

LUCIDI PER IL CORSO DI INFERENZA STATISTICA BAYESIANA

UN' INTRODUZIONE AI PROCESSI STOCASTICI (prof. A. Wedlin)

1) La nozione di processo stocastico

Dal punto di vista matematico, fissato uno spazio di probabilità $(\Omega, \mathcal{A}, P)$, un processo stocastico a valori reali 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 , $X(., t')$ è un numero aleatorio, cioè una funzione reale; al variare di t' in T si ha una famiglia di numeri aleatori;

2) per ogni fissato valore ω' di ω , in Ω , $X(\omega', .)$ è una funzione deterministica definita in T e denominata "traiettoria" del processo $X(t)$; al variare di ω' in Ω si ottiene la famiglia delle traiettorie per la "funzione aleatoria" $X(\omega, t)$.

2) Livello di specificazione del processo

Come per un numero aleatorio, anche per un processo stocastico si possono individuare più (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 $F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$ per ogni $n \geq 1$, $t_j \in T$ e per ogni $\underline{x}$. Per i processi a parametro discreto in cui $T = \mathbb{Z}^+$ 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 condizione di coerenza di Kolmogorov.

3) Teorema di A. N. Kolmogorov (1933)

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'equivalente formulazione di questo teorema richiede alle distribuzioni congiunte che soddisfino le due seguenti condizioni:

1. $F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{\pi(t_1, \dots, t_n)}[\pi(x_1, \dots, x_n)]$
ove π indica un'identica, ma arbitraria, permutazione delle due sequenze;
2. per $t_{n+1} \uparrow \infty$, $\lim F_{t_1, \dots, t_{n+1}}(x_1, \dots, x_{n+1}) = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$.

4) Specificazione del secondo ordine

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) $\sum_{s, t=1}^N a_s \cdot a_t \cdot \psi(s, t) \geq 0; \forall N, \mathbf{a} \in \mathbb{R}^N$.

Dimostriamo l'ultima proprietà con la seguente sequenza di uguaglianze:

$$0 \leq E \left[\sum_{t=1}^N a_t \cdot (X_t - EX_t) \right]^2 = E \left[\sum_{s=1}^N \sum_{t=1}^N a_s \cdot a_t \cdot (X_s - EX_s) \cdot (X_t - EX_t) \right] = \sum_{s, t} a_s a_t \psi(s, t)$$

5) Principali tipi di processi stocastici

Le principali categorie di processi stocastici sono i processi **markoviani**, quelli **stazionari** ed i processi **martingala**.

a) il processo $\{X_t; t \in \mathbb{Z}^+\}$ e' **markoviano** se per ogni scelta dell'intero n e della sequenza $t_1 < t_2 < \dots < t_n$ risulta $F(x_{t_n} / x_{t_1}, \dots, x_{t_{n-1}}) = F(x_{t_n} / x_{t_{n-1}})$;

b) il processo $\{X_t; t \in \mathbb{Z}^+\}$ e' **stazionario** se per ogni scelta dell'intero n e della sequenza $(t_1, t_2, \dots, t_n)$ risulta $F(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = F(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_n+h})$, ove h e' un qualunque intero tale che ogni $t_j + h \in \mathbb{Z}^+$;

c) il processo $\{X_t; t \in \mathbb{Z}^+\}$ e' una **martingala** se per ogni intero positivo $t \geq 1$ risulta $E[|X_t|] < \infty$ e $E(X_t / X_1, \dots, X_{t-1}) = X_{t-1}$.

6) 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**, ma (in generale) **non indipendenti**.

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 $\Psi_X(s, t) = \text{Cov}(X_s, X_t)$ dipende solo da $t - s$.

7) Teorema di H. Wold (1938)

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

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

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).

8) Analisi 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 che ora definiremo.

