni spettrali:

1. **Processi con numeri aleatori i.i.d.** e dotati di momento secondo finito: la funzione di covarianza $\Psi_X(h)$ assume il valore σ_X^2 se $h = 0$ ed è nulla per gli altri valori di h ; la corrispondente densità spettrale è identicamente uguale a $\sigma_X^2 / 2\pi$. Si noti che anche il processo WN $(0; \sigma^2)$ ha la medesima densità spettrale.
2. **Processo armonico** $X_t = Y \cdot \cos(\lambda t) + Z \cdot \sin(\lambda t)$, $0 < \lambda < \pi$, ove le speranze matematiche di Y e Z sono assunte nulle, le varianze sono assunte uguali e denotate con σ^2 e la loro

covarianza è assunta nulla; il parametro λ , denominato “frequenza angolare”, è assunto positivo e non maggiore di π . E’ immediato provare che $\varphi(t) = 0$ e si trova facilmente la

$\psi(s,t) = \sigma^2 \cdot \cos \lambda(t-s)$. Dal momento che questa funzione di covarianza **non** è assolutamente sommabile, non esiste una corrispondente densità spettrale. La funzione spettrale $F(\lambda)$ corrispondente alla $\psi(s,t)$ è una funzione a gradini con discontinuità in $-\lambda$ e λ e l’ampiezza dei salti è pari a $\sigma^2/2$. Questo stesso processo stocastico può avere altre due

rappresentazioni equivalenti; la prima è $X_t = V \cdot \sin(\lambda t + \eta)$, ove il nuovo parametro η , detto “spostamento di fase”, è definito ponendo $Y = V \cdot \sin \eta$ e $Z = V \cdot \cos \eta$, con $V = (Y^2 + Z^2)^{1/2}$, di modo che è banalmente $\sin \eta = Y/V$ e $\cos \eta = Z/V$.

La seconda rappresentazione alternativa è $X_t = A^* \cdot e^{-i\lambda t} + A \cdot e^{i\lambda t}$, ove si è posto $A = (Y - iZ)/2$ e ove A^* denota il numero complesso coniugato di A . Questa seconda rappresentazione introduce formalmente la frequenza angolare negativa $-\lambda$ e numeri aleatori a valori complessi, ma si vedrà nel seguito che tale complicazione formale consente un’effettiva semplificazione negli sviluppi matematici.

3. Processi ottenuti da **somme finite di processi armonici**.

Sia $X_t = \sum_{j=1}^N (Y_j \cdot \cos \lambda_j t + Z_j \cdot \sin \lambda_j t)$, ove $0 < \lambda_1 < \dots < \lambda_N < \pi$; i numeri aleatori Y_j e Z_j

abbiano speranza matematica nulla, varianza pari a σ_j^2 e siano tutti mutuamente non correlati. La funzione valor medio di $\{X_t\}$ è identicamente nulla e la sua funzione di covarianza è data da

$\psi(s,t) = \sum_{j=1}^N \sigma_j^2 \cdot \cos \lambda_j(t-s)$. La corrispondente funzione spettrale $F(\lambda)$ ha $2N$ discontinuità,

simmetriche rispetto a $\lambda = 0$, di ampiezza $\sigma_j^2/2$ e risulta $F(\pi) = \sum_{j=1}^N \sigma_j^2$.

Come per il precedente processo armonico, sussistono altre due rappresentazioni possibili per

l’attuale processo: $X_t = \sum_{j=1}^N V_j \cdot \sin(\lambda_j t + \eta_j)$ e $X_t = \sum_{j=-N}^N A_j \cdot \exp(i\lambda_j t)$. Nell’ultima

rappresentazione, le condizioni affinché $\{X_t\}$ sia un processo a valori reali sono $A_j = (Y_j - iZ_j)/2$ se $j \geq 0$, $A_{-j} = A_j^*$ e $\lambda_{-j} = -\lambda_j$.

Tale processo può essere esteso ad un numero infinito di addendi se si suppone finita la somma della serie delle varianze σ_j^2 ; questa estensione ha una rilevante importanza dal punto di vista teorico per la rappresentazione di processi stazionari “con spettro discreto”.

Sussiste in proposito il seguente **teorema di E.E. Slutskij** (1938): ogni processo stazionario in senso lato con funzione valor medio nulla e funzione spettrale costante a tratti (spettro discreto) può essere rappresentato secondo la

$$X_t = \sum_{j=1}^{\infty} b_j \cdot (Y_j \cdot \cos \lambda_j t + Z_j \cdot \sin \lambda_j t),$$

con $\sum_{j=1}^{\infty} b_j^2 < \infty$, ove i n.a. Y_j e Z_j sono tutti equi, mutuamente non correlati con $V(Y_j) = V(Z_j) =$

σ_j^2 . La generalizzazione di questo risultato a tutti i processi stazionari in senso lato con funzione valor medio nulla e funzione spettrale arbitraria è dovuta a Kolmogorov, Cramér e Loève.

4) **Processo MA(1)** : $X_t = \varepsilon_t + b \cdot \varepsilon_{t-1}$ con $\varepsilon_t \sim \text{WN}(0; \sigma_\varepsilon^2)$. Si ha $\Psi_X(0) = (1 + b^2) \cdot \sigma_\varepsilon^2$,

$\Psi_X(1) = b \sigma_\varepsilon^2$, mentre $\Psi_X(h) = 0$ per $h > 1$. Evidentemente esiste una densità spettrale, poiché

$\sum_{-\infty}^{+\infty} |\Psi_X(h)| = (1 + b^2) \sigma_\varepsilon^2 + 2b \sigma_\varepsilon^2$, e la sua espressione è :

$$f(\lambda) = \sigma_\varepsilon^2 (2\pi)^{-1} (1 + 2b \cos \lambda + b^2).$$

b) Filtri lineari invarianti

Si definisce “filtro lineare invariante” applicato ad un processo stocastico $\{Y_t; t \geq 1\}$ una

trasformazione lineare $X_t = \sum_{j=-\infty}^{+\infty} c_j \cdot Y_{t-j}$ del processo $\{Y_t\}$ con coefficienti c_j indipendenti da t e tali

da soddisfare una qualche condizione del tipo $\sum_{j=-\infty}^{+\infty} c_j^2 < \infty$ o $\sum_{j=-\infty}^{+\infty} |c_j| < \infty$.

Sussiste il seguente:

Teorema: se la successione di coefficienti $\{c_j\}$ è assolutamente sommabile e il processo $\{Y_t\}$ è stazionario in senso lato allora $\{X_t\}$ è stazionario in senso lato con funzione di autocovarianza

$\Psi_X(h) = \sum_{j,k=-\infty}^{+\infty} c_j \cdot c_k \cdot \Psi_Y(h - j + k)$ e funzione spettrale verificante la $dF_X(\lambda) = |H(\lambda)|^2 \cdot dF_Y(\lambda)$, ove

$H(\lambda)$ è la funzione di trasferimento del filtro definita dalla $H(\lambda) = \sum_{j=-\infty}^{+\infty} c_j \cdot e^{-ij\lambda}$. Se esiste la densità

spettrale $f_Y(\lambda)$ allora esiste anche $f_X(\lambda)$ che verifica la relazione $f_X(\lambda) = |H(\lambda)|^2 \cdot f_Y(\lambda)$.

Nell'esempio 4) è $H(\lambda) = 1 + b \cdot \exp(-i\lambda)$ e $|H(\lambda)|^2 = 1 + 2b \cos(\lambda) + b^2$ per cui si ha $f_X(\lambda) = \sigma_\varepsilon^2 \cdot (1 + 2b \cos \lambda + b^2) / 2\pi$, essendo $\sigma_\varepsilon^2 / 2\pi$ la densità spettrale del processo $\{\varepsilon_t\}$.

5. **Processo AR(1)** : $X_t - a \cdot X_{t-1} = \varepsilon_t$, con $\{\varepsilon_t\} \sim \text{WN}(0; \sigma_\varepsilon^2)$ e $|a| < 1$.

Sostituendo in $X_t = a \cdot X_{t-1} + \varepsilon_t$ a X_{t-1} l'espressione $a \cdot X_{t-2} + \varepsilon_{t-1}$ si ottiene $X_t = a^2 \cdot X_{t-2} + \varepsilon_t + a \cdot \varepsilon_{t-1}$;

continuando allo stesso modo le sostituzioni si perviene alla $X_t = a^n \cdot X_{t-n} + \varepsilon_t + \sum_{k=1}^{n-1} a^k \cdot \varepsilon_{t-k}$.

Passando al limite per n tendente all'infinito si ha $X_t = \varepsilon_t + \sum_{k=1}^{\infty} a^k \cdot \varepsilon_{t-k}$, dalla quale il processo $\{X_t\}$

appare ottenuto con l'applicazione al processo $\{\varepsilon_t\}$ di un filtro lineare invariante. La funzione di

trasferimento di questo filtro è la $H(\lambda) = 1 + \sum_{k=1}^{\infty} a^k \cdot e^{-ik\lambda} = (1 - a \cdot e^{-i\lambda})^{-1}$ e risulta $|H(\lambda)|^2 =$

$(1 - 2a \cos \lambda + a^2)^{-1}$. Si ha dunque $f_X(\lambda) = |H(\lambda)|^2 \cdot f_Y(\lambda) = \sigma_\varepsilon^2 \cdot (1 - 2a \cos \lambda + a^2)^{-1} / 2\pi$.

6. **Processo ARMA(1, 1)** : $X_t - a \cdot X_{t-1} = \varepsilon_t + b \cdot \varepsilon_{t-1}$, con $\{\varepsilon_t\} \sim \text{WN}(0; \sigma_\varepsilon^2)$ e $|a| < 1$.

E' opportuno riguardare il modello suddetto come un modello AR(1), $X_t - a.X_{t-1} = Y_t$, con $Y_t = \varepsilon_t + b.\varepsilon_{t-1}$. Per quanto già visto, $\{Y_t\}$ è un processo stazionario con densità spettrale $f_Y(\lambda) = \sigma_\varepsilon^2 (1 + 2b \cos \lambda + b^2) / 2\pi$ per cui è:

$$f_X(\lambda) = |H(\lambda)|^2 \cdot f_Y(\lambda) = \frac{\sigma_\varepsilon^2 (1 + 2b \cos \lambda + b^2)}{2\pi (1 - 2a \cos \lambda + a^2)}$$

c) Cenni sui processi scambiabili in senso lato

Vedremo che i processi scambiabili sono una sottoclasse di processi stazionari; se si suppongono finiti i momenti secondi dei n.a. X_t di un processo scambiabile, ha senso considerare la condizione di scambiabilità in senso lato: essa è caratterizzata dal fatto che la funzione valor medio $\phi_X(t)$ è costante e la funzione di covarianza $\psi_X(h)$ ha soltanto i due valori $\psi_X(0) = \sigma_X^2$ e $\psi_X(h) = \gamma > 0$ se h è diverso da 0. Evidentemente la $\psi_X(h)$ non è assolutamente sommabile e si prova che la funzione spettrale $F(\lambda)$ ha una discontinuità nell'origine e che la sua espressione è $F(\lambda) = \gamma \cdot F_1(\lambda) + (\sigma_X^2 - \gamma) \cdot F_2(\lambda)$, ove $F_1(\lambda)$ è una funzione di ripartizione che concentra la massa unitaria nell'origine ed $F_2(\lambda)$ una funzione di ripartizione con densità uniforme in $[-\pi, \pi]$.

7. Sia $X_t = Y + \varepsilon_t$ ove $\{\varepsilon_t\}$ è un processo WN $(0; \sigma_\varepsilon^2)$. Il n.a. Y abbia valor medio nullo,

varianza σ_Y^2 e sia non correlato con ciascun n.a. ε_t . E' facile verificare che $E(X_t) \equiv 0$ e che la funzione di covarianza $\Psi_X(h)$ ha due valori: $\Psi_X(0) = \sigma_Y^2 + \sigma_\varepsilon^2$ e $\Psi_X(h) = \sigma_Y^2$ se $h \neq 0$. Siamo in presenza di una funzione di covarianza a due valori, per cui, ponendo $\sigma_Y^2 + \sigma_\varepsilon^2 = \sigma^2$ e $\sigma_Y^2 = \gamma$, la funzione spettrale è la $F(\lambda) = \gamma \cdot F_1(\lambda) + (\sigma^2 - \gamma) \cdot F_2(\lambda)$. In altri termini, il processo $\{X_t\}$ è scambiabile in senso lato.

Si può provare il seguente **teorema di rappresentazione**: tutti e soli i processi stocastici scambiabili in senso lato sono rappresentabili come $X_t = Y + \varepsilon_t$ ove i n.a. a secondo membro hanno le caratteristiche suddette con $\text{Var}(Y) = \Psi_X(h)$, con $h \neq 0$, e $\sigma_\varepsilon^2 = \Psi_X(0) - \Psi_X(h)$.

d) Processi stazionari a valori complessi

Esempio: $X_t = \sum_{j=1}^N A_j \cdot \exp(i\lambda_j t)$, ove i n.a. complessi A_j sono equi, con varianza σ_j^2 e non correlati, mentre i numeri certi λ_j appartengono all'intervallo $(-\pi, \pi]$. Si trova facilmente che i n.a. X_t sono equi e che la loro funzione di covarianza ha l'espressione $\psi_X(h) = \sum_{j=1}^N \sigma_j^2 \cdot \exp(i\lambda_j h)$. La corrispondente funzione spettrale è costante a tratti con discontinuità di ampiezza σ_j^2 nei punti $\lambda = \lambda_j$; naturalmente è $F_X(\lambda_N) = F_X(\pi) = \sum_{j=1}^N \sigma_j^2$.

