

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 (Ω, \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:

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

Nozione di modello dinamico.

In generale un modello dinamico è costituito da due elementi:

- 1) **lo spazio degli stati** i cui elementi possono essere numeri (reali o complessi), vettori numerici o altro;
- 2) **la regola di evoluzione** che descrive le relazioni tra gli stati del sistema.

Chiameremo **funzione di evoluzione** quella che descrive la dipendenza dello stato dal tempo; in generale tale funzione è soluzione di un'equazione.

Bisogna distinguere i sistemi dinamici **a tempo discreto** da quelli **a tempo continuo**: nel primo caso la funzione di evoluzione è soluzione di **equazioni alle differenze finite**; nel secondo è soluzione di **equazioni differenziali**. Entrambe le categorie di equazioni sono **equazioni funzionali**, cioè equazioni nelle quali l'incognita è una funzione (e non un numero come nelle equazioni algebriche). Vediamo un primo semplice esempio in cui la regola di evoluzione è lineare ed affine, cioè $y = a.x + b$.

1. Nei modelli a tempo discreto la funzione $y = a.x + b$ descrive il passaggio del sistema dallo stato attuale x allo stato y nel prossimo periodo (unitario); nel periodo successivo dallo stato y si passerà allo stato $z = a.y + b$. Per semplicità indicheremo lo stato nel periodo ennesimo con x_n e con $x_{n+1} = a.x_n + b$ quello nel periodo $n+1$: l'ultima equazione indicherà la generale equazione di evoluzione del sistema e si tratta di un'equazione alle differenze finite del primo ordine.

Indicato con x_0 lo stato di partenza (nel periodo zero), dopo n transizioni si arriverà allo stato $x_n = a^n . x_0 + \frac{b}{1-a} . (1 - a^n)$ e tale relazione costituisce la soluzione dell'equazione $x_{n+1} = a.x_n + b$ relativa alla "condizione iniziale" o stato di partenza x_0 .

2. Nei modelli a tempo continuo indicheremo con $x(t)$ lo stato del sistema al tempo t e la stessa regola di evoluzione lineare e affine $y = a.x + b$ descrive le evoluzioni del sistema con le soluzioni dell'equazione differenziale $\frac{d}{dt}x(t) = a.x(t) + b$.

La soluzione corrispondente alla condizione iniziale o stato di partenza $x(0)$ è data dalla funzione $x(t) = x(0) . e^{a.t} - \frac{b}{a} . (1 - e^{a.t})$.

Se l'equazione alle differenze finite o l'equazione differenziale è stocastica, cioè se l'input dell'equazione è un processo stocastico, allora la soluzione dell'equazione e quindi la funzione di evoluzione del sistema dinamico sarà ancora un processo stocastico (e si parlerà di sistema dinamico stocastico).

Cenni sull'oscillatore armonico

È detto oscillatore armonico ogni sistema dinamico che spostato da una sua posizione di equilibrio tende a ritornarvi, per effetto di una forza di ripristino proporzionale all'entità dello spostamento, con un eventuale movimento oscillatorio attorno alla posizione di equilibrio. Alcuni esempi: corpo sospeso ad una molla, pendolo semplice, circuito RCL, Nel seguito faremo riferimento al primo esempio.

Il modello matematico si costruisce attraverso la seconda legge di Newton $F_i = m \cdot a_i = m \cdot x''(t)$, ove $x(t)$ rappresenta l'entità dello spostamento al tempo t dalla posizione di riposo e $a_i = x''(t)$ è l'accelerazione.

Considereremo tre casi:

- 1) **l'oscillatore armonico semplice** in cui interviene la sola forza di ripristino.

Il modello matematico è il seguente: $m \cdot \frac{d^2}{dt^2} x(t) = -k \cdot x(t)$ ove $k > 0$ indica la costante caratteristica della molla. La soluzione corrispondente alle condizioni iniziali $x(0) = c_1$ e $x'(0) = v$ è data dall'espressione $x(t) = c_1 \cdot \cos(\omega_0 \cdot t) + c_2 \cdot \sin(\omega_0 \cdot t)$ ove $\omega_0 = \sqrt{k/m}$ e $c_2 = v/\omega_0$.

- 2) **l'oscillatore armonico smorzato** in cui, oltre alla forza di ripristino, interviene la perdita di energia dovuta alla resistenza dell'aria, o all'attrito o ad altri fenomeni dissipativi; si assume che tale perdita di energia sia proporzionale alla derivata prima $x'(t)$ dello spostamento, cioè alla velocità. Il modello

matematico è il seguente: $m \cdot \frac{d^2}{dt^2} x(t) = -k \cdot x(t) - \gamma \cdot \frac{d}{dt} x(t)$ con $\gamma > 0$. Ci limitiamo a dire che, sotto opportune condizioni, la soluzione $x(t)$ ha un andamento oscillatorio smorzato attorno alla posizione di equilibrio.

3) **l'oscillatore armonico forzato** in cui, oltre alla forza di ripristino e d'attrito, interviene un forza esterna $\Phi(t)$, deterministica o stocastica.

Il modello matematico è costituito dall'equazione differenziale

$$m \cdot \frac{d^2}{dt^2} x(t) = -k \cdot x(t) - \gamma \cdot \frac{d}{dt} x(t) + \Phi(t)$$

Ci limitiamo a dire che la soluzione, corrispondente a date condizioni iniziali, ha due componenti: quella transitoria, dipendente dalle condizioni iniziali, e quella di lungo periodo che risulta sostanzialmente proporzionale a $\Phi(t)$ e indipendente dalle condizioni iniziali.

Per qualche maggior dettaglio formale assumiamo che sia $\Phi(t) = m \cdot \cos(\omega \cdot t)$, con $\omega \neq \omega_0$, e riscriviamo la precedente equazione differenziale nella forma

$$\frac{d^2}{dt^2} x(t) + 2\xi\omega_0 \frac{d}{dt} x(t) + \omega_0^2 \cdot x(t) = \cos(\omega \cdot t),$$

ove si è posto $\xi = \frac{\gamma}{2 \cdot \sqrt{k \cdot m}}$ e, come già detto, $\omega_0^2 = \frac{k}{m}$. Se infine poniamo per semplicità $\omega_0 = 1$ la componente transitoria della soluzione $x(t)$ è

$$x_T(t) = \exp\{-\xi \cdot t\} \cdot \left[c_1 \cdot \cos(t \cdot \sqrt{1 - \xi^2}) + c_2 \cdot \sin(t \cdot \sqrt{1 - \xi^2}) \right]$$

sotto l'ipotesi $\xi < 1$.

La componente di lungo periodo $x_S(t)$ è invece data dall'espressione

$$x_S(t) = A \cdot \cos(\omega \cdot t + \varphi), \text{ con } \operatorname{tg}(\varphi) = \frac{2 \cdot \omega \cdot \xi}{\omega^2 - 1} \text{ e } A = \frac{\cos \varphi}{1 - \omega^2}, \text{ ed infine si ha}$$

$$x(t) = x_T(t) + x_S(t).$$

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 $\{X_t; t \in Z^+\}$ 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 $\{X_t; t \in Z^+\}$ 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 $t_j + h \in Z^+$;

c) il processo $\{X_t; t \in Z^+\}$ 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 $\Psi_X(s, t) = Cov(X_s, X_t)$ dipende solo da $t - s$.

APPENDICE N. 3 : Problema ergodico

Sono dette "leggi deboli dei grandi numeri" quei teoremi che riguardano la convergenza in probabilità delle "medie temporali" (o momenti campionari)

$\frac{1}{n} \sum_{j=1}^n X_j$ al divergere di n . Il lettore ricorderà il teorema di Bernoulli ove le variabili aleatorie sono indicatori di eventi indipendenti e ugualmente probabili; si ha

$$\frac{1}{n} \sum_{j=1}^n |E_j| \rightarrow P(E_1).$$

Se gli indicatori sono invece scambiabili, il teorema di B. de Finetti afferma che

$\frac{1}{n} \sum_{j=1}^n |E_j| \rightarrow X$ ove il limite X è un numero aleatorio con valor medio $\omega_1 = P(E_1)$.