Teorema di G. Herglotz: se $\{X_t; t \geq 1\}$ è un processo stazionario in senso lato, a valori reali, con funzione valor medio identicamente nulla e funzione di covarianza $\Psi_X(h)$, $h \in \mathbb{Z}$, esiste una funzione (spettrale) $F(\lambda)$, a valori reali, monotona non decrescente e limitata, tale che $\Psi_X(h)$ ha la seguente rappresentazione

$$\Psi_X(h) = \int_{-\pi}^{\pi} \cos(\lambda h) dF(\lambda);$$

se esiste la derivata prima $f(\lambda) = F'(\lambda)$, detta densità spettrale, allora

$$\Psi_X(h) = \int_{-\pi}^{\pi} \cos(\lambda h) \cdot f(\lambda) d\lambda \quad \text{e} \quad f(\lambda) = (2\pi)^{-1} \cdot \sum_{h=-\infty}^{+\infty} \cos(\lambda h) \cdot \Psi_X(h).$$

9) Alcuni esempi.

- 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 $(0; \sigma^2)$, i cui n.a. sono equi, con la stessa varianza e non correlati, 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 è assunto positivo e non maggiore di π . E' immediato provare che $\varphi(t) = 0$ e si trova facilmente la $\psi(s, t) = \sigma^2 \cdot \cos \lambda(t-s)$. La funzione spettrale $F(\lambda)$ è una funzione a gradini con discontinuità in $-\lambda$ e λ e l'ampiezza dei salti è pari a $\sigma_X^2 / 2$.
- 3. Processo MA(1)**: $X(t) = \varepsilon_t + b \cdot \varepsilon_{t-1}$ con $\varepsilon_t \sim \text{WN}(0; \sigma_\varepsilon^2)$. Si ha $\Psi_X(0) = (1+b^2) \cdot \sigma_\varepsilon^2$, $\Psi_X(1) = b \cdot \sigma_\varepsilon^2$, mentre $\Psi_X(h) = 0$ per $h > 1$. Esiste una densità spettrale e la sua espressione è $f(\lambda) = \sigma_\varepsilon^2 (2\pi)^{-1} \cdot (1 + 2b \cdot \cos \lambda + b^2)$.

10) Altri esempi.

- 4. 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

$X_t = \varepsilon_t + \sum_{k=1}^{\infty} a^k \cdot \varepsilon_{t-k}$: si provi che questo processo è stazionario in senso lato e che

$f_X(\lambda) = \sigma_\varepsilon^2 \cdot (1 - 2a \cdot \cos \lambda + a^2)^{-1} / 2\pi$ è la sua densità spettrale.

- 5. 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), $X_t - a \cdot X_{t-1} = Y_t$, con $Y_t = \varepsilon_t + b \cdot \varepsilon_{t-1}$. Per quanto già visto negli esempi 3 e 4 si cerchi di giustificare la

densità spettrale del processo $X(t)$ espressa dalla $f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{(1 + 2b \cdot \cos \lambda + b^2)}{(1 - 2a \cdot \cos \lambda + a^2)}$

11) Cenni sui 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 $\{Y_t; t \geq 1\}$ con coefficienti c_j indipendenti da t e tali da soddisfare la condizione $\sum_{j=-\infty}^{+\infty} c_j^2 < \infty$.

Teorema: se $\{Y_t; t \geq 1\}$ è stazionario in senso lato allora $\{X(t)\}$ è stazionario in senso lato con funzione spettrale verificante la condizione $dF_X(\lambda) = |H(\lambda)|^2 \cdot dF_Y(\lambda)$, 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 $f_X(\lambda)$ verificante la $f_X(\lambda) = |H(\lambda)|^2 \cdot f_Y(\lambda)$.

Nell'esempio 3) 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 $\{\varepsilon_t\}$.

12) Processi stocastici scambiabili

Dal punto di vista formale, i processi scambiabili sono processi stazionari in senso stretto 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 $(x_1, \dots, x_n)$.

Tale definizione implica la $F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_n(x_1, \dots, x_n)$.

13) Continuazione

Per quanto concerne i momenti fino al secondo ordine si ha evidentemente

$$\varphi_X(t) = E(X_t) \equiv E(X_1) \quad \text{e} \quad \psi_X(s, t) = \text{Cov}(X_s, X_t) = \sigma^2 \quad \text{oppure} \quad = c > 0$$

a seconda che sia $s - t = 0$ oppure, rispettivamente, $s - t \neq 0$.