Forniremo ora due espressioni formali equivalenti per $\psi_X(h)$ e per il processo $\{X_t\}$: poiché è $F_X(\lambda) = \sum_{j:\lambda_j \leq \lambda} \sigma_j^2$ possiamo scrivere $\psi_X(h) = \int_{-\pi}^{\pi} \exp(i\lambda h) dF_X(\lambda)$ e rappresentare quindi $\psi_X(h)$

come un integrale di Stieltjes con funzione peso $F_X(\lambda)$. Ponendo invece $Z(\lambda) = \sum_{j:\lambda_j \leq \lambda} A_j$ si può

scrivere $X_t = \int_{-\pi}^{\pi} \exp(i\lambda t) dZ(\lambda)$: si tratta di un integrale stocastico, del tipo detto “integrale di

Wiener”, rispetto al processo $\{Z(\lambda)\}$. Quest’ultimo è un processo a parametro continuo definito in $(-\pi, \pi]$ con funzione valor medio identicamente nulla e con incrementi non correlati; inoltre è $E|Z(\lambda)|^2 = F_X(\lambda)$. Viene denominato “processo spettrale” associato a $\{X_t\}$.

Il fatto notevole è che le due suddette rappresentazioni sussistono, con diverse funzioni peso $F(\lambda)$ e $Z(\lambda)$, per tutti i processi stazionari in covarianza con valori complessi e con momenti secondi assoluti finiti. Sono entrambe rappresentazioni nel dominio frequenziale o spettrale e costituiscono decomposizioni della funzione di covarianza e del processo stazionario in componenti cicliche di frequenze angolari $\lambda \in (-\pi, \pi]$.

Diamo ora gli enunciati dei corrispondenti teoremi:

Teorema di Herglotz: la funzione di covarianza di un processo $\{Y_t\}$ stazionario in senso lato con $E|Y_t|^2 < +\infty$, in quanto funzione semidefinita positiva, è rappresentabile come

$$\psi_Y(h) = \int_{-\pi}^{\pi} \exp(i\lambda h) dF_Y(\lambda),$$

ove $F_Y(\lambda)$ è a valori reali, monotona non decrescente e limitata. Solitamente si assume anche che essa sia continua a destra negli eventuali punti di discontinuità e che $F_Y(-\pi) = 0$.

Teorema di Kolmogorov, Cramér, Loève: ad ogni processo $\{Y_t\}$ stazionario in senso lato con $E|Y_t|^2 < +\infty$ e funzione valor medio nulla può essere associato un processo stocastico $\{Z_Y(\lambda)\}$, $\lambda \in (-\pi, \pi]$, con incrementi non correlati tale che

$$Y_t = \int_{-\pi}^{\pi} \exp(i\lambda t) dZ_Y(\lambda),$$

ove $E[Z_Y(\lambda)] = 0$, $E|Z_Y(\lambda)|^2 = F_Y(\lambda)$, $E|dZ_Y(\lambda)|^2 = dF_Y(\lambda)$

Infine enunceremo un terzo fondamentale teorema concernente i processi stazionari in covarianza con funzione valor medio identicamente nulla:

Teorema di H. Wold: ogni processo $\{Y_t\}$ stazionario in senso lato con $E|Y_t|^2 < +\infty$ e funzione valor medio nulla può essere rappresentato al modo seguente

$$Y_t = V_t + \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-j}$$

ove $\{\varepsilon_t\} \sim \text{WN}(0; \sigma_\varepsilon^2)$, $\sum_{j=0}^{\infty} \gamma_j^2 < \infty$, $\gamma_0 = 1$, $\text{Cov}(V_t, \varepsilon_s) \equiv 0$ e ove il processo stocastico $\{V_t\}$ è “deterministico” nel senso che esso è completamente determinato dal suo passato.

Si può affermare che il teorema di H. Wold costituisce il fondamento probabilistico della tecnica statistica nota come Analisi delle serie temporali (o delle serie storiche).

e) Processi scambiabili e parzialmente scambiabili

La nozione di **processo scambiabile** è stata introdotta da B. de Finetti nel 1928 con una comunicazione al Congresso Internazionale dei Matematici di Bologna dal titolo “Funzione caratteristica di un fenomeno aleatorio” con la quale:

- introdusse la nozione di processo stocastico scambiabile,
- dimostrò il corrispondente teorema di rappresentazione e
- risolse il problema dell’inferenza statistica in ipotesi di scambiabilità delle variabili osservabili fornendo una giustificazione razionale del principio di induzione.

Mentre la modesta frase iniziale della suddetta comunicazione – “Scopo di questa comunicazione è di mostrare come il metodo della funzione caratteristica, già così vantaggiosamente introdotto nella teoria delle variabili casuali, si presti pure assai utilmente allo studio dei fenomeni aleatori” – non lascia trasparire i notevoli contributi innovativi in essa contenuti, la frase finale - “Queste conclusioni e questi esempi possono chiarire l’influenza che sulla valutazione di una probabilità esercitano i dati dell’esperienza.” – fa intravedere che la ragione principale della introduzione della nozione di scambiabilità era stata l’intenzione di chiarire le condizioni in cui l’osservazione di una frequenza su una sequenza di prove fornisce coerentemente la base per una valutazione di probabilità o per una previsione della frequenza su una sequenza di prove ancora da eseguire. In altri termini l’autore chiarisce l’insieme di ipotesi che giustifica il “principio di induzione”, cioè la propensione a valutare la probabilità di un evento E_{n+1} in termini della frequenza relativa di successo osservata su n eventi **analoghi** ad E_{n+1} ; formalmente:

$$P(E_{n+1} / K) \cong \frac{m}{n} \quad \text{se } K = \left\{ \sum_{j=1}^n |E_j| = m \right\} .$$

Dal punto di vista formale, i processi scambiabili sono caratterizzati dalla condizione di invarianza delle distribuzioni congiunte rispetto a permutazioni arbitrarie degli indici dei n.a., cioè dall’invarianza delle distribuzioni rispetto all’ordine dei n.a. . Si richiede cioè che sia

$$F_{1, \dots, n}(x_1, \dots, x_n) = F_{j_1, \dots, j_n}(x_1, \dots, x_n)$$

per ogni intero positivo $n \geq 1$, per ogni permutazione (j_1, \dots, j_n) della sequenza $(1, \dots, n)$ e per ogni sequenza di argomenti (x_1, \dots, x_n) . Per quanto concerne i momenti fino al secondo ordine si ha evidentemente $\varphi_X(t) = E(X_t) \equiv E(X_1)$ e $\psi_X(s, t) = \text{Cov}(X_s, X_t) = \psi_X(s - t) = \sigma^2$ o $\psi_X(s - t) = \gamma > 0$ a seconda che sia $s - t = 0$ o, rispettivamente, $s - t \neq 0$. La suddetta definizione implica che ogni insieme finito di n n.a. distinti X_t abbia la stessa distribuzione di probabilità congiunta, cioè

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_n(x_1, \dots, x_n),$$

qualunque sia l'intero positivo n , la sequenza di interi positivi distinti (t_1, \dots, t_n) e la sequenza di argomenti (x_1, \dots, x_n) e ove F_n indica la funzione di ripartizione congiunta per un qualunque insieme di n n.a. distinti.

Dal punto di vista interpretativo la condizione di scambiabilità traduce l'idea di **analogia o equivalenza** tra i n.a. osservabili X_t in modo ben più efficace della condizione di indipendenza stocastica e uguale distribuzione degli stessi: infatti, se gli X_t sono assunti mutuamente indipendenti l'apprendimento dall'esperienza non può avvenire. Per il caso $X_t = |E_t|$ l'assunzione di indipendenza tra gli eventi equivale ad assumere che sia $P(E_{n+1} / K) \equiv P(E_{n+1})$ qualunque sia l'evento osservabile K riguardante gli eventi E_1, \dots, E_n . Per contro, l'assunzione di scambiabilità tra gli eventi introduce tipicamente una dipendenza stocastica tra essi e questa implica la $P(E_{n+1} / K) \equiv \frac{m}{n}$, come facilmente si può provare.

Sussiste un importante teorema di rappresentazione per i processi scambiabili illimitati, cioè costituiti da una infinità numerabile di n.a. X_t .

Teorema di B. de Finetti: tutti e soli i processi scambiabili illimitati $\{X_t; t \geq 1\}$ sono combinazioni lineari (o misture) di processi stocastici $\{X_t^{(\omega)}; t \geq 1\}$, $\omega \in \Omega$, i cui n.a. sono indipendenti e dotati di una comune funzione di ripartizione $F^{(\omega)}(x)$ nel senso che esiste una funzione di ripartizione $G(\omega)$ tale che per ogni intero positivo n ed ogni sequenza finita di indici (t_1, \dots, t_n) sussiste la

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \int_{\Omega} F^{(\omega)}(x_1, \dots, x_n) dG(\omega) = \int_{\Omega} \left[\prod_{j=1}^n F^{(\omega)}(x_j) \right] dG(\omega).$$

In questa sede ci limitiamo a fornire due semplici esemplificazioni del suddetto teorema di rappresentazione: nel primo esempio le funzioni di ripartizione univariate $F^{(\omega)}$ siano tutte di tipo normale o Gaussiano con valor medio ω e varianza unitaria; supponiamo ancora che $G(\omega)$ sia la funzione di ripartizione di una distribuzione normale con parametri μ e σ^2 . Si ha allora:

$$\begin{aligned} F_{t_1, \dots, t_n}(x_1, \dots, x_n) &= \int_R \left[\prod_{j=1}^n (2\pi)^{-1/2} \cdot \exp\left\{-\frac{(x_j - \omega)^2}{2}\right\} \right] dG(\omega) = \\ &= (2\pi)^{-n/2} \int_R \left[\exp\left\{-\frac{1}{2} \sum_{j=1}^n (x_j - \omega)^2\right\} \right] \cdot \left[(2\pi\sigma^2)^{-1/2} \cdot \exp\left\{-\frac{\sigma^2}{2}(\omega - \mu)^2\right\} \right] d\omega \end{aligned}$$

dalla quale si ricava, con facili calcoli, che $F_{t_1, \dots, t_n}(x_1, \dots, x_n)$ corrisponde ad una distribuzione normale n – dimensionale caratterizzata da valori medi tutti uguali a μ , varianze tutte uguali a $1 + \sigma^2$ e covarianze tutte uguali a σ^2 .

Nel secondo esempio i n.a. X_t siano indicatori di eventi $|E_t|$ scambiabili per i quali si considera il prodotto logico $A = \{E_1' \wedge E_2' \wedge \dots \wedge E_n'\}$ implicante l'evento $\left\{ \sum_{j=1}^n |E_j| = m \right\}$; supponiamo anche che sia noto il valore logico (vero o falso) di ogni evento del prodotto logico A . Le funzioni di

ripartizione $F^{(\omega)}(x)$ sono identicamente nulle per $x < 0$, uguali a ω per $0 \leq x < 1$ e identicamente uguali a 1 per $x > 1$. Se assumiamo che $G(\omega)$ sia la funzione di ripartizione di una distribuzione Beta con parametri numerici α e β l'applicazione del teorema di rappresentazione fornisce la

$$P(A) = \int_0^1 [\omega^m \cdot (1-\omega)^{n-m}] \cdot \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \omega^{\alpha-1} \cdot (1-\omega)^{\beta-1} \right] d\omega$$

dalla quale si ricava, con facili calcoli, che è $P(A) = \frac{(\alpha)_m \cdot (\beta)_{n-m}}{(\alpha + \beta)_n}$, ove $(\alpha)_m = \prod_{j=0}^{m-1} (\alpha + j)$.

La condizione di **scambiabilità parziale** è stata introdotta sempre da B. de Finetti nel 1937 ed è più flessibile rispetto a quella di scambiabilità e quindi più adatta a modellizzare schemi nei quali l'analogia tra le unità campionarie non è considerata perfetta. Si pensi a misurazioni ripetute di una stessa grandezza effettuate con strumenti di misura aventi precisioni diverse, a rischi assicurativi (per esempio contratti R.C.A.) giudicati non omogenei (per esempio per la diversa cilindrata degli autoveicoli), ai risultati aleatori di lanci ripetuti di monete diverse e così via.

Il modello più semplice di processo parzialmente scambiabile consiste di un **insieme finito** (per esempio una coppia, o una terna,.....) di **processi stocastici scambiabili, tra loro correlati**. Si considerino, per semplicità, due processi stocastici distinti $\{X_t; t \geq 1\}$ e $\{Y_t; t \geq 1\}$ ciascuno dei quali è supposto essere scambiabile. Circa i legami di dipendenza stocastica tra i due processi si possono ipotizzare tante situazioni diverse: i casi limite sono quello di indipendenza tra essi e, all'opposto, quello di scambiabilità semplice generalizzata (in cui non c'è bisogno di considerare distinti i n.a. del primo da quelli del secondo processo). Tra queste situazioni limite c'è tutta la gamma di situazioni di scambiabilità parziale non banale: ciò che le accomuna è la condizione di uguale covarianza tra ogni n.a. X_t del primo processo e tutti i n.a. Y_s del secondo, cioè $Cov(X_t, Y_s) = \gamma$ per ogni coppia di valori (t,s) degli indici. Ovviamente dovrà essere $|\gamma| \leq \sqrt{\sigma_x^2 \sigma_y^2}$.