Le "leggi forti dei grandi numeri" riguardano invece la convergenza quasi certa di medie temporali; il primo esempio che riportiamo concerne numeri aleatori stazionari in senso stretto, dotati di momento primo assoluto finito; si ha

$$\frac{1}{n} \sum_{j=1}^n X_j \rightarrow X^*, \quad E(X^*) = E(X_1),$$

e si noti l'analogia tra questo risultato e il teorema di de Finetti.

Si avrebbe invece con probabilità uno $\frac{1}{n} \sum_{j=1}^n X_j \rightarrow E(X_1)$ se il processo stocastico stazionario fosse anche "ergodico", cioè se la funzione di covarianza $\psi_X(h)$ tendesse a zero, al crescere di h , molto rapidamente, per esempio esponenzialmente.

E' chiaro allora che non tutti i processi stazionari sono ergodici; per esempio i processi scambiabili, che sono stazionari, non sono ergodici: infatti la funzione di covarianza dei processi scambiabili è costante e, in generale, non nulla.

Per definizione, un processo stazionario in senso stretto è anche ergodico quando, per ogni intero positivo k , si ha la convergenza (in probabilità o quasi certamente) dei momenti campionari $\frac{1}{n} \sum_{j=1}^n X_j^k$ ai corrispondenti momenti spaziali $E(X_1^k)$.

Nella maggior parte dei testi sui teoremi ergodici si usa una differente impostazione formale. Si parte da uno spazio di probabilità (Ω, F, P) e da un'applicazione di applicazione misurabile $\theta: \Omega \rightarrow \Omega$ che soddisfa la proprietà espressa dalla $P(\theta^{-1}A) = P(A)$, $\forall A \in F$. Cioè la probabilità di A si conserva nell'applicazione della θ agli elementi di A (in lingua inglese : measure preserving property). Questa proprietà corrisponde alla stazionarietà in senso stretto del processo stocastico le cui variabili sono definite dalle $X_1 = X(\omega)$, $X_2 = X(\theta\omega)$, $X_3 = X(\theta^2\omega)$ e in generale, $X_n = X(\theta^{n-1}\omega)$, essendo $X(\omega)$ una fissata funzione misurabile da Ω in \mathbb{R} . Assumeremo che sia $E[X(\omega)] = 0$, e ciò implica $E(X_n) = 0$, e che $E[X^2(\omega)] < \infty$, e ciò implica $E(X_n^2) < \infty$. Per semplicità, verifichiamo che per il processo così definito la funzione di covarianza dipende solo dalla differenza dei deponenti delle variabili:

$$E(X_{n+m} \cdot X_n) = E[X(\theta^{n+m-1}\omega) \cdot X(\theta^{n-1}\omega)] = E[X(\theta^m\omega) \cdot X(\omega)] = E(X_{m+1} \cdot X_1).$$

L'uguaglianza centrale richiede un chiarimento: a partire dalla $X_n = X(\theta^{n-1}\omega)$ poniamo $\theta X_n = X(\theta^n\omega) = X_{n+1}$ interpretando l'operatore θ , quando applicato alle variabili X_n , quale operatore di traslazione di un passo.

Poiché $E(X_{n+m} \cdot X_n) = \langle X_{n+m}, X_n \rangle$ si prova che dalla proprietà $P(\theta^{-1}A) = P(A)$ discende l'invarianza del prodotto interno rispetto alla traslazione: $\langle X_{n+m}, X_n \rangle = \langle \theta^k X_{n+m}, \theta^k X_n \rangle$ e da ciò si ottiene $Cov(X_{n+m}, X_n) = Cov(X_{m+1}, X_1)$.

Teorema ergodico per processi stazionari in senso lato $\{X_n : n \geq 1\}$:

Esiste un numero aleatorio Y tale che $\frac{1}{n} \sum_{j=1}^n X_j^k \xrightarrow{q.m.} Y$ e $E(Y) = E(X_1)$.

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

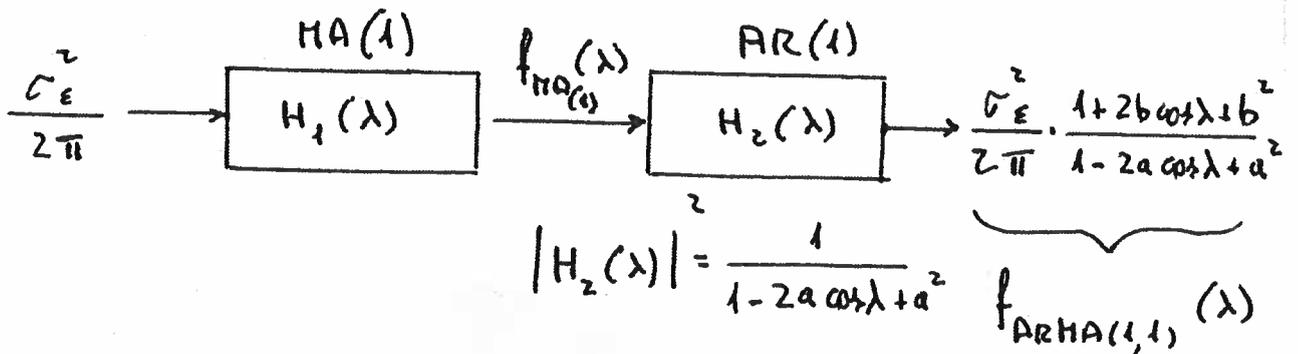
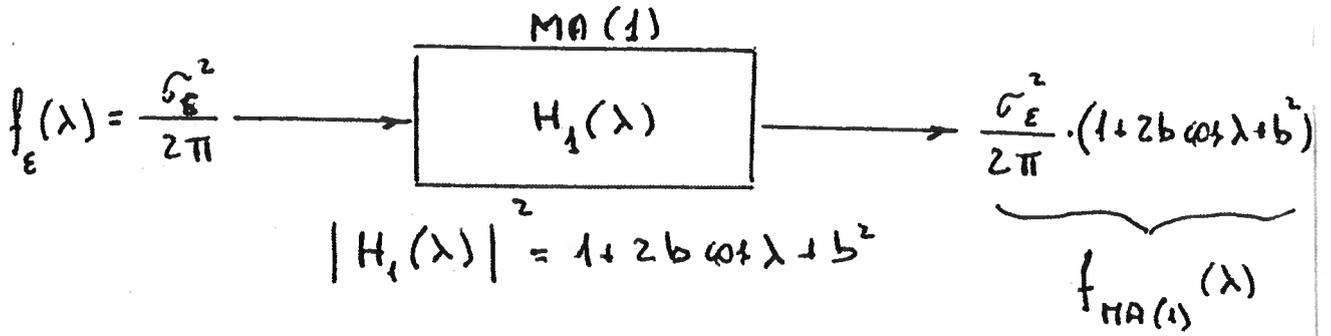
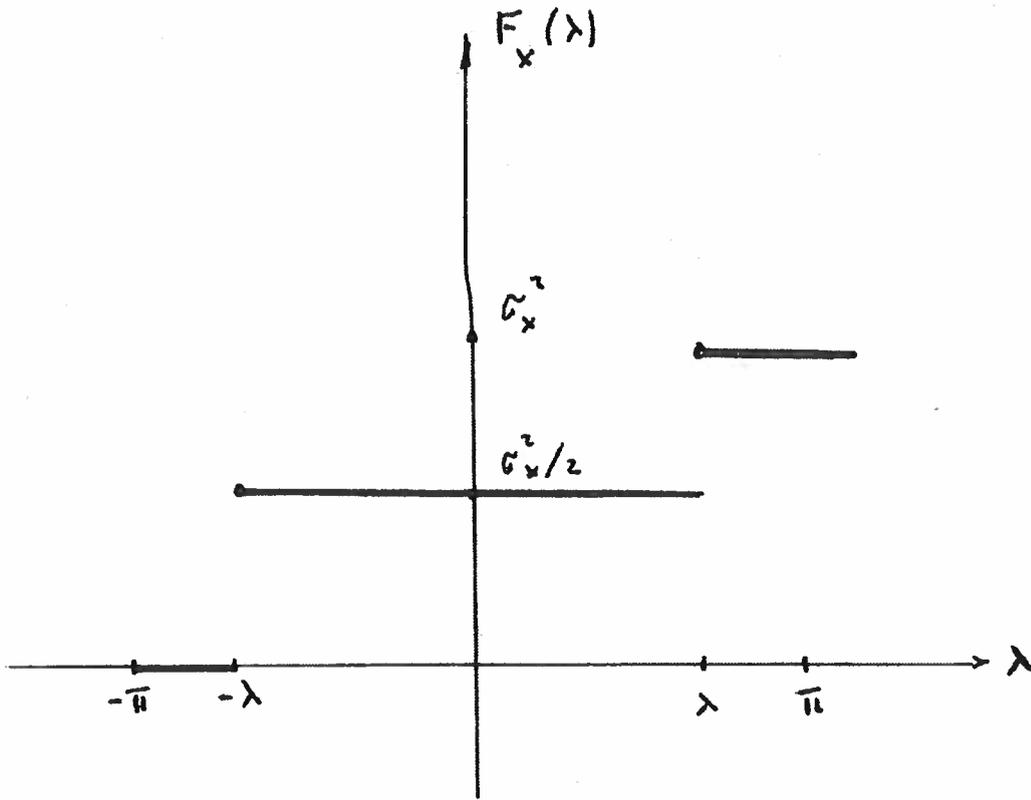
$X_t = \varepsilon_t + \sum_{k=1}^{\infty} a^k \cdot \varepsilon_{t-k}$: si provi che questo processo è stazionario in senso lato e che