Dal punto di vista interpretativo la condizione di scambiabilità traduce l'idea di **analogia o equivalenza informativa** dei n.a. osservabili X_t (come anche la condizione i.i.d.) ma l'assunzione di scambiabilità introduce tipicamente una **dipendenza stocastica** tra essi consentendo così l'**apprendimento dall'osservazione statistica**. A parole, solo se i n.a. $X_1, \dots, X_n, \dots$ sono mutuamente dipendenti l'osservazione dei valori dei primi $n - 1$ di essi consente di precisare la nostra opinione su X_n e sui n.a. successivi.

14) Il teorema di rappresentazione (1928)

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

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

In questa espressione il simbolo ω rappresenta un parametro reale e non un evento elementare e Ω è l'insieme dei suoi valori possibili.

La combinazione lineare dei processi si realizza esprimendo la probabilità di un evento relativo ai n.a. scambiabili, nel nostro caso $P\left\{\bigcap_{j=1}^n (X_{t_j} \leq x_j)\right\} = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$, con la combinazione lineare delle probabilità dello stesso evento per n.a. assunti indipendenti e ugualmente distribuiti.

15) Processi scambiabili di indicatori di eventi

Se i n.a. X_t del processo scambiabile illimitato sono **indicatori di eventi** $|E_t|$ il teorema di rappresentazione ha una formulazione più semplice: **tutti e soli i processi scambiabili $\{E_t; t \geq 1\}$ sono combinazioni lineari (o misture) di processi Bernoulliani**. Precisamente, per ogni processo scambiabile di questo tipo esiste una funzione di ripartizione $G(x)$ nell'intervallo $[0, 1]$ che realizza la combinazione lineare suddetta di processi Bernoulliani: considerato, per esempio, l'evento

$$\left\{ \sum_{j=1}^n |E_j| = m \right\} \text{ si ha } P \left\{ \sum_{j=1}^n |E_j| = m \right\} = \int_0^1 \binom{n}{m} x^m (1-x)^{n-m} dG(x), \quad m \leq n.$$

16) Cenni di dimostrazione nel caso $m = n$.

Oltre al processo degli indicatori scambiabili $\{E_t; t \geq 1\}$ si considerano i processi corrispondenti delle frequenze di successo S_n e delle frequenze relative S_n/n . Si ha:

- a) al divergere di n , si ha $p - \lim S_n/n = X$, ove $E(X) = P(E_1)$,
- b) al divergere di n , le distribuzioni delle S_n/n convergono alla $G_X(x)$,
- c) per ogni intero h , le $E(S_n/n)^h$ convergono sia alla $P(S_h = h)$ che al momento h -mo della distribuzione limite $G_X(x)$.

Dalla $P(S_n = n) = \left\{ \sum_{j=1}^n |E_j| = n \right\} = \int_0^1 x^n dG(x)$ ora dimostrata segue facilmente il caso

$m < n$ considerato al numero **15**).

Si osservi che la a) esprime la legge debole dei grandi numeri per il processo scambiabile $\{E_t; t \geq 1\}$. La corrispondente legge per un processo Bernoulliano con indicatori i.i.d., cioè il teorema di Bernoulli, stabilisce la convergenza delle medie campionarie S_n/n al numero certo costituito dalla comune probabilità degli eventi.

17) Ancora sul processo scambiabile degli indicatori di eventi.

Circa la legge temporale del processo $\{E_t; t \geq 1\}$ affermiamo che essa è costituita dalle probabilità $\omega_n = P(S_n = n)$, $n = 0, 1, 2, \dots$ con $\omega_0 = 1$. La condizione di coerenza per le ω_n è detta “di completa monotonia” e consiste nelle disuguaglianze $(-1)^r \Delta^r \omega_n \geq 0$ per $r = 0, 1, 2, \dots$ e per $n = 0, 1, 2, \dots$, ove è

$$\Delta^r \omega_n = \sum_{j=0}^r \binom{r}{j} (-1)^{r-j} \omega_{n+j}.$$

Teorema di Hausdorff (1928): una successione numerica è completamente monotona se e solo se essa è la successione dei momenti di una distribuzione di probabilità sull'intervallo $[0, 1]$.

18) Cenni sui processi parzialmente scambiabili.