Dal punto di vista formale, i processi parzialmente scambiabili (costituiti da due processi scambiabili) sono caratterizzati dalla seguente proprietà di invarianza delle distribuzioni congiunte:

$$F(x_{t_1}, \dots, x_{t_n}; y_{s_1}, \dots, y_{s_m}) = F(x_1, \dots, x_n; y_1, \dots, y_m)$$

per ogni coppia di interi non negativi n, m e ogni coppia di sequenze di interi positivi distinti t_1, \dots, t_n e s_1, \dots, s_m .

Sussiste il seguente risultato dovuto a B. De Finetti:

Teorema di rappresentazione: tutti e soli i processi parzialmente scambiabili illimitati (costituiti da due processi scambiabili componenti) sono combinazioni lineari (o misture) di processi stocastici i cui n.a. sono indipendenti, con una funzione di ripartizione $F_1^{(\omega)}(x)$ comune a tutti i n.a. X_t e una funzione di ripartizione $F_2^{(\eta)}(y)$ comune a tutti i n.a. Y_s nel senso che esiste una distribuzione congiunta $G(\omega, \eta)$ tale che sia

$$F(x_{t_1}, \dots, x_{t_n}; y_{s_1}, \dots, y_{s_m}) = \iint_{\{\omega, \eta\}} \left[\prod_{i=1}^n F_1^{(\omega)}(x_{t_i}) \right] \cdot \left[\prod_{j=1}^m F_2^{(\eta)}(y_{s_j}) \right] dG(\omega, \eta)$$

per ogni coppia di interi non negativi (n,m) e ogni coppia di sequenze di interi positivi t_1, \dots, t_n e s_1, \dots, s_m .

In un primo esempio di applicazione del teorema precedente in cui X_t e Y_s sono indicatori di eventi sia $g(\theta, \eta)$, la densità congiunta corrispondente a $G(\theta, \eta)$, una densità di probabilità di tipo Beta bivariata con parametri reali positivi ν_1, ν_2 e ν_3 , cioè

$$g(\theta_1, \theta_2) = \frac{\Gamma(\nu_1 + \nu_2 + \nu_3)}{\Gamma(\nu_1)\Gamma(\nu_2)\Gamma(\nu_3)} \theta_1^{\nu_1-1} \theta_2^{\nu_2-1} (1 - \theta_1 - \theta_2)^{\nu_3-1}.$$

Si prova allora che i processi $\{X_t\}$ e $\{Y_s\}$ riescono entrambi scambiabili con

$$P\left\{\sum_{t=1}^n X_t = n\right\} = \frac{(\nu_1)_n}{(\nu_1 + \nu_2 + \nu_3)_n} \text{ e, rispettivamente, } P\left\{\sum_{t=1}^n Y_t = n\right\} = \frac{(\nu_2)_n}{(\nu_1 + \nu_2 + \nu_3)_n}; \text{ si ha infine}$$

$$P\left\{\left[\sum_{t=1}^n X_t = n\right] \wedge \left[\sum_{s=1}^m Y_s = m\right]\right\} = \frac{(\nu_1)_n \cdot (\nu_2)_m}{(\nu_1 + \nu_2 + \nu_3)_{n+m}}.$$

Quest'ultima probabilità è quella che n eventi del primo tipo (con indicatori X_t) ed m eventi del secondo tipo (con indicatori Y_s) siano tutti veri, cioè che la frequenza di successo su n eventi del primo tipo ed m eventi del secondo tipo sia pari a $n + m$.

In un secondo esempio di applicazione le funzioni di ripartizione univariate $F_1^{(\omega)}(x)$, comune a tutti i n.a. X_t , ed $F_2^{(\eta)}(y)$, comune a tutti i n.a. Y_s , siano entrambe di tipo Gaussiano con valori medi ω ed η e varianze unitarie; supponiamo ancora che $G(\theta, \eta)$ sia la funzione di ripartizione di una distribuzione Gaussiana bivariata con vettore medio $\mu = (\mu_1, \mu_2)^T$ e matrice di varianze e covarianze $\Gamma = \begin{bmatrix} \gamma_1 & \gamma \\ \gamma & \gamma_2 \end{bmatrix}$. Si verifica allora facilmente che $F(x_1, \dots, x_n; y_1, \dots, y_m)$ è la funzione di ripartizione di una distribuzione Gaussiana $(n+m)$ -dimensionale caratterizzata dai momenti $E(X_t) \equiv \mu_1$, $E(Y_s) \equiv \mu_2$, $Var(X_t) \equiv 1 + \gamma_1$, $Var(Y_s) \equiv 1 + \gamma_2$, $Cov(X_t, Y_s) \equiv \gamma$.

Per ulteriori approfondimenti sui processi scambiabili e parzialmente scambiabili si possono consultare, per esempio:

- 1) B. de Finetti – “Teoria delle probabilità” (G. Einaudi, 1970 oppure, la corrispondente edizione in lingua inglese edita da J. Wiley, 1974).
- 2) L. Daboni e A. Wedlin – “Statistica: un'introduzione all'impostazione neo-bayesiana” (UTET, 1982).
- 3) J.M. Bernardo and A.F.M. Smith – “Bayesian Theory” (J. Wiley, 1994).

Quale testo di riferimento per la teoria dei processi stocastici a parametro discreto indichiamo:

A.N. Shiryaev – “Probability” (Springer-Verlag, 1984).

APPENDICE n. 1 : Cenni sui processi di ramificazione (Branching processes)

E' un esempio di processo a **catena markoviana con un insieme numerabile di stati**. Si pensi ad un individuo (essere umano, animale, particella subatomica, ...) idoneo a "generare" un numero aleatorio di discendenti: esso costituisce la generazione zero mentre i suoi discendenti formano la prima generazione; i discendenti dei discendenti formano la seconda generazione e così via.

Il parametro operativo individua le successive generazioni cosicché si ha: $X_0 = 1$, X_1 denota il numero di discendenti costituenti la prima generazione, X_2 il numero di individui della seconda generazione, etc. Supporremo che i numeri di discendenti di differenti individui, appartenenti alla stessa o a differenti generazioni, siano stocasticamente indipendenti

tra loro e dotati della medesima distribuzione di probabilità $\{p_j; j \geq 0\}$ della quale indicheremo con $G(s) = \sum_j p_j \cdot s^j$

la funzione generatrice (p.g.f.). Per evitare casi banali assumeremo $p_0 > 0$ e $p_0 + p_1 < 1$.

Per quanto detto si ha:

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i = Z_1 + Z_2 + \dots + Z_{X_n},$$

ove Z_i indica il numero di discendenti generati dall' i -esimo individuo della generazione X_n ; gli addendi Z_i sono assunti i.i.d. con la comune distribuzione $\{p_j\}$ e p.g.f. $G(s)$.

Il carattere markoviano del processo $\{X_n\}$ è chiaramente rivelato dalle:

$$\text{Prob} [X_{n+1} = k_{n+1} / \bigcap_{i=1}^n (X_i = k_i)] = \text{Prob} [\sum_{i=1}^{k_n} Z_i = k_{n+1}] = \text{Prob} [X_{n+1} = k_{n+1} / X_n = k_n].$$

Per quanto concerne le probabilità non condizionate si ha:

$$\begin{aligned} p_{k(n+1)} &= \text{Prob} [X_{n+1} = k] = \sum_{j=0}^{\infty} \text{Prob} (X_{n+1} = k / X_n = j) \cdot \text{Prob} (X_n = j) = \\ &= \sum_j \text{Prob} (Z_1 + Z_2 + \dots + Z_j = k) \cdot \text{Prob} (X_n = j) = \sum_j p_{jk} \cdot p_j(n), \end{aligned}$$

avendo indicato con p_{jk} le probabilità subordinate $\text{Prob} (\sum_{i=1}^j Z_i = k) = \text{Prob} (X_{n+1} = k / X_n = j) = \{p_k\}^{j*}$ ove l'ultima espressione indica la convoluzione j -ma di $\{p_j\}$ con se stessa.

Indicando con $F_{n+1}(s)$ la p.g.f. di X_{n+1} , la precedente relazione equivale alla

$$F_{n+1}(s) = \sum_j G^j(s) \cdot p_j(n) = F_n[G(s)],$$

relazione ricorrente per le funzioni generatrici delle probabilità dei n.a. X_n che può anche scriversi

$$F_{n+1}(s) = G[F_n(s)]$$

poiché, essendo $X_0 = 1$ e quindi $F_0(s) = s$, si ha $F_1(s) = F_0[G(s)] = G(s)$, $F_2(s) = F_1[G(s)] = G[G(s)] = G^{(2)}(s)$ (seconda iterata di G) e, in generale, $F_{n+1}(s) = G^{(n+1)}(s) = G[G^{(n)}(s)] = G[F_n(s)]$.

La relazione ricorrente $F_{n+1}(s) = F_n[G(s)] = G[F_n(s)]$ ha un'importanza fondamentale nella teoria dei processi di ramificazione anche se, da un punto di vista operativo, non può essere impiegata per la determinazione delle $F_n(s)$. Essa può però fornire informazioni importanti sui momenti dei n.a. X_n : indicati con μ e v la speranza matematica e la varianza della distribuzione $\{p_j\}$, si ricavano facilmente le relazioni $E(X_{n+1}) = \mu \cdot E(X_n)$ e $\text{Var}(X_{n+1}) = \mu^2 \text{Var}(X_n) + v E(X_n)$.

Un problema importante concerne la possibilità di estinzione dell'insieme dei discendenti, cioè la possibilità che i n.a. X_n siano definitivamente nulli. In proposito sussiste il seguente risultato dovuto a Steffensen (1930):

Teorema : la probabilità ξ di estinzione in tempo finito è data dalla più piccola radice positiva dell'equazione $G(x) = x$. Se $\mu = G'(1) \leq 1$ allora è $\xi = 1$ e l'estinzione è certa; se $\mu > 1$ allora $\xi < 1$ e la probabilità che la popolazione cresca indefinitamente è data da $1 - \xi$.

Una traccia di dimostrazione si ricava dalle seguenti considerazioni:

- 1) innanzitutto la successione delle $p_0(n) = \text{Prob}(X_n = 0)$ è monotona e limitata cosicché esiste il relativo limite ξ per $n \rightarrow \infty$;
- 2) dalla relazione $F_{n+1}(s) = G[F_n(s)]$ si ricava la $p_0(n+1) = G[p_0(n)]$ e passando al limite si ottiene $\xi = G(\xi)$;
- 3) si verifica facilmente che $G(s)$ è concava verso l'alto e quindi monotona crescente; poiché $G(0) = p_0 > 0$ per ipotesi e $G(1) = 1$, si ha che l'esistenza di una o due radici dell'equazione $G(x) = x$ in $[0, 1]$ dipende dal fatto che $G'(1) = \mu$ sia ≤ 1 o > 1 ; quindi, se $\mu \leq 1$ è $\xi = 1$ mentre se $\mu > 1$ allora $\xi < 1$ e la probabilità che la popolazione cresca indefinitamente è $1 - \xi$.

Un esempio: $G(s) = (ps + q) / (1 + r - rs)$, con $p + q = 1$ e p, q, r non negativi.

Si ricava $\mu = G'(s) = p + r$; inoltre è $p_0 = q / (1-r)$, $p_1 = (p+r) / (1+r)^2$, $p_2 = r / (1+r)^3$ e in generale $p_n = r^{n-1} / (1+r)^{n+1}$. L'equazione $G(s) = s$ ha due radici pari a q/r e 1 ; essendo $\xi = \min(q/r, 1)$ è $\xi = (q/r)$ se $p+r > 1$.

In una ricerca riguardante l'estinzione dei cognomi familiari per i maschi bianchi degli USA nel 1920, Lotka trovò che la p.g.f. suddetta con $q/(1+r) = 0.4981$ e $r/(1+r) = 0.5586$ rappresentava bene il fenomeno concreto. Si ottiene, per tale funzione, $\mu = 1.14$ e $\xi = 0.89$: a parole, un cognome di un maschio bianco aveva negli USA del 1920 una probabilità di estinzione pari a 0.89, mentre quella di non estinzione era 0.11.

APPENDICE N. 2 : Urna di Pòlya ed alcuni processi stocastici associati

Com'è noto, è detta "urna di Pòlya" un sistema di estrazioni casuali da un'urna contenente inizialmente b palline bianche ed r palline rosse; dopo un'estrazione casuale di una pallina, questa viene rimessa nell'urna assieme a c palline dello stesso colore di quella estratta: è chiaro che la composizione dell'urna varia colpo per colpo. Considereremo tre processi stocastici associati alla sequenza potenzialmente illimitata di estrazioni successive dall'urna:

- a) il processo $\{Y_n; n \geq 1\}$ dei risultati delle successive estrazioni, ove $Y_n = 1$ o 0 a seconda che l'ennesima estrazione dia pallina bianca oppure rossa;
- b) il processo $\{X_n; n \geq 1\}$ relativo al numero di palline bianche nell'urna dopo le estrazioni successive, ove X_n indica il numero di palline bianche presenti nell'urna dopo le prime n estrazioni;
- c) il processo $\{Z_n; n \geq 1\}$ delle frazioni o percentuali di palline bianche nell'urna dopo le successive estrazioni; è chiaramente $Z_n = X_n / (b + r + n.c)$, in quanto dopo n estrazioni le palline nell'urna sono $b+r+n.c$.