$f_X(\lambda) = \sigma_\varepsilon^2 \cdot (1 - 2a \cdot \cos \lambda + a^2)^{-1} / 2\pi$ 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 $f_X(\lambda) = \sigma_\varepsilon^2 \cdot (1 + 2b \cdot \cos \lambda + b^2) / 2\pi$, 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

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

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:

- al divergere di n , si ha $p - \lim S_n/n = X$, ove $E(X) = P(E_1)$,
- al divergere di n , le distribuzioni delle S_n/n convergono alla $G_X(x)$,
- 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.

Pr. STAZIONARI: $(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})$

Pr. SCAMBIABILI: $(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h_1}, \dots, X_{t_n+h_n})$



Pr. SCAMBIABILE DI INDICATORI DI EVENTI: $X_n = |E_n|$

$$S_n = \sum_{j=1}^n |E_j|$$

TEOREMA DI RAPPRESENTAZIONE:

$$P(S_n = u) = \int_0^1 \binom{n}{u} x^u (1-x)^{n-u} dG(x) = \omega_u^{(n)}$$

$$P(S_n = u) = \int_0^1 x^u dG(x) = \omega_u^{(n)} = \omega_u$$

ESEMPI

1) PROCESSO BERNOULLIANO: $\omega_u = P(S_n = u) = x^u$

2) PROCESSO DI BAYES-LAPLACE: $\omega_u = \frac{1}{n+1} = \omega_u^{(n)}$

$$\omega_u^{(n)} = \frac{1}{n+1}$$

3) PROCESSO DI PÓLYA: $\omega_u = \frac{(\alpha)_n}{(\alpha+\beta)_n}$

$$\omega_u^{(n)} = \binom{n}{u} \frac{(\alpha)_u (\beta)_{n-u}}{(\alpha+\beta)_n}$$

ove $(\alpha)_u = \alpha(\alpha+1)(\alpha+2) \dots (\alpha+u-1)$

$G(x)$ {
 concentrata su x in 1)
 uniforme su $(0, 1)$ in 2)
 di tipo Beta (α, β) in 3)

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

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

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

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

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

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

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

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

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 $a(1) = a(0).P$ e quella di X_n dalle

$$a(n) = a(n-1).P = a(0).P^n$$

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 u la generica riga di U , è $u \cdot P = u$, per cui 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 $u \cdot P = u$ 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

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

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:

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

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 $N(t)$.

Modello AR(1) e stazionarietà del processo stocastico

Scriviamo il modello nel modo seguente:

$$X_n = a.X_{n-1} + \varepsilon_n, \quad |a| < 1, \quad \{\varepsilon_n; n \geq 1\} \approx WN(0, \sigma_\varepsilon^2), \quad X_0 = k \in R.$$

Per n=1 $\longrightarrow X_1 = a.k + \varepsilon_1,$

n=2 $\longrightarrow X_2 = a.X_1 + \varepsilon_2 = a^2.k + a.\varepsilon_1 + \varepsilon_2,$

n=3 $\longrightarrow X_3 = a.X_2 + \varepsilon_3 = a^3.k + a^2.\varepsilon_1 + a.\varepsilon_2 + \varepsilon_3,$

.....

per n > 3 $\longrightarrow X_n = a^n.k + (\varepsilon_n + a.\varepsilon_{n-1} + a^2.\varepsilon_{n-2} + \dots + a^{n-1}.\varepsilon_1).$

Si ottiene facilmente: $E(X_n) = a^n.k$ e $X_n - E(X_n) = \sum_{h=0}^{n-1} a^h.\varepsilon_{n-h}.$

Meno facilmente si ottiene per s > 0: $Cov(X_{n+s}, X_n) = \frac{\sigma_\varepsilon^2.a^s.(1-a^n)}{1-a^2}.$ Dalle

$$Cov(X_{n+s}, X_n) = E\left\{ [X_{n+s} - E(X_{n+s})][X_n - E(X_n)] \right\} = E\left\{ \left[\sum_{h=0}^{n+s-1} a^h.\varepsilon_{n+s-h} \right] \left[\sum_{j=0}^{n-1} a^j.\varepsilon_{n-j} \right] \right\}$$

si consideri il valor medio del generico prodotto $E\left\{ (a^h.\varepsilon_{n+s-h})(a^j.\varepsilon_{n-j}) \right\}$: per la non correlazione delle variabili ε_n tale valor medio sarà positivo solo quando

$\varepsilon_{n+s-h} = \varepsilon_{n-j}$ e in tal caso si avrà

$$E\left\{ \left[\sum_{h=0}^{n+s-1} a^h.\varepsilon_{n+s-h} \right] \left[\sum_{j=0}^{n-1} a^j.\varepsilon_{n-j} \right] \right\} = a^s.\sigma_\varepsilon^2.(1 + a^2 + a^4 + \dots + a^{2(n-1)})$$
 e quindi il

risultato sopra indicato.

Nell'espressione sono presenti entrambi gli interi s ed n per cui la covarianza non dipende solo dalla differenza (n+s)-n = s come richiede la stazionarietà in senso lato.

Solo per $n \rightarrow \infty$ si ottiene $Cov(X_{n+s}, X_n) = \frac{\sigma_\varepsilon^2.a^s}{1-a^2}$ per cui il processo X_n è detto

“asintoticamente stazionario in senso lato”.

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 $\langle X, Y \rangle = Cov(X, Y)$, $\|X\| = \sqrt{Var(X)}$, $d(X, Y) = \sqrt{Var(X - Y)}$ e infine $X \perp Y \leftrightarrow Cov(X, Y) = 0$. 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) : $\exists Z = \int_a^b Y(t) dt$, $\lim E \left\{ \sum_{i=0}^{n-1} Y(t_i) \cdot [t_{i+1} - t_i] - Z \right\}^2 = 0$

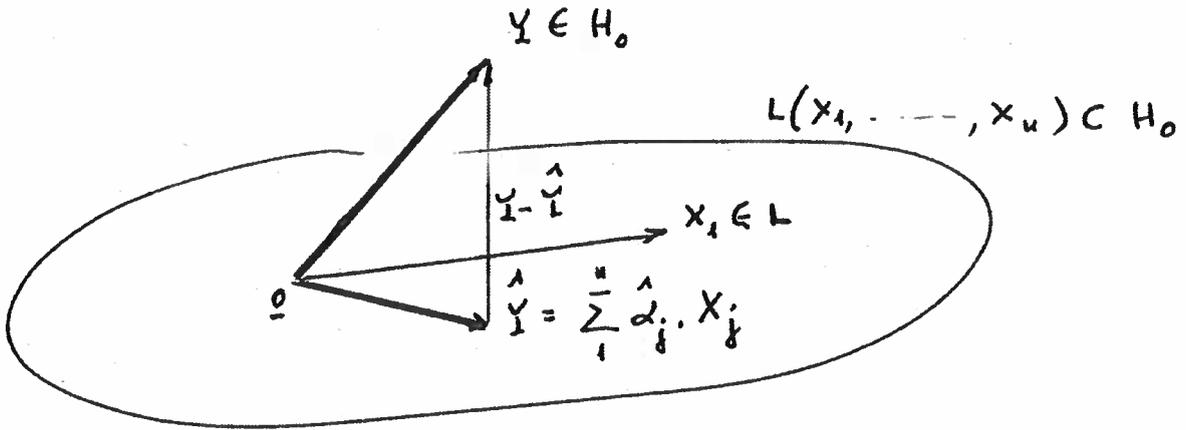
per $\max_i [t_{i+1} - t_i] \rightarrow 0$ in corrispondenza ad una sequenza di partizioni P_n sempre più fini di (a, b) , ove $P_n = (a = t_0 < t_1 < \dots < t_n = b)$.