La condizione di scambiabilità parziale è più flessibile di quella della scambiabilità perché ammette **differenti gradi di analogia** tra i n.a. osservabili. Il modello più semplice è quello in cui le variabili osservabili $X(t)$ sono suddivise in due sottoprocessi $Y(t)$ e $Z(t)$ ciascuno dei quali è un processo scambiabile. Si assume anche che sussista una correlazione non nulla tra le variabili del primo e quelle del secondo processo. Lo schema è quindi il seguente:

- 1) processo scambiabile $Y(t)$,
- 2) processo scambiabile $Z(t)$,
- 3) $\text{Cov}[Y(s), Z(t)] \neq 0$.

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

$$F(y_{t_1}, \dots, y_{t_n}; z_{s_1}, \dots, z_{s_m}) = F(y_1, \dots, y_n; z_1, \dots, z_m)$$

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

19) Teorema di rappresentazione (B. de Finetti, 1937)

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 $F_1^{(\omega)}(y)$ comune a tutti i n.a. Y_n e una funzione di ripartizione $F_2^{(\eta)}(z)$ comune a tutti i n.a. Z_m , nel senso che esiste una distribuzione congiunta $G(\omega, \eta)$ tale che

$$F(y_1, \dots, y_n; z_1, \dots, z_m) = \int \int \left[\prod_{i=1}^n F_1^{(\omega)}(y_i) \right] \cdot \left[\prod_{j=1}^m F_2^{(\eta)}(z_j) \right] G(d\omega, d\eta)$$

per ogni coppia di interi non negativi (n, m) .

Anche qui, i simboli ω e η rappresentano parametri reali e non eventi elementari.

Per questo teorema non forniremo cenni di dimostrazione.

In un primo esempio, a tempo discreto, del teorema precedente in cui le variabili Y_i e Z_j sono indicatori di eventi $|E_i|$ ed $|H_j|$, sia $g(\omega, \eta)$, la densità congiunta corrispondente a $G(\omega, \eta)$, una densità di tipo Beta bivariata con parametri reali positivi ν_1, ν_2 e ν_3 , cioè

$$g(\omega, \eta) = \frac{\Gamma(\nu_1, \nu_2, \nu_3)}{\Gamma(\nu_1) \cdot \Gamma(\nu_2) \cdot \Gamma(\nu_3)} \omega^{\nu_1-1} \cdot \eta^{\nu_2-1} \cdot (1 - \omega - \eta)^{\nu_3-1}.$$

Si prova allora che i processi Y_i e Z_j riescono entrambi scambiabili con

$$P\left\{\sum_{i=1}^n Y_i = n\right\} = \frac{(\nu_1)_n}{(\nu_1 + \nu_2 + \nu_3)_n} \text{ e, rispettivamente, } P\left\{\sum_{i=1}^n Z_i = n\right\} = \frac{(\nu_2)_n}{(\nu_1 + \nu_2 + \nu_3)_n};$$

$$\text{si ha infine } P\left\{\left[\sum_{i=1}^n Y_i = n\right] \wedge \left[\sum_{j=1}^m Z_j = m\right]\right\} = \frac{(\nu_1)_n \cdot (\nu_2)_m}{(\nu_1 + \nu_2 + \nu_3)_{n+m}}.$$

Quest'ultima probabilità è quella che la frequenza di successo su n eventi del primo tipo ed m eventi del secondo tipo sia pari a $n + m$.

Per esempio si consideri per una fissata collettività di individui maschi e femmine le due percentuali, non note, di individui con capelli biondi e si assuma di essere interessati a precisare la nostra opinione su esse. Un modello matematico possibile potrebbe essere il precedente e la dipendenza tra gli indicatori dei due tipi indicherebbe che, per qualche motivo di carattere biologico, non riteniamo indipendenti tra loro le due percentuali.

20) Un secondo esempio

In un secondo esempio, a tempo continuo, del teorema di rappresentazione assumeremo che le funzioni di ripartizione univariate $F_1^{(\omega)}(y)$, comune a tutti i n.a. $Y(t)$, ed $F_2^{(\eta)}(z)$, comune a tutti i n.a. $Z(s)$, siano entrambe di tipo Gaussiano con valori medi ω ed η e varianze unitarie; supponiamo ancora che $G(\omega, \eta)$ 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 $F(y_1, \dots, y_n; z_1, \dots, z_m)$ e' la funzione di ripartizione di una distribuzione Gaussiana, $(n+m)$ – dimensionale, caratterizzata dai momenti

$$E[Y(t)] = \mu_1, \quad E[Z(s)] = \mu_2, \quad \text{Var}[Y(t)] = 1 + \gamma_1, \quad \text{Var}[Z(s)] = 1 + \gamma_2, \quad \text{Cov}[Y(t), Z(s)] = \gamma.$$

21) Cenni sui processi di Markov.