Si vedrà che il processo sub a) è **stazionario**, anzi scambiabile, il processo sub b) è **markoviano** e che infine il processo sub c) è una **martingala**; ovviamente tutti i processi sono a parametro discreto. Sono quindi rappresentate, nello schema di estrazioni suddetto, le tre principali categorie di processi stocastici. Per motivi di semplicità, nel seguito supporremo $c = 1$ di modo che il numero complessivo di palline nell'urna, dopo ogni estrazione, aumenta di un'unità; dopo n estrazioni le palline nell'urna passano dalle iniziali $b + r$ a $b + r + n$, quelle bianche da b a $b + S_n$ e quelle rosse da r a $r + n - S_n$, ove S_n indica il numero aleatorio $Y_1 + \dots + Y_n$, cioè il numero di palline bianche uscito dall'urna nelle prime n estrazioni successive. Porremo inoltre $X_0 = b$ e $Z_0 = b / (b + r)$.

E' abbastanza evidente che lo schema dell'urna di Pòlya può costituire un modello semplificato per **fenomeni di contagio**: infatti dopo l'estrazione di una pallina bianca (o rossa) aumenta il numero e la percentuale di bianche (o rosse) nell'urna per cui aumenta anche la probabilità di un'altra estrazione di pallina bianca (o rossa) al colpo successivo.

Cominceremo a fare alcune considerazioni di carattere generale: con riferimento ad un generico prodotto logico dei primi n eventi $E'_1 \wedge E'_2 \wedge \dots \wedge E'_n$, ove h eventi sono affermati ed $n-h$ sono negati, in un ordinamento qualsiasi delle affermazioni e negazioni, e valutando la probabilità $P(E'_1 \wedge E'_2 \wedge \dots \wedge E'_n)$ secondo la procedura sequenziale

$$P(E'_1 \wedge E'_2 \wedge \dots \wedge E'_n) = P(E'_1) \cdot P(E'_2 / E'_1) \cdot \dots \cdot P(E'_n / E'_1 \wedge \dots \wedge E'_{n-1})$$

ci si rende conto facilmente che dei prodotti logici subordinanti conta, per la valutazione, soltanto il numero degli eventi coinvolti e quello degli eventi affermati, o negati; l'ordine delle affermazioni e negazioni è irrilevante. Ciò comporta, come si può dimostrare, che ogni prodotto logico di n eventi distinti (non necessariamente i primi n) implicante l'evento $\{S_n = h\}$ ha la medesima probabilità che dipende solo dagli interi n ed h :

$$P(E'_1 \wedge E'_2 \wedge \dots \wedge E'_n) = \frac{(b+h-1)_h \cdot (r+n-h-1)_{n-h}}{(b+r+n-1)_n},$$

ove il simbolo $(n)_h$ sta per $(n)_h = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-h+1)$. Evidentemente, la probabilità dell'evento $\{S_n = h\}$ è data dalla

$$P(S_n = h) = \binom{n}{h} \cdot P(E'_1 \wedge E'_2 \wedge \dots \wedge E'_n) = \binom{n}{h} \cdot \frac{(b+h-1)_h \cdot (r+n-h-1)_{n-h}}{(b+r+n-1)_n}$$

che viene denominata "distribuzione di Pòlya". I valori della speranza matematica e della varianza sono dati dalle $n \cdot b / (b+r)$ e, rispettivamente, $n \cdot b \cdot r \cdot (b+r+n) / (b+r)^2 \cdot (b+r+1)$.

Considerando ora i processi $\{X_n; n \geq 1\}$ e $\{Z_n; n \geq 1\}$, avendo posto $X_0 = b$ e $Z_0 = b/(b+r)$, si ha che il n.a. X_n può assumere tutti i valori interi compresi tra b e $b+n$ con probabilità che, come facilmente si verifica, dipendono soltanto dal valore di S_n ; precisamente, l'evento $\{X_n = b+h\}$ si verifica quando e solo quando $S_n = h$ e ciò accade con la probabilità espressa sopra. In corrispondenza è vero l'evento $Z_n = (b+h) / (b+r+n)$. Per quanto concerne le probabilità subordinate $P(X_{n+1} = j / X_n = b+h)$, queste sono diverse da zero solo se $j = b+h$ e $j = b+h+1$ e risulta:

$$P(X_{n+1} = j / X_n = b+h) = \frac{(r+n+h)/(b+r+n)}{(b+h)/(b+r+n)}$$

a seconda che $j = b+h$ oppure $j = b+h+1$.

Proveremo ora alcune proposizioni riguardanti il carattere dei tre processi indicati in precedenza:

Proposizione 1: il processo $\{Y_n; n \geq 1\}$ è stazionario scambiabile.

Dobbiamo provare che le distribuzioni congiunte finite-dimensionali dei n.a. sono invarianti per traslazione rigida nei valori del parametro operativo (condizione di stazionarietà) e addirittura che ogni n -pla di variabili distinte ha la stessa distribuzione congiunta (condizione di scambiabilità). Poiché i n.a. Y_n sono indicatori di eventi, la distribuzione congiunta della sequenza (Y_1, \dots, Y_n) è costituita dalle probabilità dei prodotti logici $E'_1 \wedge E'_2 \wedge \dots \wedge E'_n$ che dipendono, come si è già stabilito, dalla coppia (n, h) , oppure dalla coppia equivalente $(h, n-h)$, e non da quali eventi sono stati considerati. Pertanto non solo è $P(E'_1 \wedge E'_2 \wedge \dots \wedge E'_n) = P(E'_{i_1} \wedge E'_{i_2} \wedge \dots \wedge E'_{i_n})$ ove (i_1, \dots, i_n) è un'arbitraria permutazione degli indici $(1, \dots, n)$ ma anche $P(E'_1 \wedge E'_2 \wedge \dots \wedge E'_n) = P(E'_{j_1} \wedge E'_{j_2} \wedge \dots \wedge E'_{j_n})$ ove (j_1, \dots, j_n) è un'arbitraria n -pla di interi distinti, essendo comunque vero l'evento $\left\{ \sum_{k=1}^n |E_{j_k}| = h \right\}$.

L'invarianza delle valutazioni di probabilità rispetto alla scelta dei numeri aleatori del processo rivela la scambiabilità dello stesso e, a maggior ragione, la sua stazionarietà.

Proposizione 2: il processo $\{X_n; n \geq 1\}$ è markoviano.

Ricordiamo che la markovianità del processo sussiste se, per ogni intero n e ogni scelta dei valori dei n.a., le valutazioni di probabilità subordinate soddisfano le condizioni $P \left[X_{n+1} = x_{n+1} / \bigcap_{j=1}^n (X_j = x_j) \right] = P[X_{n+1} = x_{n+1} / X_n = x_n]$. Nel nostro caso porremo $x_j = b + h_j$ essendo $h_1 \leq h_2 \leq \dots \leq h_{n+1}$. Nell'ipotesi che, dopo le prime n estrazioni, si sia arrivati ad avere nell'urna $b + h_n$ palline bianche, qualunque sia stata la sequenza (h_1, \dots, h_n) , avremo

$$P \left[X_{n+1} = b + h_{n+1} / \bigcap_{j=1}^n (X_j = b + h_j) \right] = \frac{r + n - h_n}{b + r + n} \cdot \frac{b + h_n}{b + r + n},$$

a seconda che sia $h_{n+1} = h_n$ oppure $h_{n+1} = h_n + 1$. Dal fatto che nelle espressioni suddette non compaiano gli h_j precedenti ad h_n discende la dipendenza markoviana tra i n.a. X_n ; dal fatto che nelle espressioni compare l'indice n discende che la dipendenza markoviana non è omogenea.

Proposizione 3: il processo $\{Z_n; n \geq 1\}$ è una martingala.

Per dimostrare che sussistono, per ogni intero n , le condizioni $E(Z_{n+1} / Z_1, Z_2, \dots, Z_n) = Z_n$ occorre provare che in corrispondenza ad ogni possibile evento concernente i n.a. subordinanti la speranza matematica subordinata di Z_{n+1} coincide con il valore ipotizzato per Z_n . Considereremo pertanto il valor medio condizionato relativo ad una

ipotesi qualsiasi, $\bigcap_{j=1}^{n-1} (Z_j = z_j)$, per i primi $n-1$ numeri Z_j congiuntamente alla $Z_n = (b+h) / (b+r+n)$; si ha allora:

$$E \left\{ Z_{n+1} / \left[\bigcap_{j=1}^{n-1} (Z_j = z_j) \right] \bigcap \left(Z_n = \frac{b+h}{b+r+n} \right) \right\} = \frac{b+h}{b+r+n+1} \cdot \frac{r+n-h}{b+r+n} + \frac{b+h+1}{b+r+n+1} \cdot \frac{b+h}{b+r+n} = \frac{b+h}{b+r+n}$$

e quindi la tesi.

Il modello di Ehrenfest : un approfondimento

Si è già introdotto brevemente il modello dell'urna di Ehrenfest presentando i primi esempi di catene markoviane: esso era caratterizzato dai parametri $c = -1$ e $d = 1$. A parole, ad ogni estrazione casuale di una pallina da un'urna contenente inizialmente b palline bianche ed r palline rosse si sostituiva nell'urna la pallina estratta con un'altra del colore opposto. Indicato con X_n il numero di palline bianche nell'urna dopo n estrazioni successive, il processo $\{X_n\}$ risultava essere una catena di Markov, con $b+r+1$ stati, caratterizzata dalle probabilità subordinate di transizione:

$$p_{i,i-1} = i / (b+r), \quad p_{i,i+1} = (b+r-i) / (b+r) \quad \text{e} \quad p_{i,j} = 0 \quad \text{se } j \text{ è diverso da } i-1 \text{ e } i+1.$$

Considereremo ora una generalizzazione del suddetto schema consistente nell'ipotesi che la pallina scelta possa venire rimessa nell'urna, senza cambiamento di colore, con probabilità assegnata; precisamente si suppone che la **dinamica della composizione dell'urna sia stocastica** e regolata dalla matrice di transizione

$$(*) \quad \begin{matrix} & \text{B} & \text{R} \\ \text{B} & \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \\ \text{R} & & \end{matrix},$$

secondo la quale vi è probabilità pari a $1-\alpha$, se la pallina scelta è di colore bianco, e $1-\beta$, se di colore rosso, che essa sia rimessa nell'urna, mantenendo quindi la composizione di palline bianche e rosse precedente. Essa cambia invece di colore con probabilità rispettive α e β .

Un diverso tipo di descrizione dello schema di Ehrenfest prevede due urne (I e II) al posto di una, in sostituzione dei due colori, bianco e rosso, delle palline: le $b+r = N$ palline, ora del medesimo colore, sono distribuite tra le due urne, b nella I ed r nella II. Quando una di esse viene scelta, con probabilità $1/N$, viene poi posta nell'altra urna con probabilità α o β a seconda che essa sia stata estratta dall'urna I o II, o rimessa al suo posto con probabilità rispettive $1-\alpha$ o $1-\beta$.

In entrambi i tipi di descrizione un'ulteriore perfezionamento può essere introdotto dichiarando eventualmente **distinguibili** tra loro le palline, per esempio pensandole numerate da 1 a N ; in questo caso il numero degli stati della catena di Markov sale da $N+1$ a 2^N .

Se le palline **non sono distinguibili**, allo scopo di determinare la matrice P di transizione della catena di Markov $\{X_n\}$, si supponga che siano $i, i \neq 0$ e $i \neq N$, le palline nell'urna I (oppure i le palline bianche nell'unica urna); si trovano facilmente i seguenti valori per gli elementi della riga corrispondente di P:

$$p_{i,i-1} = \frac{i}{N}\alpha, \quad p_{i,i} = \frac{i}{N}(1-\alpha) + \frac{N-i}{N}(1-\beta), \quad p_{i,i+1} = \frac{N-i}{N}\beta \quad \text{e} \quad p_{i,j} = 0 \quad \text{se} \quad j \neq i-1, i, i+1;$$

ovviamente, se $i = 0$ oppure $i = N$ è $p_{0,1} = p_{N,N-1} = 1$.

Supponendo invece che le palline siano tra loro **distinguibili** e che nell'urna I si trovino quelle numerate con n_1, \dots, n_i , e nell'urna II quelle numerate con m_1, \dots, m_{N-i} (denoteremo con K lo stato appena descritto se i è diverso da 0 o da N) allora la riga corrispondente della matrice di transizione P avrà i seguenti elementi:

- α/N in corrispondenza ad ogni stato K_h ($h=1, \dots, i$) che differisce da K per il fatto che le palline nell'urna I sono le $n_1, \dots, n_{h-1}, n_{h+1}, \dots, n_i$ e quelle nell'urna II sono le m_1, \dots, m_{N-i}, n_h (la pallina numero n_h è stata estratta e posta nell'urna II);
- $\frac{i}{N}(1-\alpha) + \frac{N-i}{N}(1-\beta)$ in corrispondenza allo stato K (si tratta della $P(K/K)$);
- β/N in corrispondenza ad ogni stato K_k ($k=1, \dots, N-i$) che differisce da K per il fatto che le palline nell'urna I sono le n_1, \dots, n_i, m_k e quelle nell'urna II sono le $m_1, \dots, m_{k-1}, m_{k+1}, \dots, m_{N-i}$ (la pallina numero m_k è stata estratta e posta nell'urna I).

Nell'ipotesi di N palline **non distinguibili**, si prova che gli N+1 autovalori di P (matrice quadrata con N+1 righe e colonne) sono tutti distinti e appartenenti all'intervallo $[-1, 1]$; gli estremi -1 e 1 sono autovalori e la differenza tra due autovalori successivi è pari a $2/N$. Indicando con Λ la matrice diagonale degli autovalori e con V la matrice dei corrispondenti autovettori (di lunghezza unitaria) si ha la seguente ben nota rappresentazione spettrale della potenza n-ma di P: $P^n = V \Lambda^n V^{-1}$; in teoria essa consente la determinazione delle probabilità subordinate di transizione in n passi e quindi, assegnata la distribuzione di X_0 , quella della distribuzione di ogni $X_n, n \geq 1$.