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 $\varphi(\mathbf{X}) = E(Y/\mathbf{X})$.



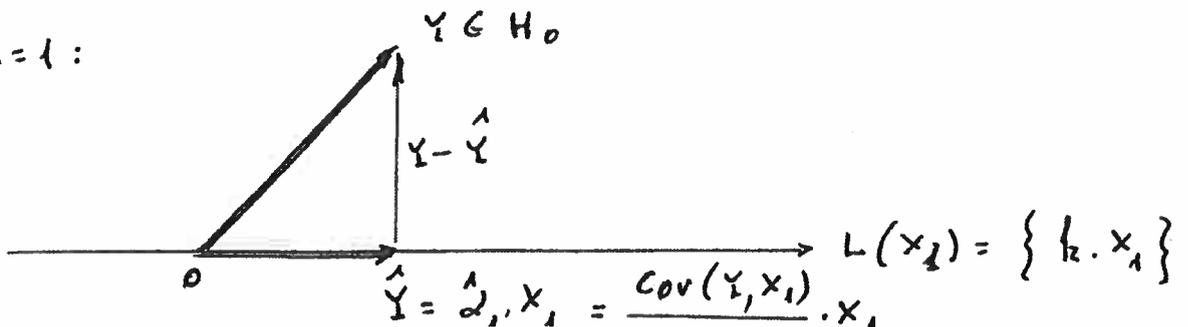
LEAST SQUARES APPROXIMATIONS

$$E \left(Y - \sum_{j=1}^n \hat{\alpha}_j X_j \right)^2 \left\{ \begin{array}{l} = [d(Y - \hat{Y})]^2 \\ \leq E \left(Y - \sum_{j=1}^n \alpha_j X_j \right)^2 = [d(Y - l)]^2, \forall l \in L \end{array} \right.$$

$$\hat{Y} = \sum_{j=1}^n \hat{\alpha}_j X_j = \mathcal{P}(Y|L) \left\{ \begin{array}{l} \in L(X_1, \dots, X_n) \\ Y - \mathcal{P}(Y|L) \perp X_j, j=1, \dots, n \end{array} \right.$$

$$\underline{\hat{\alpha}} = \underline{X}^T \underline{\hat{\alpha}} = \underline{X}^T \underbrace{[\text{Cov}(\underline{X})]^{-1}}_{\underline{\hat{\alpha}}} \cdot E(\underline{X} \cdot Y)$$

See n=1:



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, $\sum_{j=1}^n \alpha_j X_j$, dei n.a. X_1, \dots, X_n , si devono trovare gli n coefficienti

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

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 previsore lineare ottimale per Y è allora $\hat{Y} = (X_1, \dots, X_n) \cdot [\text{Cov}(\mathbf{X})]^{-1} \cdot E(\mathbf{Y} \cdot \mathbf{X})$.

34) Interpretazione geometrica del previsore lineare ottimale.

Il n.a. $\hat{Y} = (X_1, \dots, X_n) \cdot [\text{Cov}(\mathbf{X})]^{-1} \cdot E(\mathbf{Y} \cdot \mathbf{X})$ 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. $d[Y, P(Y/L)] = \sqrt{E \left(Y - \sum_{j=1}^n \alpha_j X_j \right)^2}$ 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 $\underline{\alpha} = \left(\text{Cov}(\underline{X}) \right)^{-1} \cdot E(\underline{Y} \cdot \underline{X})$ per cui

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

Approssimazione lineare ottimale per Y :

$$1) \hat{Y}_n = (x_1, \dots, x_n) \cdot [\text{Cov}(\underline{X})]^{-1} \cdot E(Y, \underline{X})$$

$$2) \hat{Y}_n = \hat{Y}_{n-1} + \underbrace{O\left\{Y \mid x_n - O(x_n \mid x_1, \dots, x_{n-1})\right\}}_{\text{Innovazione di } x_n}$$

3) Processo osservabile: $x_1, x_2, x_3, \dots, x_{n-1}, x_n, \dots$

Innovazioni:

$$z_1 = x_1$$

$$z_2 = x_2 - O(x_2 \mid x_1)$$

$$z_3 = x_3 - O(x_3 \mid x_1, x_2)$$



$$a) L(x_1, \dots, x_n) = L(z_1, \dots, z_n), \quad \forall n$$

$$b) z_i \perp z_j, \quad \forall i \neq j$$

$$\hat{Y}_n = (x_1, \dots, x_n) \cdot [\text{Cov}(\underline{X})]^{-1} \cdot E(Y, \underline{X}) =$$

$$= (z_1, \dots, z_n) \cdot [\text{Cov}(\underline{Z}_n)]^{-1} \cdot E(Y, \underline{Z}) =$$

$$= (z_1, \dots, z_n) \cdot \begin{bmatrix} v(z_1) & 0 & \dots & 0 \\ 0 & v(z_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v(z_n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \text{Cov}(Y, z_1) \\ \vdots \\ \text{Cov}(Y, z_n) \end{bmatrix}$$

$$= (z_1, \dots, z_n) \begin{bmatrix} \text{Cov}(Y, z_1) / v(z_1) & & \\ & \vdots & \\ & & \text{Cov}(Y, z_n) / v(z_n) \end{bmatrix} = \sum_{j=1}^n \frac{\text{Cov}(Y, z_j)}{v(z_j)} \cdot z_j =$$

$$\sum_{j=1}^n O(Y \mid z_j)$$

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

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

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

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

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

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

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.

Osservazioni sul filtro di R.E. Kalman (1960)

$$\hat{Y}_{n|m} = \mathcal{G}(Y_n | x_1, \dots, x_m)$$

$$P_{n|m} = \text{Var}(Y_n - \hat{Y}_{n|m}) = E \left[(Y_n - \hat{Y}_{n|m})^2 \right]$$

Condizioni iniziali: $\hat{Y}_{0|0}, P_{0|0}$

Dopo n osservazioni: $\hat{Y}_{n|n}, P_{n|n}$

1) Fase di previsione: $\hat{Y}_{n+1|n} = \mathcal{G}(Y_{n+1} | x_1, \dots, x_n)$

$$P_{n+1|n} = \text{Var}(Y_{n+1} - \hat{Y}_{n+1|n})$$

Equazioni (3) e (4)

2) Fase di aggiornamento:

$$\hat{Y}_{n+1|n} \longrightarrow \hat{Y}_{n+1|n+1} \quad \text{Equazione (5)}$$

$$P_{n+1|n} \longrightarrow P_{n+1|n+1} \quad \text{Equazione (6)}$$

Equazioni (3) e (5):

$$\hat{Y}_{n+1|n} = \mathcal{G}(aY_n + V_n | L_n) = a \mathcal{G}(Y_n | L_n) + \mathcal{G}(V_n | L_n) = a \cdot \hat{Y}_{n|n}$$

$$\begin{aligned} \hat{Y}_{n+1|n+1} &= \hat{Y}_{n+1|n} + \mathcal{G}(Y_{n+1} | X_{n+1}^*) = \\ &= \hat{Y}_{n+1|n} + \frac{\text{Cov}(Y_{n+1}, X_{n+1}^*)}{\text{Var}(X_{n+1}^*)} \cdot X_{n+1}^* \end{aligned}$$