La condizione di **dipendenza stocastica Markoviana** è stata introdotta da A. Markov nel 1906. Fissati n valori $t_1 < t_2 < \dots < t_n$ del parametro operativo tale condizione è espressa dalle $F(x_{t_n} / x_{t_1}, \dots, x_{t_{n-1}}) = F(x_{t_n} / x_{t_{n-1}})$ o dalla corrispondente uguaglianza tra le densità di probabilità se queste esistono. Una prima implicazione per le distribuzioni congiunte che supporremo dotate di densità è la seguente:

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

Una seconda notevole implicazione è la seguente: se $s < \tau < t$ 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 è nota in letteratura come condizione di Chapman – Kolmogorov.

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

Sono costituite 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:

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

Si verifica facilmente che in tal caso la struttura probabilistica del processo è univocamente determinata dalla matrice delle probabilità subordinate $P = [p_{ij}]$ e dalla distribuzione $a(0)$ (vettore riga) del n.a. X_0 , detta “distribuzione iniziale”. La distribuzione di X_1 è data dal prodotto $a(1) = a(0) \cdot P$ e quella di X_n dalle

$$a(n) = a(n-1) \cdot P = a(0) \cdot P^n$$

L'elemento $p_{ij}^{(n)}$ di P^n è detto “probabilità subordinata di transizione da i a j in n

passi” e soddisfa la condizione di Chapman – Kolmogorov $p_{ij}^{(n)} = \sum_{h=1}^N p_{ih}^{(n-m)} \cdot p_{hj}^{(m)}$.

Teorema di Markov: se la matrice P^n ha elementi definitivamente positivi essa converge, al divergere di n , ad una matrice U con elementi positivi e 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 ogni 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**.

Una matrice P ha elementi definitivamente positivi se esiste un intero N tale che per ogni $n > N$ gli elementi p_{ij}^n sono > 0 ; in questo caso la catena di Markov è detta “regolare”. In una catena regolare tutti gli stati “comunicano” tra loro in un numero di passi finito. Dal momento che le righe della matrice limite U sono tutte uguali tra loro la condizione di stazionarietà $\mathbf{u} \cdot P = \mathbf{u}$ è equivalente alla $U \cdot P = U$.

23) Processi a tempo continuo con un insieme discreto di stati .

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:

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

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:

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

24) Generatore del processo a tempo continuo con stati finiti.

La legge temporale del processo richiederebbe che si fissi il vettore $a(0)$ e, in modo coerente, la totalità delle matrici di transizione $\{P(t); t \geq 0\}$; in alternativa basterà fissare, oltre ad $a(0)$, un'unica matrice G chiamata "generatore del processo".

Teorema: posto $P(0) = I$, se $\lim_{t \downarrow 0} P(t) = P(0)$ allora esiste una matrice $G = [g_{ij}]$ tale

che $P(t) = \exp\{t \cdot G\} = \sum_{n=0}^{\infty} \frac{(t \cdot G)^n}{n!}$. Gli elementi g_{ij} di G verificano le condizioni $g_{ij} \geq 0$

se i è diverso da j e, per $i = j$, $g_{ii} = -\sum_{j \neq i} g_{ij}$ in modo che $G \cdot \mathbf{1} = \mathbf{0}$.

25) Esempio

Se a e b sono numeri reali positivi e se gli autovalori a_j della matrice

$$G = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \text{ sono diversi tra loro, allora è } P(t) = \exp\{t.G\} = V \cdot \text{diag}(e^{t.a_j}) \cdot V^{-1}$$

la generica matrice di transizione del processo di cui G è il generatore. V è la matrice le cui colonne sono gli autovettori corrispondenti agli autovalori a_j , $j = 1, 2$.

Ponendo uguale a zero il determinante della matrice $(G - \lambda.I_2)$ e risolvendo l'equazione si trovano i due autovalori $\lambda_1 = 0$ e $\lambda_2 = -(a+b)$, diversi tra loro.