In alternativa, la funzione generatrice delle probabilità $G_n(s)$ di X_n è data, come si può provare, dalla $G_n(s) = [(1-\alpha_n)s + \alpha_n]^{X_0} \cdot [\beta_n s + (1-\beta_n)]^{N-X_0}$, ove α_n e β_n sono elementi della matrice

$$(**) \quad \begin{bmatrix} 1-\alpha_n & \alpha_n \\ \beta_n & 1-\beta_n \end{bmatrix} = \frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}$$

potenza n-ma della matrice (*) sopra specificata; si noti che i numeratori dei due coefficienti scalari sono gli autovalori di (*) e che il secondo membro della (**) corrisponde alla rappresentazione spettrale della matrice a primo membro.

Per una intuitiva giustificazione dell'espressione di $G_n(s)$ si supponga che sia $X_0 = j$ (j palline inizialmente presenti nell'urna I): il primo fattore del prodotto è allora $[(1-\alpha_n)s + \alpha_n]^j$ che è la funzione generatrice di una distribuzione binomiale $B(j, 1-\alpha_n)$, le cui probabilità riguardano le eventualità che un certo numero delle j palline inizialmente nell'urna I si trovino nella stessa urna dopo n passi. Il secondo fattore è dato da $[\beta_n s + (1-\beta_n)]^{N-j}$ ed è la funzione generatrice di $B(N-j, \beta_n)$ le cui probabilità riguardano le eventualità che un certo numero di palline dell'urna II si trovino nella I dopo n passi. I due fattori sono moltiplicati a causa dell'indipendenza delle traiettorie delle singole palline.

Al divergere di n la matrice (**) tende alla $\begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix}$ le cui righe esprimono la distribuzione stazionaria

di equilibrio (se α e β sono entrambi diversi da 0 e da 1). In corrispondenza, la funzione generatrice $G_n(s)$ tende alla:

$$G_n(s) = \left(\frac{\beta}{\alpha + \beta} \cdot s + \frac{\alpha}{\alpha + \beta} \right)^N,$$

che è la funzione generatrice della distribuzione binomiale $B(N, \frac{\beta}{\alpha + \beta})$.

Cenni sui modelli lineari dinamici stocastici a tempo discreto

1) Rappresentazioni per modelli lineari dinamici deterministici

I modelli lineari sono la categoria più semplice di modelli dinamici e ciò spiega la loro ampia utilizzazione nelle applicazioni. Esiste per essi una teoria sufficientemente generale della quale tenteremo di fornire gli elementi introduttivi. Bisogna subito distinguere i modelli a tempo discreto da quelli a tempo continuo: i primi sono tipicamente impiegati nella rappresentazione dei fenomeni economici; i secondi in quella dei fenomeni fisici. I tipi principali di modelli matematici lineari per l'analisi dinamica sono i seguenti:

- 1) equazioni (e sistemi di equazioni) differenziali e alle differenze finite;
- 2) modelli nello spazio degli stati (o markoviani);
- 3) modelli input – output (definiti in termini della funzione di trasferimento).

Per prima cosa tratteremo dei modelli lineari dinamici **a tempo discreto** per i quali, con riferimento all'elenco precedente, il primo tipo di rappresentazione è costituito da equazioni alle differenze finite, lineari, con coefficienti costanti:

$$y_t - \sum_{j=1}^p a_j \cdot y_{t-j} = u_t, \quad t = 1, 2, \dots, \dots, \dots,$$

ove $\{u_t\}$ è una successione nota ed i coefficienti a_j sono fissati, mentre $\{y_t\}$ è la successione incognita che va determinata risolvendo l'equazione dopo aver fissato una "condizione iniziale", per esempio la sequenza iniziale (y_1, \dots, \dots, y_p) della $\{y_t\}$.

La seconda rappresentazione, corrispondente alla precedente, prevede l'introduzione delle "variabili di stato" $x_t^{(h)}$, $h = 1, \dots, p$, per il cui vettore $\mathbf{x}_t = [x_t^{(1)}, \dots, x_t^{(p)}]^T$ sussistono le equazioni seguenti:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F} \cdot \mathbf{x}_{t-1} + \mathbf{b} \cdot u_t \\ y_t &= \mathbf{h}^T \cdot \mathbf{x}_t \end{aligned}$$

ove \mathbf{F} è una matrice quadrata di dimensione (p,p) , \mathbf{b} ed \mathbf{h} sono vettori colonna p -dimensionali. La scelta delle variabili di stato, per uno stesso modello, non è unica; nel nostro caso una scelta possibile è la seguente: $x_t^{(1)} = y_t, \dots, x_t^{(p)} = y_p$. In corrispondenza, si ha:

$$\mathbf{F} = \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_p \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

In questa seconda rappresentazione, l'equazione alle differenze lineare di ordine p della prima è trasformata in una equazione vettoriale alle differenze, lineare, del primo ordine.

Il terzo tipo di rappresentazione, molto usato nelle scienze ingegneristiche, si ottiene applicando la “trasformazione Z” ad entrambi i membri della prima equazione alle differenze ottenendo l’espressione

$$Y(z) = H(z)U(z)$$

ove $H(z)$ rappresenta la “funzione di trasferimento” del sistema; essa ha la seguente espressione in termini dei coefficienti dell’equazione alle differenze:

$$H(z) = (1 - \sum_{j=1}^p a_j z^{-j})^{-1} .$$

La denominazione di rappresentazione “input – output” corrisponde all’interpretazione di $U(z)$ come input di un “trasformatore”, descritto dalla funzione di trasferimento $H(z)$, e a quella di $Y(z)$ come output dello stesso.

Dal momento che nel seguito non impiegheremo questa rappresentazione ci limitiamo a dire che la “trasformazione Z” sostituisce alle successioni numeriche $\{y_t; t \geq 0\}$ e $\{u_t; t \geq 0\}$ le funzioni di

variabile complessa $Y(z) = \sum_{j=0}^{\infty} y_j z^{-j}$ e $U(z) = \sum_{j=0}^{\infty} u_j z^{-j}$

Esempio: l’equazione della produzione aggregata nel modello di P.A. Samuelson.

Prendiamo in considerazione il modello economico di P.A. Samuelson costituito dalle seguenti equazioni:

$$C_t = \beta \cdot Y_{t-1}, \quad I_t = \gamma \cdot (C_t - C_{t-1}), \quad Y_t = C_t + I_t + G_t,$$

ove Y_t, C_t e I_t (produzione, consumo e investimento aggregati) sono variabili endogene, mentre G_t (spesa della pubblica amministrazione) è l’unica variabile esogena del modello. La terza equazione, con le sostituzioni indicate dalle

$$Y_t = C_t + I_t + G_t = \beta \cdot Y_{t-1} + \{ \gamma [(\beta \cdot Y_{t-1}) - (\beta \cdot Y_{t-2})] \} + G_t = \beta \cdot (1 + \gamma) Y_{t-1} - \beta \cdot \gamma \cdot Y_{t-2} + G_t ,$$

costituisce un’equazione alle differenze finite del secondo ordine, lineare, a coefficienti costanti nella successione incognita dei livelli della produzione aggregata $\{Y_t\}$.

A questa prima rappresentazione corrisponde la seguente rappresentazione in termini delle variabili di stato $X_t^{(1)} = Y_t$ e $X_t^{(2)} = Y_{t-1}$, componenti del vettore \mathbf{X}_t :

$$\mathbf{X}_t = \begin{bmatrix} \beta \cdot (1 + \gamma) & - \beta \cdot \gamma \\ 1 & 0 \end{bmatrix} \cdot \mathbf{X}_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot G_t ,$$

$$Y_t = (1 \ 0) \cdot \mathbf{X}_t .$$

Affermiamo che la risoluzione del suddetto sistema di equazioni fornisce, se le condizioni iniziali poste nei due casi sono coerenti, la stessa soluzione della precedente equazione alle differenze del secondo ordine.

2) Rappresentazioni per modelli lineari dinamici stocastici

Quando qualche elemento del modello dinamico, uno o piu' coefficienti oppure l'input $\{u_t\}$ oppure la condizione iniziale (o anche piu' elementi tra questi) non e' noto con certezza allora si parla di **modello dinamico stocastico**. L'elemento non completamente noto dev'essere allora caratterizzato in senso probabilistico e cio' implica che anche la soluzione $\{y_t\}$ possa essere caratterizzata solo probabilisticamente.

Se, per esempio, l'input $\{u_t\}$ del modello non e' noto con certezza esso va considerato un processo stocastico $\{U_t\}$ e definito probabilisticamente sulla base delle informazioni disponibili: se queste consentono di specificare soltanto la funzione valor medio $\phi_U(t) = \{E(U_t); t \geq 1\}$ e la funzione di covarianza $\psi_U(s,t) = \{Cov[U_s, U_t]; s, t \geq 1\}$ di $\{U_t\}$ anche la specificazione del processo stocastico output $\{Y_t\}$ sara' limitata alle funzioni $\phi_Y(t)$ e $\psi_Y(s,t)$. Ovviamente, il livello di specificazione probabilistica dell'output non puo' essere piu' elevato di quello dell'input.

Le rappresentazioni gia' menzionate nel paragrafo precedente diventano ora, se l'input e' un processo stocastico $\{U_t\}$:

- 1) un'equazione alle differenze finite, lineare, stocastica

$$Y_t - \sum_{j=1}^p a_j Y_{t-j} = U_t, \quad t = 1, 2, \dots$$

e, rispettivamente,

- 2) un sistema di due equazioni lineari stocastiche

$$\mathbf{X}_t = \mathbf{F} \cdot \mathbf{X}_{t-1} + \mathbf{b} \cdot U_t,$$

$$Y_t = \mathbf{h}^T \cdot \mathbf{X}_t.$$

La risoluzione del modello in ciascuna delle due forme precedenti, fissata che sia la condizione iniziale, non fornisce la corrispondente successione dei livelli per l'output, ma la corrispondente successione dei livelli aleatori $\{Y_t; t \geq 1\}$ con un livello di specificazione probabilistica non superiore a quello del processo stocastico input $\{U_t; t \geq 1\}$. Un semplice esempio si ottiene supponendo che $\{U_t; t \geq 1\}$ sia un processo White Noise $(0; \sigma_U^2)$: l'equazione 1) o il corrispondente sistema 2) caratterizzano il processo $\{Y_t; t \geq 1\}$ come un processo AR(p) la cui eventuale stazionarieta' asintotica dipende dalla ben nota condizione che le radici dell'equazione caratteristica $\lambda^t - \sum_{j=1}^p a_j \lambda^{t-j} = 0$ siano tutte minori di 1 in modulo.

Esempio: la produzione aggregata nella versione stocastica del modello di P.A. Samuelson

Le tre note equazioni ora diventano

$$C_t = \beta \cdot Y_{t-1} + \varepsilon_t, \quad I_t = \gamma \cdot (C_t - C_{t-1}) + \eta_t, \quad Y_t = C_t + I_t + G_t,$$

ove le nuove variabili ε_t e η_t sono "perturbazioni aleatorie" che riassumono l'influenza su C_t e, rispettivamente, su I_t di tutti quei fattori economici e di altro tipo (sociologici, politici,.....) che nel modello sono trascurati. Molto spesso si assume che le perturbazioni aleatorie siano

processi stocastici di tipo White Noise, $\varepsilon_t \approx WN(0, \sigma_\varepsilon^2)$ e $\eta_t \approx WN(0, \sigma_\eta^2)$, eventualmente correlati tra loro. L'equazione per la produzione aggregata diventa

$$Y_t - \beta \cdot (1 + \gamma) \cdot Y_{t-1} + \beta \cdot \gamma \cdot Y_{t-2} = G_t + U_t,$$

ove $U_t = (1 + \gamma) \cdot \varepsilon_t - \gamma \cdot \varepsilon_{t-1} + \eta_t$, cioè un'equazione alle differenze finite, lineare, stocastica la cui soluzione fornisce il processo stocastico $\{Y_t; t \geq 1\}$. Evidentemente è $E(U_t) \equiv 0$ per cui, dalla precedente equazione, si ottiene

$$E(Y_t) - \beta \cdot (1 + \gamma) \cdot E(Y_{t-1}) + \beta \cdot \gamma \cdot E(Y_{t-2}) = E(G_t)$$

che è un'equazione alle differenze deterministica nella successione incognita $\phi_Y(t)$ dei valori medi dei livelli della produzione aggregata. È anche facile ottenere la funzione di covarianza, $\psi_Y(s, t)$, del processo $\{Y_t; t \geq 1\}$ il che fornisce per quest'ultimo una specificazione del secondo ordine.