Attilio Wedlin

Dipartimento di Scienze economiche e statistiche dell'Università di Trieste

DAL PROCESSO DI POISSON AL PROCESSO PSEUDO- POISSONIANO: ASPETTI PROBABILISTICI E STATISTICI

Riassunto

Da una decina d'anni è ripreso l'interesse dei ricercatori per i processi di P. Lévy, considerati tanto dal punto di vista teorico che da quello applicativo. Tra le applicazioni una posizione preminente spetta alla Finanza matematica: uno sviluppo moderno delle antiche discipline Matematica finanziaria e Matematica attuariale. L'attuale Seminario riguarda una modalità di presentazione dei processi stocastici di Lévy che li mette in relazione con i ben noti processi di passeggiata aleatoria e con i processi di Markov – Feller. Gli aspetti strutturali riguardano in particolare due teoremi di approssimazione dei processi di Lévy e dei processi di Markov – Feller: il primo teorema risale ai lavori di B. de Finetti del biennio 1929 – 1930 e sostanzialmente coincide con un primo teorema di rappresentazione per le distribuzioni infinitamente divisibili. Il Seminario affronta anche alcuni problemi di inferenza statistica su un particolare tipo di processi di Lévy, i processi di Poisson composto; vengono considerati sia il noto approccio inferenziale classico che quello bayesiano.

DE FINETTI - LEVY STOCHASTIC PROCESSES

$\{Z(t)\}$ has stationary and independent increments $Z(t) - Z(s)$,

$$\{Z(t)\} \rightarrow \lim_{s \rightarrow t} \Pr\{|Z(s) - Z(t)| > \varepsilon\} = 0 \text{ and } Z(0) = 0.$$

Some examples:

- the **Wiener process** $\begin{cases} \text{Gaussian pr.s with cont. trajectories} \\ E[Z(t)] \equiv 0, \text{ Cov}[Z(s), Z(t)] = \min(s, t) \end{cases}$
- the **Poisson process**: $P[Z(t) = n] = \exp\{-\lambda t\} \cdot \frac{(\lambda t)^n}{n!}$
- the **compound-Poisson pr.:** $Z(t) = \sum_{j=1}^{N(t)} U_j \begin{cases} N(t) \approx \text{Poisson}(\lambda t) \\ U_j / N(t) = n \approx \text{i.i.d.} \end{cases}$
- the **Gamma process**; α - **stable process**; etc.

As they have stationary and independent increments :

1) $Z(t) \sim$ **infinitely divisible distribution**: $\forall n, \varphi_{Z(t)}(\xi) = [\varphi_n(\xi)]^n$;

2) $\varphi_{Z(t+s)}(\xi) = \varphi_{Z(t)}(\xi) \cdot \varphi_{Z(s)}(\xi) \Rightarrow \varphi_{Z(t)}(\xi) = \exp\{t \cdot \psi(\xi)\}$;

3) $\varphi_{Z(t)}(\xi) = E[e^{i\xi Z(t)}] = \exp\left\{ibt\xi - \frac{ct\xi^2}{2} + t \cdot \int_R [e^{i\xi x} - 1 - i\xi x I_{(-1,1)}(x)] \nu(dx)\right\}$

if $\nu(\{0\}) = 0$ and $\int_R (|x|^2 \wedge 1) \nu(dx) < \infty$; $\varphi_{Z(t)}(\xi) \Leftrightarrow (b, c, \nu)$

4) $Z(t) = m(t) + Z_c(t) + Z_j(t), \quad Z_c(t) \coprod Z_j(t).$

ALCUNI RIFERIMENTI BIBLIOGRAFICI DEI PRIMI CONTRIBUTI
SUI PROCESSI CON INCREMENTI UGUALMENTE DISTRIBUITI
E INDIPENDENTI

- 1928) B. de Finetti: Funzione caratteristica di un fenomeno aleatorio.
(Atti Congr. Int. dei Matematici, Bologna, 1928)
- 1929) B. de Finetti: Sulle funzioni ad incremento aleatorio. (Rend.
Acc. Naz. Lincei)
- 1929) B. de Finetti: Integrazione delle funzioni a incremento
aleatorio. (Rend. Acc. Naz. Lincei)
- 1929) B. de Finetti: Sulla possibilità di valori eccezionali per una
legge di incrementi aleatori. (Rend. Acc. Naz. Lincei)
- 1930) B. de Finetti: Le funzioni caratteristiche di legge istantanea.
(Rend. Acc. Naz. Lincei)
- 1931) B. de Finetti: Le funzioni caratteristiche di legge istantanea
dotate di valori eccezionali. (Rend. Acc. Naz.
Lincei)
- 1931) B. de Finetti: Le leggi differenziali e la rinuncia al
determinismo. (Univ. di Roma)
- 1932) A.N. Kolmogorov: Sulla forma generale di un processo
stocastico omogeneo (Un problema di
Bruno de Finetti). (Rend. Acc. Naz. Lincei)
- 1934-35) P. Lévy: Sur les intégrales dont les éléments sont des variables
aléatoires indépendantes. (Ann. Scuola Norm. Sup.
Pisa)
- 1937-1942) B. de Finetti ("Funzioni aleatorie", Congresso UMI, 1937)
W. Feller,
A.Ya. Khintchine,
K. Ito and others.

* Funzione caratteristica

$$\text{distrib. } F(x) \left\{ \begin{array}{l} E_F(e^{\xi x}) = \int_{\mathbb{R}} e^{\xi x} F_x(dx) \sim \text{moment gen. function} \\ E_F(e^{i\xi x}) = \int_{\mathbb{R}} [\cos \xi x + i \sin \xi x] F_x(dx) \quad \text{character. function} \\ \underbrace{\phantom{E_F(e^{i\xi x})}}_{\Phi_x(\xi)} \end{array} \right.$$

a) Φ continua in $\xi \in \mathbb{R}$

b) $\Phi(0) = 1$ e $|\Phi(\xi)| \leq 1$

c) $Y = aX + b \rightarrow \Phi_Y(\xi) = e^{ib\xi} \cdot \Phi_X(a\xi)$

d) Se $X \perp\!\!\!\perp Y \rightarrow \Phi_{X+Y}(\xi) = \Phi_X(\xi) \cdot \Phi_Y(\xi)$

e) $E(X^n) = i^{-n} \cdot \Phi_X^{(n)}(0)$

Alcuni esempi:

1. $X \sim N(\mu, \sigma^2) \rightarrow \Phi_X(\xi) = e^{i\mu\xi - \frac{\sigma^2}{2}\xi^2}$

2. $X \sim \text{Poisson}(\lambda) \rightarrow \Phi_X(\xi) = e^{\lambda(e^{i\xi} - 1)}$

3. $X \sim \text{Gamma}(\alpha, \beta) \rightarrow \Phi_X(\xi) = \left(1 - \frac{i\xi}{\beta}\right)^{-\alpha}$

4. $X_t \sim \text{Poisson composto} \rightarrow \Phi_{X_t}(\xi) = \exp\left\{\lambda t [\Phi_Y(\xi) - 1]\right\}$

5. $Z_t \sim \text{distrib. infinitamente divisibile}$, allora è

$$\Phi_{Z_t}(\xi) = \exp\left\{t \cdot \psi(\xi)\right\}$$

↑
funzione a valori complessi

* Distribuzioni infinitamente divisibili

$$F_X \sim \text{inf. div.} \begin{cases} \forall n, \exists X_{j_n} \text{ i.i.d.}, X \stackrel{d}{=} \sum_{j=1}^n X_{j_n} \\ \forall n, \exists F_n : F_X = (F_n)^{n^*} \\ \forall n, \exists \phi_n(\xi) : \phi_X(\xi) = [\phi_n(\xi)]^n \end{cases}$$

$F_X \sim \text{stabile} \Rightarrow F_X \sim \text{inf. div.}$ con $F_n \approx F_X$ e più precisamente $\exists \alpha, \gamma_n : \sum_{j=1}^n X_{j_n} \stackrel{d}{=} n^{1/\alpha} X + \gamma_n$