Gli autovettori corrispondenti sono le colonne della matrice $V = \begin{bmatrix} 1 & 1 \\ 1 & -b/a \end{bmatrix}$ e per

$P(t)$ si trova la somma di matrici:

$$P(t) = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix} + \begin{bmatrix} \frac{a}{a+b} & -\frac{a}{a+b} \\ -\frac{b}{a+b} & \frac{b}{a+b} \end{bmatrix} \cdot \exp\{-t.(a+b)\}.$$

26) Il processo di Poisson.

L'insieme degli stati è 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 $P\{N(t) = n\} = e^{-\lambda t} \cdot (\lambda t)^n / n!$ con $\lambda > 0$.

Ritourneremo sul processo di Poisson parlando dei processi di Lévy.

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

Assumeremo che lo spazio degli stati sia R . Nel caso di un processo markoviano a tempo *discreto* le probabilità di transizione sono espresse da una funzione detta, stochastic kernel, $K(x,S) = \text{Prob} \{X_{n+1} \in S / X_n = x\}$ dipendente di due argomenti:

$x \in R$ ed $S \subset R$; fissato x , $K(x, \cdot)$ è una probabilità sui sottoinsiemi boreliani S di R , mentre fissato S , $K(\cdot, S)$ è una funzione misurabile.

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:

$$K^{(m+n)}(x,S) = \int_R K^{(m)}(x,dy) \cdot K^{(n)}(y,S) \quad \text{e} \quad k^{(m+n)}(x,y) = \int_R k^{(m)}(x,z) \cdot k^{(n)}(z,y) dz.$$

Esempio : il processo definito dal modello AR(1) $X_{n+1} = a \cdot X_n + \varepsilon_n$, $X_0 = x \in R$, $\varepsilon_n \approx NWN(0, \sigma^2)$. In questo caso esiste la densità subordinata di transizione $k(x, y)$

di tipo $N(a \cdot x; \sigma^2)$; si ha ovviamente $k^n(x, y) = \int_R k^{n-1}(x, z) \cdot k(z, y) dz$.

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

Esse soddisfano le condizioni $Q_t(x,S) \geq 0$, $Q_t(x,R) = 1$ e le relazioni di Chapman – Kolmogorov : $Q_t(x,S) = \int_R Q_{t-\tau}(x,dy) \cdot Q_\tau(y,S)$, $0 < \tau < t$.

Esempio : il processo di Poisson composto.

Si tratta di un processo markoviano a parametro continuo definito dalla $X(t) = \sum_{j=0}^{N(t)} Y_j$

ove $N(t)$ è un processo di Poisson con intensità λ e i n.a. Y_j sono i.i.d. (con una comune distribuzione $F_Y(y)$) e indipendenti dagli $N(t)$.

29) Cenni sui processi di P. Lévy

Si tratta di un'importante classe di processi di Markov a tempo continuo della quale fanno parte

- 1) Il processo di Poisson,
- 2) Il processo di Poisson composto,
- 3) Il processo di Moto Browniano o processo di Wiener,
- 4) Il processo Gamma

e tanti altri processi stocastici utilizzati in molti campi applicativi.

Sostanzialmente sono processi $\{X(t); t \geq 0\}$ i cui incrementi $X(t) - X(s)$ sono stocasticamente indipendenti se riguardano intervalli di tempo disgiunti e la cui distribuzione dipende dalle epoche t ed s soltanto attraverso la differenza $t - s$. Si parla anche di processi con **incrementi indipendenti e omogenei**.

30) Processi stocastici del secondo ordine

La totalità dei n.a. definiti in uno spazio di probabilità (Ω, A, P) e dotati di momenti secondi finiti si indica con $H = L_2(\Omega, A, P)$ e tale insieme di n.a. costituisce uno spazio di Hilbert con funzione prodotto – interno $\langle X, Y \rangle = E(X.Y)$ 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 $\langle X, Y \rangle = 0$. Se questi n.a. hanno tutti valor medio nullo allora è

$\langle X, Y \rangle = \text{Cov}(X, Y)$, $\|X\| = \sqrt{\text{Var}(X)}$, $d(X, Y) = \sqrt{\text{Var}(X - Y)}$ e infine $X \perp Y \leftrightarrow \text{Cov}(X, Y) = 0$.