Inferenza statistica sulle variabili di stato

Siamo infine arrivati a trattare l'argomento centrale della nostra esposizione. Il modello nello spazio degli stati è stato introdotto nel paragrafo precedente nella forma

$$\mathbf{X}_t = \mathbf{F} \cdot \mathbf{X}_{t-1} + \mathbf{b} \cdot U_t,$$

$$Y_t = \mathbf{h}^T \cdot \mathbf{X}_t,$$

che costituisca una rappresentazione equivalente dell'equazione alle differenze stocastica

$$Y_t - \sum_{j=1}^p a_j \cdot Y_{t-j} = U_t.$$

Nel seguito considereremo la seguente generalizzazione dello schema suddetto:

$$\mathbf{X}_t = \mathbf{F} \cdot \mathbf{X}_{t-1} + \mathbf{U}_t,$$

$$\mathbf{Y}_t = \mathbf{H} \cdot \mathbf{X}_t + \mathbf{V}_t,$$

ove la prima equazione, detta "equazione di evoluzione o di stato", è un'equazione alle differenze lineare e stocastica nella successione dei vettori di stato $\{\mathbf{X}_t\}$, mentre la seconda equazione, detta "di osservazione", mette in relazione il vettore di stato con il vettore delle variabili osservabili \mathbf{Y}_t . I vettori \mathbf{U}_t e \mathbf{V}_t sono perturbazioni aleatorie vettoriali di tipo White Noise non correlate tra loro, cioè si assume che sia $E(\mathbf{U}_t) = \mathbf{0}$, $E(\mathbf{V}_t) = \mathbf{0}$ e $\text{Cov}(\mathbf{U}_t) \equiv \Sigma_U$, $\text{Cov}(\mathbf{V}_t) \equiv \Sigma_V$, $\text{Cov}(\mathbf{U}_t, \mathbf{V}_s) \equiv [\mathbf{0}]$. Nella più semplice di tali generalizzazioni tutti i vettori hanno le stesse dimensioni e le matrici quadrate \mathbf{F} ed \mathbf{H} , costanti nel tempo, sono note, come anche le matrici di dispersione Σ_U e Σ_V .

Il problema inferenziale che considereremo, detto anche problema di "filtraggio", consiste nella stima puntuale lineare del vettore di stato non osservabile \mathbf{X}_t sulla base della conoscenza dei vettori $\mathbf{Y}_1, \dots, \mathbf{Y}_t$. Il metodo di stima, o di approssimazione, che adotteremo è quello dei minimi quadrati; ricordiamo al lettore che tale metodo richiede soltanto la specificazione del secondo ordine dei processi stocastici coinvolti nel problema.

1) Approssimazioni lineari dei minimi quadrati per numeri aleatori

Si consideri lo spazio lineare H_0 (di dimensione infinita) dei numeri aleatori X, Y, Z, \dots che supponiamo dotati di speranza matematica nulla e momento secondo finito. Sia Y il n.a. di interesse per il quale si voglia costruire una stima (o previsione o approssimazione) in termini di una qualche funzione $\varphi(\mathbf{X})$ dei n.a. X_1, \dots, X_n costituenti il vettore aleatorio \mathbf{X} .

Se non si introducono vincoli particolari per la funzione stimatore $\varphi(\mathbf{X})$, eccetto quello $E[\varphi(\mathbf{X})]^2 < \infty$, si prova che la funzione ottimale, nel senso dei minimi quadrati, è la funzione di regressione $E(Y/\mathbf{X})$; formalmente, per ogni ammissibile funzione $\varphi(\cdot)$ si ha:

$$E[Y - E(Y/\mathbf{X})]^2 \leq E[Y - \varphi(\mathbf{X})]^2.$$

Se per $\varphi(\mathbf{X})$ si impone il vincolo di **linearità**, cioè se si assume che $\varphi(\mathbf{X})$ sia una funzione lineare, $\sum_{j=1}^n \hat{\alpha}_j X_j$, dei n.a. X_1, \dots, X_n , si devono trovare gli n coefficienti $\hat{\alpha}_j, j = 1, \dots, n$, per i quali risulta:

$$E\left[Y - \sum_j \hat{\alpha}_j X_j\right]^2 \leq E\left[Y - \sum_j \alpha_j X_j\right]^2$$

in corrispondenza ad ogni n-pla $\alpha_1, \dots, \alpha_n$ di numeri reali. Si dimostra facilmente che il vettore dei coefficienti ottimali $\hat{\underline{\alpha}}$ è dato da $\hat{\underline{\alpha}} = [\text{Cov}(\mathbf{X})]^{-1} \cdot E(\mathbf{Y} \cdot \mathbf{X})$ sotto la condizione che la matrice di dispersione di \mathbf{X} sia invertibile (il che accade se i n.a. X_j sono linearmente indipendenti).

Per dimostrarlo si tratta di porre uguali a 0 le derivate parziali rispetto ad ogni α_j di

$E\left[Y - \sum_j \alpha_j X_j\right]^2$: il sistema lineare che si ottiene ha l'espressione $\text{Cov}(\mathbf{X}) \cdot \hat{\underline{\alpha}} = E(\mathbf{Y} \cdot \mathbf{X})$ e la sua soluzione è unica se $\text{Cov}(\mathbf{X})$ è invertibile. Il previsore (o stimatore o approssimatore) lineare ottimale per Y è allora $\hat{Y} = (X_1, \dots, X_n) \cdot [\text{Cov}(\mathbf{X})]^{-1} \cdot E(\mathbf{Y} \cdot \mathbf{X})$.

Tale numero aleatorio ha un'interessante interpretazione geometrica: \hat{Y} coincide con la **proiezione ortogonale** $P(Y/L)$ di Y sul sottospazio lineare L di H_0 generato dai n.a. X_1, \dots, X_n .

Per attribuire un significato preciso a tale proposizione è necessario introdurre le seguenti definizioni: in H_0 la **lunghezza** (o norma) del vettore geometrico associato ad un n.a. Z è definita dalla $\|Z\| = [\text{Var}(Z)]^{1/2}$; la **distanza** tra due n.a. Z e V è definita dalla $d(Z, V) = [\text{Var}(Z - V)]^{1/2}$; la **condizione di ortogonalità** tra Z e V è espressa dalla $E(Z \cdot V) = 0$ (poiché $E(Z) = E(V) = 0$, Z e V sono ortogonali se $\text{Cov}(Z, V) = 0$).

La suddetta interpretazione geometrica di \hat{Y} si consegue applicando il "principio di ortogonalità" il cui enunciato in questo caso stabilisce che **considerato un qualunque n.a. Y di H_0 , il n.a. di L a minima distanza da esso coincide con la proiezione ortogonale di Y su L** quando si osservi che

$E \left[Y - \sum_j \alpha_j X_j \right]^2$ esprime il quadrato della distanza tra Y ed il generico elemento del sottospazio L .

Importa osservare che mentre l'utilizzazione dell'approssimatore ottimale di Y , costituito dalla funzione di regressione $E(Y/X)$, richiede la conoscenza della distribuzione congiunta $F(y, x_1, \dots, x_n)$ dei n.a. considerati, o almeno della distribuzione subordinata $F(y/x_1, \dots, x_n)$, la costruzione dell'approssimatore lineare ottimale di Y , costituito dal n.a. $\hat{Y} = P(Y/L)$, richiede la conoscenza (o meglio la specificazione) dei soli momenti del primo e secondo ordine dei n.a. Y, X_1, \dots, X_n . Una seconda osservazione rilevante è che per le distribuzioni implicanti una funzione di regressione $E(Y/X)$ lineare negli elementi di \mathbf{X} accade che gli approssimatori $E(Y/X)$ e $\hat{Y} = P(Y/L)$ coincidono; le distribuzioni più note aventi questa caratteristica sono quella normale e quella Student - t multivariate.

2) Proprietà iterativa dell'approssimazione lineare dei minimi quadrati

Si considerino tre spazi lineari tali che sia $L_{n-1} \subset L_n \subset H_0$, ove l'ultimo spazio (di dimensione infinita) contiene tutti i numeri aleatori (n.a.) equi e dotati di varianza finita, mentre $L_{n-1} = L(X_1, \dots, X_{n-1})$ ed $L_n = L(X_1, \dots, X_n)$, essendo i n.a. X_j linearmente indipendenti ed osservabili. Sia $Y \in H$ il n.a. non osservabile per il quale si vogliono determinare le approssimazioni lineari dei minimi quadrati basate sul processo osservabile $\{X_j; j = 1, 2, \dots\}$.

Proveremo la seguente **proprietà iterativa**:

$$\hat{Y}_n = P(Y/L_n) = P(Y/L_{n-1}) + P[Y/X_n - P(X_n/L_{n-1})],$$

ove la differenza $X_n - P(X_n/L_{n-1})$ è il n.a. "innovazione di X_n " che risulta ortogonale a L_{n-1} .

Indicando per brevità con Y_n^* la differenza $Y - \hat{Y}_n$, sussistono evidentemente le due decomposizioni ortogonali $Y = \hat{Y}_{n-1} + Y_{n-1}^*$ e $Y = \hat{Y}_n + Y_n^*$ ed inoltre si ha $\hat{Y}_n = Z_1 + Z_2$, ove $Z_1 = P(\hat{Y}_n/L_{n-1}) = \hat{Y}_{n-1}$ perché $P(\hat{Y}_n/L_{n-1}) = P[P(Y/L_n)/L_{n-1}] = P(Y/L_{n-1})$ per una nota proprietà del proiettore ortogonale. Si ha dunque:

$$Y = \hat{Y}_n + Y_n^* = (Z_1 + Z_2) + Y_n^* = \hat{Y}_{n-1} + (Z_2 + Y_n^*)$$

cosicché dev'essere $Z_2 + Y_n^* = Y_{n-1}^*$ la quale implica che sia $\hat{Y}_n = \hat{Y}_{n-1} + Z_2$.

Rimane da provare che $Z_2 = P[Y/X_n - P(X_n/L_{n-1})] = P(Y/X_n - \hat{X}_{n/n-1})$ e ciò si ottiene confrontando la precedente uguaglianza con la seguente sequenza:

$$\begin{aligned} \hat{Y}_n &= P(Y / X_1, \dots, X_{n-1}, X_n) = P[Y / X_1, \dots, X_{n-1}, X_n - X_{n/n-1}] = \\ &= P(Y / X_1, \dots, X_{n-1}) + P(Y / X_n - \hat{X}_{n/n-1}) = \hat{Y}_{n-1} + P(Y / X_n - \hat{X}_{n/n-1}). \end{aligned}$$

3) Altre proprietà del proiettore ortogonale

$$P[P(X/L_n)/L_n] = P(X/L_n)$$

$$P(a_1 X_1 + a_2 X_2 / L_n) = a_1 P(X_1) + a_2 P(X_2),$$

$\langle P(X/L_n), Y \rangle = \langle X, P(Y/L_n) \rangle$, ove $\langle X, Y \rangle = E(X \cdot Y)$ indica il prodotto interno tra X e Y,

$$P[P(X/L_{n-1})/L_n] = P[P(X/L_n)/L_{n-1}] = P(X/L_{n-1}),$$

$$P(X/Y_1, Y_2, \dots, Y_n) = \sum_{j=1}^n P(X/Y_j) \text{ se i n.a. } Y_j \text{ sono due a due ortogonali.}$$

4) Filtro di R. E. Kalman a tempo discreto

Assumiamo di essere interessati ai n.a. di un processo stocastico $\{Y_n; n = 1, 2, \dots\}$, non osservabile, generato da un modello stocastico lineare espresso dalle equazioni alle differenze finite

$$(1) \quad Y_{n+1} = a Y_n + V_n, \text{ ove } V_n \approx WN(0; \sigma_v^2) \text{ e ove } V_n \perp Y_m, \forall m \leq n.$$

Si supponga di poter osservare i n.a. X_n definiti dalla trasformazione lineare affine e stocastica

$$(2) \quad X_n = b Y_n + W_n, \text{ ove } W_n \approx WN(0; \sigma_w^2) \text{ e ove } W_n \perp V_m, \forall n, m.$$

Il modello stocastico espresso dalle (1) e (2) è denominato “modello lineare dinamico” o anche “modello nello spazio degli stati” in quanto le variabili Y_n sono indicate come “variabili di stato”.

Ci porremo il problema di determinare le **approssimazioni lineari dei minimi quadrati** $\hat{Y}_{n/n}$ per le variabili Y_n basate sull'osservazione delle (X_1, \dots, X_n) ; in letteratura questo problema è noto come “problema di filtraggio”. Il n.a. $\hat{Y}_{n/n}$ è definito dalle condizioni $E(Y_n - \hat{Y}_{n/n})^2 \leq$

$E(Y_n - \sum_{j=1}^n \alpha_j X_j)^2$ per ogni n-pla di coefficienti reali $(\alpha_1, \dots, \alpha_n)$ ed è anche indicato con il

simbolo $P(Y_n / L_n)$ a causa della sua interpretazione geometrica quale “proiezione ortogonale di Y_n sullo spazio lineare L_n generato dai n.a. osservabili X_1, \dots, X_n ”. Ricordiamo infine al lettore

che se si assume che tutti i n.a. considerati abbiano valor medio nullo e varianze finite e si pone $\underline{X} = (X_1, \dots, X_n)^T$ allora $\hat{Y}_{n/n}$ è dato dalla

$$(3) \quad \hat{Y}_{n/n} = P(Y_n / L_n) = E(Y_n \underline{X}) \cdot (\text{Cov } \underline{X})^{-1} \cdot \underline{X}.$$

Ci proponiamo di presentare il procedimento ricorsivo di approssimazione lineare noto come “filtro di Kalman”, apparso in letteratura nel 1960; esso è costituito dalle seguenti equazioni, ove si è indicato in generale con $P_{n/m}$ la varianza $E(Y_n - \hat{Y}_{n/m})^2$ dell'errore di approssimazione $Y_n - \hat{Y}_{n/m}$:

$$(4) \quad \hat{Y}_{n+1/n} = a \cdot \hat{Y}_{n/n} ,$$

$$(5) \quad P_{n+1/n} = E(Y_{n+1} - \hat{Y}_{n+1/n})^2 = a^2 \cdot P_{n/n} + \sigma_V^2 ,$$

$$(6) \quad \hat{Y}_{n+1/n+1} = \hat{Y}_{n+1/n} + \frac{b \cdot P_{n+1/n}}{b^2 \cdot P_{n+1/n} + \sigma_W^2} \cdot (X_{n+1} - b \cdot \hat{Y}_{n+1/n}) ,$$

$$(7) \quad P_{n+1/n+1} = \frac{\sigma_W^2 \cdot P_{n+1/n}}{b^2 \cdot P_{n+1/n} + \sigma_W^2} .$$

Le precedenti equazioni presuppongono che si siano già determinati ricorsivamente $\hat{Y}_{n/n}$ e $P_{n/n}$ a partire dalle fissate posizioni iniziali $\hat{Y}_{0/0}$ e $P_{0/0}$; le (4) e (5) sono dette “equazioni di previsione” mentre le (6) e (7) sono dette “equazioni di aggiornamento” cosicché ogni stadio del procedimento ricorsivo consiste di una fase di previsione e una di aggiornamento.