1) B. de Finetti (1930): rappresentazione generale della funzione caratteristica di una distr. inf. divisibile

$$\phi(\xi) = \exp \left\{ (i b \xi - \frac{c}{2} \xi^2) + \lambda \cdot [\phi_Y(\xi) - 1] \right\}$$

2) A.N. Kolmogorov (1932):

$$\phi(\xi) = \exp \left\{ (i a \xi + m) + \int_{\mathbb{R}} (e^{i \xi y} - 1 - i \xi y) \tilde{\mu}(dy) \right\}, \text{ ove } \int_{\mathbb{R}} y^2 \tilde{\mu}(dy) < \infty$$

3) P. Lévy (1934):

$$\phi(\xi) = \exp \left\{ (i a \xi + m) + \int_{\mathbb{R}} (e^{i \xi y} - 1 - i \xi y I_{[-1,1]}(y)) \nu(dy) \right\}$$

ove $\nu\{0\} = 0, \int_{\mathbb{R}} (y^2 \wedge 1) \nu(dy) < \infty, \nu(0, \infty) < \infty$

↑
misura di Lévy

$$X(t) = bt + B_c(t) + \sum_{s \leq t} \Delta X(s)$$

$$\Delta X(s) = X(s) - X(s^-) \neq X(s+s) - X(s)$$

Let $W_n = \sum_{j=1}^n T_j$, where $T_j \approx i.i.d. - Neg.exp(\lambda)$ be a random walk process.

1) Poisson process $\{N(t); t \geq 0\}$

a) $N(0) = 0$ and $N(t) = n$ iff $W_n \leq t < W_{n+1} \rightarrow$

$$\rightarrow P\{N(t) = n\} = e^{-\lambda t} \cdot (\lambda t)^n / n! \quad \Phi_{N(t)}(\xi) = \exp\{\lambda t (e^{i\xi} - 1)\};$$

b) It is a Lévy process characterized by $E[N(t)] = \lambda t$ and $\nu_t(dx) = \lambda t \cdot I_{\{1\}}(dx)$

2) Compound-Poisson process $\{X(t); t \geq 0\}$: it is a random walk $S_k =$

$= \sum_{j=1}^k Y_j$ with transition times given by the random walk process $\{W_n\}$.

a) $X(t) = \sum_{j=1}^{N(t)} Y_j = \sum_{n \geq 1} S_n \cdot |N(t) = n|$, where $Y_j / N(t) = n \approx i.i.d. - F_Y(dy) \rightarrow$

$$\rightarrow F_{X(t)}(x) = e^{-\lambda t} \cdot \sum_{n \geq 0} \frac{(\lambda t)^n}{n!} F_Y^{n*}(x) \quad , \quad \Phi_{X(t)}(\xi) = \exp\{\lambda t [\Phi_Y(\xi) - 1]\};$$

b) It is a Lévy process characterized by $E[X(t)] = \lambda t \cdot E(Y)$, $\nu_t(dx) = \lambda t \cdot F_Y(dx)$

Approximation theorem (1)

3) Pseudo-Poisson process $X(t) = \sum S_n \cdot |N(t) = n|$: it is a Markov

chain $\{S_k\}$ with transition times given by the random walk process $\{W_n\}$.

$$p_{ij}(t) = e^{-\lambda t} \cdot \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \cdot p_{ij}^n \quad \text{where } [p_{ij}^n] = P^n;$$

$$P\{X(t+s) \in \Gamma / X(s) = x\} = Q_t(x, \Gamma) = e^{-\lambda t} \cdot \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} K^{(n)}(x, \Gamma),$$

$$K^{(n)}(x, \Gamma) = \int_R K(x, dy) \cdot K^{(n-1)}(y, \Gamma).$$

Approximation theorem (2)

1) Exponential formula of semi-group of contractions theory

$$Q^*(t) = e^{-\lambda t} \cdot \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} (T^*)^n = e^{\lambda t (T^* - I)}; \quad (T^*)^{(n)} u(x) = \int_R K^{(n)}(x, dy) \cdot u(y).$$

$$\text{COMPOUND-POISSON process } X(t) \begin{cases} N(t) \approx \text{Poisson process } (\lambda t) \\ \text{random walk } S_k = \sum_1^k Y_j \\ \forall n, Y_j / N(t) = n \approx \text{i.i.d.} \end{cases}$$

An example:

$$f_Y(y) \approx \Gamma(1, \beta) \Rightarrow f_{X(t)}(x) = \sum_{n=0}^{\infty} P[N(t) = n] \cdot f_{X(t)/N(t)=n}(x) = \sum_{n=0}^{\infty} \left(\frac{e^{-\lambda t} \cdot (\lambda t)^n}{n!} \right) \cdot f_Y^{n*}(x),$$

where $f_Y^{n*}(x) \approx \Gamma(n, \beta)$, and this implies $E[X(t)] = E[N(t)] \cdot E(Y_j) = \lambda t / \beta$
and $Var[X(t)] = 2\lambda t / \beta^2$.

Maximum likelihood inferential procedure about the parameters λ and β

$$\ell(\lambda, \beta / \text{data}) \propto (\lambda \cdot \beta)^{\sum_{k=1}^m n_k} \cdot \exp \left\{ -m \cdot \lambda - \beta \cdot \sum_{k=1}^m \sum_{j=1}^{n_k} y_{jk} \right\}, \text{ where}$$

$m \approx$ number of the unitary periods of observation,

$n_k \approx$ number of arrivals in the k-th period of observation,

$\sum_{j=1}^{n_k} y_{jk} \approx$ cumulative effect in the k-th period of observation.

$$\text{Maximum likelihood estimates } \begin{cases} \hat{\lambda} = \sum_{k=1}^m n_k / m \\ \hat{\beta} = \sum_{k=1}^m n_k / \sum_{k=1}^m \sum_{j=1}^{n_k} y_{jk} \end{cases}$$

PSEUDO-POISSON process $X(t)$ $\left\{ \begin{array}{l} \text{discrete-time Markov chain (K)} \\ \text{Poisson process } (\alpha t) \end{array} \right.$

$$P\{X(t+s) \in \Gamma / X(s) = x\} = Q_t(x, \Gamma) = e^{-\alpha t} \cdot \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} K^{(n)}(x, \Gamma),$$

$$K^{(n)}(x, \Gamma) = \int_R K(x, dy) \cdot K^{(n-1)}(y, \Gamma), \quad n > 1; \quad K^{(1)}(x, \Gamma) = K(x, \Gamma).$$

$$Q_{t+\tau}(x, \Gamma) = \int_R Q_t(x, dy) Q_\tau(y, \Gamma)$$

As an example we may consider a particle travelling at uniform speed through homogeneous matter occasionally scoring a collision. Each collision produces a change of energy regulated by a stochastic kernel K . The transition probabilities have the above expression if the number of collisions obeys a Poisson process. It is usually assumed that the fraction of energy lost at each collision is independent of the initial amount. Let us first assume that the fraction of energy lost is uniformly distributed:

$$k(x, y) \equiv 1/x, \quad x > 0, \quad \Rightarrow \quad k^{(n)}(x, y) = \frac{1}{\Gamma(n) \cdot x} \left(\ln \frac{x}{y} \right)^{n-1}, \quad y \in (0, x);$$

$$r = \frac{y}{x} \approx f_n(r) = \frac{1}{\Gamma(n)} \left(\ln \frac{1}{r} \right)^{n-1}, \quad r \in (0, 1);$$

$$q_t(x, y) = \frac{\alpha t}{x \cdot e^{\alpha t}} \sum_{n=0}^{\infty} \frac{[\alpha t \cdot \ln(x/y)]^n}{n! \Gamma(n+2)} \quad \leftrightarrow \quad f_t(r) = \frac{\alpha t}{e^{\alpha t}} \sum_{n=0}^{\infty} \frac{[\alpha t \cdot \ln(1/r)]^n}{n! \Gamma(n+2)}$$

More generally: $k(x, y) \equiv \lambda \cdot y^{\lambda-1} / x, \quad \Rightarrow \quad k^{(n)}(x, y) = \frac{\lambda^n \cdot y^{\lambda-1}}{\Gamma(n) \cdot x^\lambda} \left(\ln \frac{x}{y} \right)^{n-1};$

$$q_t(x, y) = \frac{e^{-\alpha t} \cdot y^{\lambda-1} \cdot \sqrt{\lambda \cdot \alpha t}}{x^\lambda \cdot \sqrt{\ln(x/y)}} J_1 \left(2 \cdot \sqrt{\lambda \cdot \alpha t \cdot \ln(x/y)} \right); \quad I_1(x) = \sum_{n=0}^{\infty} \frac{(x/2)^{2n+1}}{n! \Gamma(n+2)}, \quad x \in R.$$

Theoretical hypothesis: if the initial energy may be considered infinitely large, then the total energy loss in $[0, t]$, $X(t)$, has a compound-Poisson distribution.