L'eventuale convergenza in norma per questi n.a. coincide con la convergenza in media quadratica.

Un processo stocastico i cui n.a. appartengono ad H si dice “del secondo ordine”. Nonostante che la specificazione riguardi soltanto i loro valori medi, le loro varianze e le covarianze (e non le loro distribuzioni congiunte) tali processi trovano molte applicazioni nei modelli stocastici. L'esempio classico è costituito dal processo White Noise (o di rumore bianco) che modella matematicamente i “rumori”, i “disturbi casuali” o le “perturbazioni” non prevedibili in sistemi dinamici.

31) Calcolo differenziale in media quadratica

Per processi stocastici $\{Y(t); t \geq 0\}$ del tipo suddetto **a tempo continuo**, detti anche “funzioni aleatorie” del secondo ordine, 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 $P_n = (a = t_0 < t_1 < \dots < t_n = b)$.

32) Approssimazioni dei minimi quadrati per numeri aleatori

Si consideri lo spazio lineare dei n.a. $H = L_2(\Omega, A, P)$ che supponiamo dotati di speranza matematica nulla (e momento secondo finito). Sia Y un 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 $E[Y - E(Y/\mathbf{X})]^2 \leq E[Y - \varphi(\mathbf{X})]^2$.

Ci proponremo allora di minimizzare $E[Y - \varphi(\mathbf{X})]^2$ rispetto alla funzione $\varphi(\cdot)$: si ha chiaramente $E[Y - \varphi(\mathbf{X})]^2 = E\{[Y - E(Y/\mathbf{X})] + [E(Y/\mathbf{X}) - \varphi(\mathbf{X})]\}^2$. Innalzando al quadrato la somma delle parentesi quadre e applicando l'operatore E si trova che $E[Y - \varphi(\mathbf{X})]^2 = E[Y - E(Y/\mathbf{X})]^2 + E[E(Y/\mathbf{X}) - \varphi(\mathbf{X})]^2$ e tale somma riesce minima scegliendo $\varphi(\mathbf{X}) = E(Y/\mathbf{X})$.

33) Approssimazioni lineari dei minimi quadrati

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

$$\hat{\alpha}_j, j = 1, \dots, n, \text{ per i quali risulta } E \left[Y - \sum_j \hat{\alpha}_j \cdot X_j \right]^2 \leq E \left[Y - \sum_j \alpha_j \cdot X_j \right]^2$$

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 $\hat{\underline{\alpha}} = [\text{Cov}(\mathbf{X})]^{-1} \cdot E(Y \cdot \mathbf{X})$ sotto la condizione che la matrice di dispersione di $\mathbf{X}$ sia invertibile. Il previsore lineare ottimale per Y è allora $\hat{Y} = (X_1, \dots, X_n) \cdot [\text{Cov}(X)]^{-1} \cdot E(Y \cdot X)$.

34) Interpretazione geometrica del previsore lineare ottimale.

Il n.a. $\hat{Y} = (X_1, \dots, X_n) \cdot [\text{Cov}(X)]^{-1} \cdot E(Y \cdot X)$ coincide con la **proiezione ortogonale** di Y sul sottospazio lineare L generato dai n.a. $X_1, \dots, X_n$, indicata con $P(Y/L)$ e definita come n.a. di L a minima distanza da Y. Formalmente,

1. $P(Y/L) \in L$ e quindi è una combinazione lineare dei n.a. $\{X_j; j = 1, \dots, n\}$,
2. $d[Y, P(Y/L)] = \sqrt{E \left(Y - \sum_{j=1}^n \alpha_j \cdot X_j \right)^2}$ è minima rispetto agli $\{\alpha_j; j = 1, \dots, n\}$.