Forniremo ora una dimostrazione delle precedenti equazioni:

- la (4) si prova in base alla proprietà di linearità del proiettore ortogonale perché risulta

$$\hat{Y}_{n+1/n} = P(Y_{n+1} / L_n) = P(a \cdot Y_n + V_n / L_n) = a \cdot P(Y_n / L_n) + P(V_n / L_n) = a \cdot \hat{Y}_{n/n}$$

in quanto $P(V_n / L_n) = 0$ giacché si è supposto $V_n \perp Y_m, \forall m \leq n$, e $W_n \perp V_m, \forall n, m$;

- la (5) si prova utilizzando il risultato precedente in quanto è:

$$P_{n+1/n} = E(Y_{n+1} - \hat{Y}_{n+1/n})^2 = E[a \cdot (Y_n - \hat{Y}_{n/n}) + V_n]^2 = a^2 \cdot P_{n/n} + \sigma_V^2 ;$$

- la (6) si dimostra utilizzando la proprietà iterativa del proiettore ortogonale che scriveremo indicando con X_{n+1}^* l'innovazione $X_{n+1} - \hat{X}_{n+1/n}$:

$$\hat{Y}_{n+1/n+1} = \hat{Y}_{n+1/n} + P(Y_{n+1} / X_{n+1}^*) = \hat{Y}_{n+1/n} + \frac{\text{Cov}(Y_{n+1}, X_{n+1}^*)}{\text{Var}(X_{n+1}^*)} \cdot X_{n+1}^* = \hat{Y}_{n+1/n} + \frac{b \cdot P_{n+1/n}}{b^2 \cdot P_{n+1/n} + \sigma_W^2} \cdot X_{n+1}^*$$

e l'ultima espressione coincide con la (6) se si tiene presente che è

$$X_{n+1}^* = X_{n+1} - P(X_{n+1} / L_n) = X_{n+1} - P(b \cdot Y_{n+1} + W_{n+1} / L_n) = X_{n+1} - b \cdot P(Y_{n+1} / L_n) = X_{n+1} - b \cdot \hat{Y}_{n+1/n} ;$$

- la (7) si prova in base alla definizione di $P_{n+1/n+1}$ e ricorrendo a semplici, anche se noiose sostituzioni:

$$P_{n+1/n+1} = E(Y_{n+1} - \hat{Y}_{n+1/n+1})^2 = \frac{\sigma_W^2 \cdot P_{n+1/n}}{b^2 \cdot P_{n+1/n} + \sigma_W^2} .$$

Per completezza, riporteremo ora le equazioni della versione vettoriale del filtro di Kalman che riguarda la stima lineare ricorsiva di un processo stocastico vettoriale $\{\mathbf{Y}_n; n = 1, 2, \dots\}$ basata sull'osservabilità del processo vettoriale $\{\mathbf{X}_n; n = 1, 2, \dots\}$. Le due equazioni del modello lineare dinamico sono ora espresse dalle:

$$\mathbf{Y}_{n+1} = \mathbf{A} \cdot \mathbf{Y}_n + \mathbf{V}_n \quad \text{e} \quad \mathbf{X}_n = \mathbf{B} \cdot \mathbf{Y}_n + \mathbf{W}_n ;$$

in esse i due processi vettoriali di rumore sono ancora mutuamente non correlati e del tipo White Noise: $\mathbf{V}_n \sim \text{WN}(\mathbf{0}; \mathbf{Q})$ e $\mathbf{W}_n \sim \text{WN}(\mathbf{0}; \mathbf{R})$.

Supponendo noti $\hat{Y}_{n/n}$ e la matrice di momenti secondi $P_{n/n} = E[(Y_n - \hat{Y}_{n/n}) \cdot (Y_n - \hat{Y}_{n/n})^T]$, le equazioni del filtro di Kalman vettoriale sono espresse dalle:

$$(4') \quad \hat{Y}_{n+1/n} = \mathbf{A} \cdot \hat{Y}_{n/n} ,$$

$$(5') \quad P_{n+1/n} = \mathbf{A} \cdot P_{n/n} \cdot \mathbf{A}^T + \mathbf{Q} ,$$

$$(6') \quad \hat{Y}_{n+1/n+1} = \hat{Y}_{n+1/n} + P_{n+1/n} \cdot \mathbf{B}^T (\mathbf{B} \cdot P_{n+1/n} \cdot \mathbf{B}^T + \mathbf{R})^{-1} \cdot (X_{n+1} - \mathbf{B} \cdot \hat{Y}_{n+1/n}) ,$$

$$(7') \quad P_{n+1/n+1} = P_{n+1/n} - P_{n+1/n} \cdot \mathbf{B}^T (\mathbf{B} \cdot P_{n+1/n} \cdot \mathbf{B}^T + \mathbf{R})^{-1} \cdot \mathbf{B} \cdot P_{n+1/n} .$$

Per chi volesse approfondire queste prime nozioni suggeriamo, in ordine di completezza crescente della trattazione:

Bittanti S. – “Teoria della Predizione e del Filtraggio” (Pitagora Editrice Bologna, 2000) ,

Catlin D.E. – “Estimation, Control and the Discrete Kalman Filter” (Sprinter-Verlag, 1989) ,

Anderson B.D.O. – Moore J.B. – “Optimal Filtering” (PRENTICE-HALL, INC., 1979) .

Cenni sui modelli lineari dinamici stocastici a tempo continuo

1) Rappresentazioni per modelli lineari dinamici a tempo continuo

Dedicheremo ora qualche cenno ai modelli lineari dinamici a parametro continuo, usati sistematicamente nella Fisica, ma impiegati raramente in Economia matematica. Da qualche tempo, però, si assiste ad un loro uso massiccio nella Finanza matematica. I tipi principali di modelli lineari per l'analisi dinamica a tempo continuo sono i seguenti:

- a) equazioni (e sistemi di equazioni) differenziali,
- b) modelli input – output (definiti in termini della funzione di trasferimento);
- c) modelli nello spazio degli stati (o markoviani).

Il primo tipo di rappresentazione è costituito da equazioni differenziali ordinarie lineari con coefficienti costanti:

$$y^{(n)}(t) + \sum_{k=1}^n a_k \cdot y^{(n-k)}(t) = u(t) ,$$

ove $u(t)$ è una funzione nota (o un processo stocastico) ed i coefficienti a_k sono fissati, mentre $y(t)$ è la funzione deterministica (o stocastica) incognita.

Il secondo tipo di rappresentazione si ottiene applicando la “trasformazione di Laplace” ad entrambi i membri della precedente equazione differenziale ottenendo l'espressione

$$y(s) = H(s)u(s) ,$$

ove $H(s)$ rappresenta la “funzione di trasferimento” del sistema; essa ha la seguente espressione in termini dei coefficienti dell'equazione differenziale:

$$H(s) = \frac{1}{s^n + \sum_{k=1}^n a_k \cdot s^{n-k}} .$$

La denominazione di rappresentazione “input – output” corrisponde all'interpretazione di $u(s)$ come input del “trasformatore”, descritto dalla funzione di trasferimento $H(s)$, e a quella di $y(s)$ come output dello stesso. Non faremo uso di questa rappresentazione.

La terza rappresentazione, corrispondente alle due precedenti, prevede l'introduzione delle “variabili di stato” $x_h(t)$, $h = 1, \dots, n$, per il cui vettore $\mathbf{x}(t)$ sussistono le equazioni seguenti:

$$\begin{aligned} \frac{d}{dt} \mathbf{x}(t) &= \mathbf{F} \cdot \mathbf{x}(t) + \mathbf{b} \cdot u(t) , \\ y(t) &= \mathbf{h}^T \cdot \mathbf{x}(t) , \end{aligned}$$

ove \mathbf{F} è una matrice quadrata di dimensione n , \mathbf{b} ed \mathbf{h} sono vettori colonna n -dimensionali. La scelta delle variabili di stato, per uno stesso modello, non è unica; nel nostro caso una scelta possibile è la seguente: $x_1(t) = y(t)$, $x_2(t) = y'(t)$, $x_3(t) = y''(t)$, \dots , $x_n(t) = y^{(n-1)}(t)$. In corrispondenza, si ha:

$$F = \begin{bmatrix} 0 & 1 & 0 & \vdots & 0 \\ 0 & 0 & 1 & \vdots & 0 \\ 0 & 0 & 0 & \vdots & 0 \\ 0 & 0 & 0 & \vdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \vdots & -a_1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

In quest'ultima rappresentazione, l'equazione differenziale lineare di ordine n di partenza è trasformata in un sistema di equazioni differenziali lineari del primo ordine.

Esempio: oscillatore armonico.

Prendiamo in considerazione un punto materiale di massa m che si muove lungo l'asse x, attorno alla sua posizione di riposo, per effetto di una forza elastica di richiamo (esercitata per esempio da una molla) e che per effetto dell'attrito e/o della resistenza dell'aria tende progressivamente a fermarsi. Se indichiamo con $y(t)$ lo spostamento del punto dalla posizione di riposo, l'equazione del moto è data dalla

$$m \cdot y''(t) + a \cdot y'(t) + k \cdot y(t) = 0,$$

ove i coefficienti a e k dipendono dall'attrito e, rispettivamente, dalla forza di richiamo. Posto $a/m = \beta$ e $k/m = \omega^2$, l'equazione del moto viene espressa dalla seguente equazione differenziale lineare omogenea del secondo ordine:

$$y''(t) + \beta \cdot y'(t) + \omega^2 \cdot y(t) = 0.$$

Ponendo $y(t) = \exp(\lambda t)$ nella suddetta equazione differenziale e risolvendo l'equazione algebrica che ne risulta, e cioè $(\lambda^2 + \beta \lambda + \omega^2) \cdot \exp(\lambda t) = 0$, detta equazione caratteristica, si trovano le due radici $\lambda_1 = (-\beta + \sqrt{\beta^2 - 4\omega^2}) / 2$ e $\lambda_2 = (-\beta - \sqrt{\beta^2 - 4\omega^2}) / 2$ e con ciò la soluzione generale della equazione differenziale omogenea :

- 1) $y(t) = c_1 \cdot \exp(\lambda_1 t) + c_2 \cdot \exp(\lambda_2 t)$, se le due radici sono reali e distinte ;
- 2) $y(t) = (c_1 + c_2 t) \cdot \exp(-\beta t / 2)$, se $\lambda_1 = \lambda_2 = -\beta / 2$;
- 3) $y(t) = \exp\{-\beta t / 2\} \cdot [(c_1 + c_2) \cdot \cos b t + i \cdot (c_1 - c_2) \cdot \sin b t]$, se λ_1 e λ_2 sono complesse coniugate e ove $b = \frac{1}{2} \sqrt{4\omega^2 - \beta^2}$.

Chiaramente, i tre casi corrispondono al fatto che il discriminante $\beta^2 - 4\omega^2$ dell'equazione $\lambda^2 + \beta \lambda + \omega^2 = 0$ sia positivo, nullo o negativo. Nei primi due casi il moto è aperiodico perché l'attrito è sufficientemente grande da impedire oscillazioni; nel terzo caso il moto è oscillatorio e smorzato attorno al punto di riposo.

La rappresentazione nello spazio degli stati corrispondente all'equazione differenziale omogenea, ponendo $x_1(t) = y(t)$ e $x_2(t) = y'(t)$, è data dal sistema:

$$\frac{d}{dt} x(t) = \begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = F \cdot x(t) = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -\beta \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix},$$

$$y(t) = (1 \ 0) \cdot x(t).$$

Le radici λ_1 e λ_2 trovate precedentemente corrispondono agli autovalori della matrice F, cioè alle soluzioni dell'equazione $\det(\lambda I - F) = \lambda^2 + \beta \lambda + \omega^2 = 0$.

Nel caso che esista una funzione input deterministica uguale a $u(t) = f_0 \cos \Omega t$ la rappresentazione nello spazio degli stati diventa:

$$\frac{d}{dt} x(t) = \begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = F \cdot x(t) = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -\beta \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot f_0 \cos \Omega t ,$$

$$y(t) = (1 \ 0) \cdot x(t) .$$

Rinviamo il lettore agli innumerevoli manuali sull'argomento per la soluzione di questo sistema.

2) Modelli lineari dinamici stocastici a tempo continuo

Se la funzione a secondo membro, $u(t)$, dell'equazione differenziale lineare ordinaria

$$y^{(n)}(t) + \sum_{k=1}^n a_k \cdot y^{(n-k)}(t) = u(t)$$

non è deterministica, ma stocastica, cioè se **$u(t)$ rappresenta un processo stocastico a tempo continuo**, allora l'equazione descrive una trasformazione lineare di tale processo il cui risultato è un altro processo stocastico $y(t)$, soluzione dell'equazione differenziale.