APPROXIMATION THEOREM (1)

Lemma: there is a one-to-one correspondence between Lévy processes $\{X(t)\}$ and infinitely divisible distributions $\{\mu\}$ through $F_{X(t)}(dx) = \mu(dx)$.
The process corresponding to μ is unique up to identity in law.

Theorem: every infinitely divisible distribution (Lévy process) is the limit of a sequence of compound-Poisson distributions (processes).

A short hint (B. de Finetti – Teoria delle Probabilità, II vol., p. 444):

- 1) let $\Phi(\xi)$ be the characteristic function of an infinitely divisible distribution and $\nu(dx)$ the Lévy measure of the Lévy process corresponding to $\Phi(\xi) \rightarrow \Phi^{1/n}(\xi)$ is a characteristic function for each n ;
 - 2) $\exp\{n.[\Phi^{1/n}(\xi) - 1]\}$ is the characteristic function of a compound-Poisson distribution for each n ;
 - 3) $\exp\{n.[\Phi^{1/n}(\xi) - 1]\} \xrightarrow{n \rightarrow \infty} \Phi(\xi)$ as $n.(x^{1/n} - 1) \xrightarrow{n \rightarrow \infty} \ln x$.
-

4) $\exp\{n.[\Phi^{1/n}(\xi) - 1]\} = \exp\left\{\int_R (e^{i\xi x} - 1) n.F_{1/n}(dx)\right\}$ and it can be proved that

$$\forall \text{ bounded, continuous } f(x): \int_R f(x) n.F_{1/n}(dx) \rightarrow \int_R f(x)\nu(dx).$$

Theorem (W. Feller): The following classes of prob. distr.s are identical

- 1) Infinitely divisible distributions
- 2) Distributions of the increments in Lévy processes
- 3) Limits of sequences of compound-Poisson distributions
- 4) Limits of sequences of infinitely divisible distributions
- 5) Limit distributions of row sums in triangular arrays $\{X_{k,n}\}$ where the variables $X_{k,n}$ of the n -th row are i.i.d.

CONTINUOUS SEMI-GROUPS OF TRANSITION OPERATORS

Continuous-time Markov process $\rightarrow P\{X(t+s) \in \Gamma / X(s) = x\} = D_t(x, \Gamma)$

$$D_{t+\tau}(x, \Gamma) = \int_R D_t(x, dy) D_\tau(y, \Gamma);$$

$V \approx$ real valued bounded continuous functions with $\|u(x)\| = \sup|u(x)|$:

$$V \xrightarrow{\Delta(t)} V : \Delta(t)u(x) = \int_R u(y) D_t(x, dy) = E\{u[X(t+\tau)] / X(\tau) = x\}$$

Continuous semi-group of contractions: $\{\Delta(t); t \geq 0\}$ $\left\{ \begin{array}{l} \Delta(0) = \mathfrak{I} \\ \Delta(s+t) = \Delta(s)\Delta(t) \\ \|\Delta(t)u\| \leq \|u\| \\ \Delta(t) \xrightarrow{t \downarrow 0} \mathfrak{I} \end{array} \right.$

Generator A of $\{\Delta(t); t \geq 0\}$: $Au(x) = \lim_{t \downarrow 0} \frac{\Delta(t)u(x) - u(x)}{t}$

a) bounded A $\rightarrow \Delta(t)u(x) = \exp[t.A]u(x) = \sum_{n=0}^{\infty} \frac{t^n . A^n u(x)}{n!}$

Some examples $\left\{ \begin{array}{l} \text{Compound - Poisson process} \\ \text{Pseudo - Poisson process} \end{array} \right.$

b) unbounded A \rightarrow it does not exist the exponential representation

as for a general Feller - Markov process, but

$$\Delta_h(t) = e^{-t/h} \cdot \sum_{n=0}^{\infty} \frac{(t/h)^n}{n!} \Delta(nh) = \exp\left\{\frac{t}{h} [\Delta_h - \mathfrak{I}]\right\} \xrightarrow{h \rightarrow 0^+} \Delta(t)$$

Transition operator of the pseudo-Poisson process

APPROXIMATION THEOREMS

1) $\{X(t); t \geq 0\} \approx \text{Lévy process with } E[e^{i\xi X(1)}] = \Phi(\xi)$

$\Phi(\xi) \approx \text{inf. div.} \Rightarrow \Phi^{1/n}(\xi) \approx \text{characteristic function}$

comp. Poisson $\{X_n(t); t \geq 0\} \approx \Phi_{X_n(1)}(\xi) = \exp\{n. [\Phi^{1/n}(\xi) - 1]\}$

It can be proved that:

$$\exp\{n. [\Phi^{1/n}(\xi) - 1]\} \xrightarrow{n \rightarrow \infty} \Phi(\xi)$$

2) $\{X(t); t \geq 0\} \approx \text{Feller - Markov pr. with } \{D_t(x, \Gamma); t \geq 0\}$

$$\Delta(t)u(x) = \int_R u(y) D_t(x, dy) = E\{u[X(t+\tau)] / X(\tau) = x\} \left\{ \begin{array}{l} \Delta(0) = \mathfrak{I} \\ \Delta(s+t) = \Delta(s)\Delta(t) \\ \|\Delta(t)u\| \leq \|u\| \\ \Delta(t) \xrightarrow{t \downarrow 0} \mathfrak{I} \end{array} \right.$$

$$u(x) = I_\Gamma(x) \rightarrow \Delta(t)I_\Gamma(x) = D_t(x, \Gamma) = \Pr\{X(t+\tau) \in \Gamma / X(\tau) = x\};$$

pseudo-Poisson process $\{X_h(t); t \geq 0\}$ with transition operator $\Delta_h(t)$:

$$\Delta_h(t) = e^{-t/h} \sum_{n=0}^{\infty} \frac{(t/h)^n}{n!} \Delta(nh).$$

It can be proved that:

$$\Delta_h(t) = e^{-t/h} \sum_{n=0}^{\infty} \frac{(t/h)^n}{n!} \Delta(nh) = \exp\left\{\frac{t}{h} [\Delta_h - \mathfrak{I}]\right\} \xrightarrow{h \rightarrow 0^+} \Delta(t)$$

Some inferential aspects

Compound-Poisson processes

$$X(t) \sim \text{compound Poisson pr. if } \Rightarrow X(t) = \sum_{j=1}^{N(t)} Y_j \quad \begin{cases} N(t) \approx \text{Poisson}(\lambda t) \\ f_Y \approx \Gamma(1, \beta) \end{cases}$$

$$f_{X(t)}(x) = \sum_{n=0}^{\infty} P[N(t) = n] \cdot f_{X(t)/N(t)=n}(x) = \sum_{n=0}^{\infty} \left(\frac{e^{-\alpha t} \cdot (\alpha t)^n}{n!} \right) \cdot f_Y^{n*}(x),$$

$$f_Y^{n*}(x) \approx \Gamma(n, \beta) \quad \begin{cases} E[X(t)] = E[N(t)] \cdot E(Y_j) = \lambda t / \beta \\ \text{Var}[X(t)] = 2\lambda t / \beta^2 \end{cases}$$

1. Maximum likelihood inferential procedure about the parameters α and β

$$\ell(\alpha, \beta / \text{data}) \propto (\alpha \cdot \beta)^{\sum_{k=1}^m n_k} \cdot \exp \left\{ -m \cdot \alpha - \beta \cdot \sum_{k=1}^m \sum_{j=1}^{n_k} y_{jk} \right\}$$

$m \sim$ number of the unitary periods of observation,

$n_k \approx$ number of arrivals in the k -th period of observation,

$\sum_{j=1}^{n_k} y_{jk} \approx$ cumulative effect in the k -th period of observation.

$$\text{Maximum likelihood estimates} \quad \left\{ \begin{array}{l} \hat{\alpha} = \sum_{k=1}^m n_k / m \\ \hat{\beta} = \sum_{k=1}^m n_k / \sum_{k=1}^m \sum_{j=1}^{n_k} y_{jk} \end{array} \right.$$

* Stime di massima verosimiglianza per un processo osservabile di tipo Poisson-composto.