Questa proposizione si dimostra trovando il minimo della funzione dei coefficienti α_j sotto il segno di radice quadrata; il risultato è $\hat{\underline{\alpha}} = \left(\text{Cov}(\underline{X}) \right)^{-1} \cdot E(\underline{Y} \cdot \underline{X})$ per cui

$$\hat{Y} = (X_1, \dots, X_n) \cdot \hat{\underline{\alpha}} = (X_1, \dots, X_n) \cdot \left(\text{Cov}(\underline{X}) \right)^{-1} \cdot E(\underline{Y} \cdot \underline{X}).$$

35) Proprietà iterativa dell'approssimazione lineare dei minimi quadrati

Si considerino i due sottospazi lineari $L_{n-1} = L(X_1, \dots, X_{n-1})$ ed $L_n = L(X_1, \dots, X_n)$, generati dai n.a. X_j linearmente indipendenti ed osservabili. Sia $Y \notin L_n$ il n.a. non osservabile. Proveremo la seguente **proprietà iterativa**:

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

La dimostrazione è fornita dalla seguente semplice sequenza di uguaglianze:

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

36) Altre proprietà del proiettore ortogonale

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

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

$$\gamma) \langle P(X / L_n), Y \rangle = \langle X, P(Y / L_n) \rangle , \text{ ove } \langle X, Y \rangle = E(X \cdot Y) ,$$

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

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

L'ultima proprietà giustifica l'importanza del procedimento di ortogonalizzazione di Gram – Schmidt: applicandolo ad un insieme di n.a. $\{X_i; i = 1, \dots, n\}$ si ottiene un insieme di altri n.a. $\{Y_i; i = 1, \dots, n\}$ che generano lo stesso spazio lineare degli X_i e sono mutuamente non correlati. Per questi sussiste l'interessante proprietà ε)

37) Problema dinamico di stima lineare dei minimi quadrati

Assumiamo di essere interessati ai n.a. di un processo stocastico $\{Y_n; n \geq 1\}$ non osservabile, generato da un modello stocastico lineare espresso dalle equazioni alle differenze finite (1) $Y_{n+1} = a.Y_n + V_n$, ove $V_n \approx WN(0; \sigma_V^2)$ 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) $X_n = b.Y_n + W_n$, ove $W_n \approx WN(0; \sigma_W^2)$ e ove $W_n \perp V_m, \forall n, m$.

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".

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

Indicheremo con $\hat{Y}_{n/n}$ l'approssimazione lineare ottimale del n.a. Y_n non osservabile basata sui n.a. osservabili $X_1, \dots, X_n$. Come sappiamo essa può essere indicata anche con il simbolo $P(Y_n / L_n)$ ed è data dalla $\hat{Y} = (X_1, \dots, X_n) \cdot [Cov(X)]^{-1} \cdot E(Y \cdot X)$

Ci proponiamo di presentare un 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 $Y_n - \hat{Y}_{n/m}$:

$$(3) \quad \hat{Y}_{n+1/n} = a \cdot \hat{Y}_{n/n}, \quad (4) \quad P_{n+1/n} = E(Y_{n+1} - \hat{Y}_{n+1/n})^2 = a^2 \cdot P_{n/n} + \sigma_V^2,$$

$$(5) \quad \hat{Y}_{n+1/n+1} = \hat{Y}_{n+1/n} + \frac{b \cdot P_{n+1/n}}{b^2 \cdot P_{n+1/n} + \sigma_W^2} \cdot (X_{n+1} - b \cdot \hat{Y}_{n+1/n}), \quad (6) \quad P_{n+1/n+1} = \frac{\sigma_W^2 \cdot P_{n+1/n}}{b^2 \cdot P_{n+1/n} + \sigma_W^2}.$$

Si intuisce che tale procedimento iterativo è chiaramente adatto alla sua implementazione mediante elaboratori elettronici; questi si sono resi disponibili quasi simultaneamente alla comparsa del Filtro di R. E. Kalman.

39) Qualche osservazione sul filtro di Kalman

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 (3) e (4) sono dette **“equazioni di previsione”** mentre le (5) e (6) sono dette **“equazioni di aggiornamento”** cosicché ogni stadio del procedimento ricorsivo consiste di una fase di previsione e una di aggiornamento. La dimostrazione delle formule utilizza le proprietà sopra elencate del proiettore ortogonale. Aggiungiamo che la (5) si basa sulla proprietà ricorsiva del proiettore ortogonale e utilizza l’incremento di informazione costituito dall’osservazione del n.a. X_{n+1} .

Un’ultima osservazione riguarda la possibilità di pre-calcolare la sequenza delle varianze degli errori di approssimazione $\{p_{n/n}; n \geq 1\}$ in quanto nelle formule (4) e (6) non intervengono i valori delle osservazioni che si rendono via via disponibili.