Un primo semplice esempio di equazione differenziale lineare stocastica è fornito dall'equazione di Langevin, $v'(t) + b \cdot v(t) = u(t)$, $b > 0$, riguardante il fenomeno di "moto Browniano" di una qualunque microscopica particella sospesa in un fluido e soggetta ad innumerevoli urti da parte delle molecole del fluido stesso, a loro volta soggette all'agitazione termica. In tale equazione $v(t)$ indica la velocità aleatoria della particella, $u(t)$ il processo stocastico input del sistema definito sulla base di ragionevoli ipotesi sulle caratteristiche del fenomeno fisico, il coefficiente b il grado di viscosità del fluido. Imponendo la condizione iniziale $v(0) = V_0$ e risolvendo l'equazione si trova:

$$v(t) = V_0 \cdot e^{-bt} + \int_0^t e^{-b(t-s)} \cdot u(s) ds$$

che rappresenta il processo stocastico output.

Ritornando all'equazione differenziale generale $y^{(n)}(t) + \sum_{k=1}^n a_k \cdot y^{(n-k)}(t) = u(t)$, per il processo stocastico $u(t)$ deve essere fornita una qualche specificazione probabilistica: ci si può limitare a specificare la funzione valor medio $E[u(t)]$ e/o la funzione di covarianza $Cov[u(t), u(\tau)]$ oppure, all'estremo opposto, si può caratterizzare completamente $u(t)$ specificando la famiglia delle sue distribuzioni di probabilità congiunte. In corrispondenza del livello di specificazione prescelto per il processo input, dalla soluzione dell'equazione differenziale si otterrà lo stesso livello di specificazione per il processo output.

Nell'esempio dell'equazione di Langevin, se ci limitassimo a specificare $E[u(t)] \equiv 0$ si otterrebbe $E[v(t)] = E(V_0) \cdot e^{-bt} + \int_0^t e^{-b(t-s)} \cdot E[u(s)] ds = E(V_0) \cdot e^{-bt}$. Se oltre alla $E[u(t)] \equiv 0$ si fosse specificata anche la funzione di covarianza $Cov[u(t), u(\tau)]$ risulterebbe determinabile anche la funzione di covarianza del processo $v(t)$.

Particolarmente semplice risulta la specificazione probabilistica dei processi stocastici se si adotta, quando ciò è possibile, l'ipotesi di **stazionarietà** in quanto, com'è noto, tale ipotesi implica che la funzione valor medio del processo sia costante e che la sua funzione di covarianza dipenda da un solo argomento, l'ampiezza dell'intervallo temporale $t - \tau$. Ricordiamo anche che in questo caso alla funzione di covarianza può essere associata mediante la trasformazione di Fourier, in modo unico, la **funzione spettrale** $F_u(\omega)$ al modo seguente:

$$\text{Cov}[u(t), u(\tau)] = \psi_u(t - \tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega(t-\tau)} dF_u(\omega) .$$

Quando $F_u(\omega)$ riesce derivabile, la sua derivata $F_u'(\omega) = f_u(\omega)$ è chiamata **densità spettrale** e per essa sussistono le:

$$\psi_u(t - \tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega(t-\tau)} \cdot f_u(\omega) d\omega ,$$

$$f_u(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega\tau} \psi_u(\tau) d\tau .$$

Ricordiamo infine che se la specificazione probabilistica riguarda soltanto i momenti fino al secondo ordine, cioè se la specificazione concerne soltanto valori medi, varianze e covarianze dei numeri aleatori considerati, allora un processo stocastico stazionario $u(t)$ viene caratterizzato dalla scelta del comune valor medio e della funzione di covarianza $\psi_u(\tau)$ oppure da quella del comune valor medio e della funzione spettrale $F_u(\omega)$ (o densità spettrale $f_u(\omega)$); se si assume anche che tutte le distribuzioni congiunte del processo siano **di tipo normale** le scelte alternative precedenti specificano completamente il processo stocastico.

Nella rappresentazione in termini del vettore $\mathbf{x}(t)$ delle variabili di stato $\frac{d}{dt} \mathbf{x}(t) = F \cdot \mathbf{x}(t) + b \cdot u(t)$, $y(t) = h^T \cdot \mathbf{x}(t)$, si ha che anche il vettore $\mathbf{x}(t)$ è stocastico, essendo esso la soluzione di un sistema di equazioni differenziali stocastiche.

Presenteremo a questo punto due esemplificazioni di modelli lineari dinamici stocastici: la prima riguarda un modello fisico analogo, dal punto di vista matematico, all'oscillatore armonico e cioè un circuito elettrico RCL con input stocastico; la seconda esemplificazione, di notevolissimo interesse storico, concerne la stima di un processo di segnale basata sull'osservabilità della somma del segnale e di un disturbo casuale, cioè aleatorio.

a) Circuito RCL.

Il circuito elettrico è costituito da un resistore, un condensatore e un solenoide connessi in serie e alimentati da un generatore di tensione. Sono note le caratteristiche dei tre primi componenti e cioè la resistenza R (misurata in ohm) del resistore, la capacità C del condensatore (misurata in farad) e l'induttanza L del solenoide (misurata in henry). Ci porremo il seguente problema: cosa si può affermare sull'intensità di corrente $I(t)$ nel circuito se la tensione $V(t)$ del generatore applicato al circuito è solo parzialmente nota? Assumeremo che la tensione $V(t)$ sia un processo stocastico stazionario con funzione valor medio $\varphi_V(t)$ e funzione di covarianza $\psi_V(\tau)$ note.

Nel circuito scorre una corrente elettrica aleatoria $I(t)$ determinata dalla seguente equazione integro-differenziale

$$L I'(t) + R I(t) + \frac{1}{C} \int I(t) dt = V(t)$$

basata sulle leggi dell'Elettrotecnica.

Poiché la carica elettrica $Q(t)$ del condensatore è legata all'intensità di corrente $I(t)$ dalla relazione $I(t) = Q'(t)$ la precedente equazione può essere trasformata nella seguente equazione differenziale del secondo ordine

$$Q''(t) + 2\xi\omega \cdot Q'(t) + \omega^2 \cdot Q(t) = \frac{1}{L} \cdot V(t)$$

nella funzione (aleatoria) incognita $Q(t)$. Le costanti ξ e ω sono legate alle caratteristiche note degli elementi passivi del circuito dalle relazioni $\xi = \frac{R}{2} \cdot \sqrt{\frac{C}{L}}$ e $\omega = \frac{1}{\sqrt{CL}}$.

Ponendo $\mathbf{Z}(t) = [Q(t), Q'(t)]^T$ la precedente equazione scalare si trasforma nell'equazione vettoriale

$$\mathbf{Z}'(t) = \begin{bmatrix} Q'(t) \\ Q''(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\xi\omega \end{bmatrix} \cdot \begin{bmatrix} Q(t) \\ Q'(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{L} \cdot V(t) \end{bmatrix}.$$

Indicando con F la matrice del sistema, la soluzione dell'equazione differenziale vettoriale è data dalla

$$\mathbf{Z}(t) = e^{F \cdot t} \cdot \mathbf{Z}(0) + \int_0^t e^{F \cdot (t-s)} \cdot \begin{bmatrix} 0 \\ \frac{1}{L} \cdot V(s) \end{bmatrix} ds.$$

Gli autovalori λ_i , ($i = 1, 2$), di F sono le radici dell'equazione $\det(F - \lambda \cdot I) = 0$ ed hanno le espressioni $\lambda_{1,2} = -\xi \cdot \omega \pm \omega \cdot \sqrt{\xi^2 - 1}$. Assumendo che sia $\xi^2 < 1$ i due autovalori sono complessi coniugati

$$\lambda_{1,2} = -\xi \cdot \omega \pm i\omega \cdot \sqrt{1 - \xi^2}$$

e gli autovettori corrispondenti sono

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -\xi\omega + i\omega_1 \end{pmatrix}^T \quad \text{e} \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ -\xi\omega - i\omega_1 \end{pmatrix}^T$$

ove si è posto $\omega_1 = \omega \cdot \sqrt{1 - \xi^2}$; è facile constatare che \mathbf{v}_1 e \mathbf{v}_2 sono linearmente indipendenti.

Indicando con $V = (\mathbf{v}_1 \ \mathbf{v}_2)$ la matrice degli autovettori e ponendo $e^{\Lambda \cdot t} = \text{diag}(e^{\lambda_1 \cdot t}, e^{\lambda_2 \cdot t})$ si ha la rappresentazione $e^{F \cdot t} = V \cdot e^{\Lambda \cdot t} \cdot V^{-1}$ ed anche la $e^{F \cdot (t-s)} = V \cdot e^{\Lambda \cdot (t-s)} \cdot V^{-1}$ dalle quali, con un noioso ma

non difficile calcolo, si ottengono le seguenti soluzioni corrispondenti alla condizione iniziale $\mathbf{Z}(0) = \mathbf{0}$:

$$Q(t) = \frac{1}{\omega_1 \cdot L} \int_0^t e^{-\xi\omega(t-s)} \cdot \sin[\omega_1(t-s)] \cdot V(s) ds,$$

$$Q'(t) = I(t) = \frac{1}{L} \int_0^t \left\{ e^{-\xi\omega(t-s)} \cdot \left[\cos\omega_1(t-s) - \xi \cdot \frac{\omega}{\omega_1} \sin\omega_1(t-s) \right] \right\} \cdot V(s) ds.$$

Gli integrali stocastici che compaiono nelle due espressioni sono numeri aleatori definiti come limiti in media quadratica di opportune successioni di somme integrali. Ci limitiamo a dire che dalle funzioni $\varphi_V(t)$ e $\psi_V(\tau)$ del processo $\{V(t)\}$ si possono determinare le corrispondenti $\varphi_Q(t)$ e $\psi_Q(\tau)$ e $\varphi_I(t)$ e $\psi_I(\tau)$.

b) Stima lineare dei minimi quadrati di un processo stocastico a tempo continuo.

Indicheremo con $S(t)$ il processo stocastico di segnale che vogliamo stimare e con $U(t)$ un processo di disturbo (o di rumore) indesiderato ma non eliminabile che si sovrappone al primo; con $Y(t) = S(t) + U(t)$ indicheremo il processo di osservazione o generatore di dati. Assumeremo, per semplicità, che i due processi $S(t)$ e $U(t)$ non siano correlati tra loro e faremo le seguenti ipotesi sulle loro strutture stocastiche:

$$S(t) \sim N(0; \sigma_s^2 \cdot \exp\{-\alpha \cdot t\}), \quad U(t) \sim \text{NWN}(0; \sigma_u^2).$$

A parole, entrambi i processi sono assunti di tipo normale o Gaussiano: il primo, noto in letteratura come “processo di Ornstein-Uhlenbeck”, è un processo stazionario con funzione valor medio identicamente nulla e funzione di covarianza $\psi_s(\tau) = \sigma_s^2 \cdot \exp\{-\alpha \cdot \tau\}$; il secondo, noto in letteratura come processo “normal-white noise”, è anch'esso stazionario con variabili mutuamente indipendenti, aventi la comune varianza pari a σ_u^2 e funzione valor medio identicamente nulla.

Il modello introdotto non corrisponde apparentemente a nessuna delle tre rappresentazioni di modelli dinamici precedentemente introdotte; è però possibile fornire una equivalente rappresentazione nello spazio degli stati sostituendo alla suddetta specificazione probabilistica per il processo $\{S(t)\}$ una equazione differenziale lineare stocastica del primo ordine la cui soluzione è costituita dal processo $\{S(t)\}$:

$$\frac{d}{dt} S(t) + \alpha \cdot S(t) = \sigma \cdot W(t),$$

ove $W(t)$ è un processo normal - white noise con varianza unitaria. La rappresentazione completa nello spazio degli stati è allora la seguente:

$$\begin{aligned} \frac{d}{dt} S(t) + \alpha \cdot S(t) &= \sigma \cdot W(t), \\ Y(t) &= S(t) + U(t), \end{aligned}$$

essendo i due processi $W(t)$ e $U(t)$ non correlati tra loro. Si tratta della versione a tempo continuo del modello lineare dinamico già presentato nel paragrafo precedente. Il problema di stima dei minimi quadrati delle variabili di stato $S(t)$ sulla base delle osservazioni $\{Y(s); s \leq t\}$ si risolve utilizzando la versione a tempo continuo del filtro di E. Kalman (nota come “filtro di Kalman – Bucy”). Si dimostra che, indicando con $\hat{S}(t)$ lo stimatore lineare dei minimi quadrati per $S(t)$, esso è determinato dalla seguente equazione differenziale

$$\frac{d}{dt} \hat{S}(t) = -\alpha \cdot \hat{S}(t) + (\beta - \alpha) \cdot [Y(t) - \hat{S}(t)] = -\beta \cdot \hat{S}(t) + (\beta - \alpha) \cdot Y(t)$$

con la condizione iniziale $\hat{S}(0) = 0$ e ove si è posto $\beta = \sqrt{\frac{2\alpha \sigma_s^2}{\sigma_u^2} + \alpha^2}$.

Uno studio sistematico di questi temi può basarsi su:

- M.H.A. Davis – Linear estimation and Stochastic Control (Chapman and Hall, 1977),
- A.H. Jazwinski – Stochastic Processes and Filtering Theory (Academic Press, 1970),
- R.S.Liptser, A.N. Shiryaev – Statistics of Random Processes (Springer, 1978).