1) Osservazioni statistiche per ogni periodo di tempo unitario:

n_k : numero di eventi osservati nel periodo k -esimo;

$(y_{1,k}, y_{2,k}, \dots, y_{n_k,k})$: effetti osservati per ogni evento del periodo k -esimo;

m : periodi di osservazione.

$$\begin{aligned} \ell(\lambda, \beta | n_k; y_{1,k}, \dots, y_{n_k,k}) &= \ell_1(\lambda | \text{dati}) \cdot \ell_2(\beta | \text{dati}) \\ &= P(N_k = n_k | \lambda) \cdot f(y_{1,k}, \dots, y_{n_k,k} | \beta) = \\ &= \left(e^{-\lambda} \cdot \frac{\lambda^{n_k}}{n_k!} \right) \cdot \left[\prod_{i=1}^{n_k} \left(\beta e^{-\beta y_{i,k}} \right) \right] = \\ &= \left(\text{idem} \right) \cdot \left(\beta^{n_k} \cdot e^{-\beta \sum_{i=1}^{n_k} y_{i,k}} \right) \propto \\ &\propto \left(e^{-\lambda} \cdot \lambda^{n_k} \right) \left(\beta^{n_k} \cdot e^{-\beta \cdot \sum y_{i,k}} \right) \end{aligned}$$

2) Per tutti gli m periodi di osservazione: posto $\underline{y}_k = \begin{bmatrix} y_{1,k} \\ \vdots \\ y_{n_k,k} \end{bmatrix}$

$$\begin{aligned} \text{si ha } \ell(\lambda, \beta | n_1, \underline{y}_1; n_2, \underline{y}_2; \dots; n_m, \underline{y}_m) &= \\ &= \prod_{k=1}^m \ell(\lambda, \beta | n_k, \underline{y}_k) \propto \prod_{k=1}^m \left[\left(e^{-\lambda} \cdot \lambda^{n_k} \right) \left(\beta^{n_k} \cdot e^{-\beta \sum_{i=1}^{n_k} y_{i,k}} \right) \right] \\ &= \left(\lambda \cdot \beta \right)^{\sum_{k=1}^m n_k} \cdot \exp \left\{ -m \cdot \lambda - \beta \sum_{k=1}^m \sum_{j=1}^{n_k} y_{j,k} \right\} \end{aligned}$$

$$\text{Stime per } \lambda \text{ e } \beta \left\{ \begin{array}{l} \hat{\lambda} = \frac{1}{m} \sum_{k=1}^m n_k, \\ \hat{\beta} = \frac{\sum_{k=1}^m n_k}{\sum_{k=1}^m \sum_{j=1}^{n_k} y_{j,k}}. \end{array} \right.$$

2. Bayesian inference about the parameters α and β .

Observables:
$$S(t) = X(t) - X(t-1) = \sum_{j=1}^{N(t)-N(t-1)} Y_{tj}$$

where Y_{tj} is the j -th effect in the time interval $[t-1, t]$.

From a Bayesian viewpoint, when the parameters α and β are unknown:

- 1) the random effects Y_j and the increments of the Poisson process have to be considered *exchangeable*, that is independent conditional on every values of α and β ; moreover the two processes are to be considered mutually independent conditional on every values of α and β ;
- 2) every information which is available prior to observe the data has to be expressed by means of a joint distribution $G(\alpha, \beta / H)$, where H denotes such an information;
- 3) the pooling of the prior and the data information, which we will denote by K , has to take place through the application of the Bayes theorem:

$$dG(\alpha, \beta / H \wedge K) \propto \ell(\alpha, \beta / K) dG(\alpha, \beta / H).$$

We will consider two different situations:

- a) the statistician has no reason to assume that the effects process $\{Y_j; j \geq 1\}$ and the Poisson count process $\{N(t); t \geq 0\}$ are connected in some way so that he assume α and β to be **stochastically independent**: $G(\alpha, \beta / H) = G_1(\alpha / H).G_2(\beta / H)$;
- b) the statistician believes that there is some connection between the two processes so that he assume α and β to be **stochastically dependent**: $G(\alpha, \beta / H) \neq G_1(\alpha / H).G_2(\beta / H)$.

In the first case, sub a):

$$G(\alpha, \beta / H) = G_1(\alpha / H) \cdot G_2(\beta / H) \quad \begin{cases} G_1(\alpha / H) \approx \Gamma(\alpha / a, b) \\ G_2(\beta / H) \approx \Gamma(\beta / c, d) \end{cases}$$

$$g(\alpha, \beta / H) = g_1(\alpha / H) \cdot g_2(\beta / H) \propto [\alpha^{a-1} \cdot e^{-b \cdot \alpha}] \cdot [\beta^{c-1} \cdot e^{-d \cdot \beta}] .$$

Data:

$m \approx$ number of the unitary periods of observation,

$n_k \approx$ number of arrivals in the k-th period of observation $\rightarrow n = \sum_{k=1}^m n_k$

$y_k^* = \sum_{j=1}^{n_k} y_{jk} \approx$ cumulative effect in the k-th period of observation $\rightarrow y^* = \sum_{k=1}^m y_k^*$

$$g(\alpha, \beta / H \wedge K) \propto l(\alpha, \beta / K) \cdot g(\alpha, \beta / H) \propto$$

$$\propto (\alpha \cdot \beta)^n \cdot \exp\{-m \cdot \alpha - \beta \cdot y^*\} \cdot [\alpha^{a-1} \cdot \beta^{c-1} \cdot \exp\{-(b \cdot \alpha + d \cdot \beta)\}] =$$

$$= [\alpha^{a+n-1} \cdot e^{-\alpha(b+m)}] \cdot [\beta^{c+n-1} \cdot e^{-\beta(d+y^*)}]$$

Bayesian estimates: $E(\alpha / H \wedge K) = \frac{a+n}{b+m}$ and $E(\beta / H \wedge K) = \frac{c+n}{d+y^*}$.

Obviously, a full exploitation of the posterior joint distribution is given by the determination of the expectations of given integrable functions of the two parameters:

$$E\{f(\alpha, \beta) / H \wedge K\} = \iint_{R^2} f(\alpha, \beta) \cdot g(\alpha, \beta / H \wedge K) d\alpha d\beta .$$

In the second case, sub b): $G(\alpha, \beta / H) \neq G_1(\alpha / H).G_2(\beta / H)$

$$g(\alpha, \beta / H) = \sum_{i=1}^k u_i \cdot \Gamma(\alpha / a_i, b_i) \cdot \Gamma(\beta / c_i, d_i) ; \quad u_i \geq 0, \quad \sum_{i=1}^k u_i = 1,$$

Suppose $k = 2$:

$$\text{Cov}(\alpha, \beta / H) = (u_1 - u_1^2) \cdot \frac{a_1}{b_1} \cdot \frac{c_1}{d_1} + (u_2 - u_2^2) \cdot \frac{a_2}{b_2} \cdot \frac{c_2}{d_2} - u_1 u_2 \left(\frac{a_1}{b_1} \cdot \frac{c_2}{d_2} + \frac{a_2}{b_2} \cdot \frac{c_1}{d_1} \right),$$

so that the correlation between α, β may be either positive or negative.

For simplicity, we assume:

$$b_1 = b_2 = b, \quad d_1 = d_2 = d, \quad a_2 = a_1 + r, \quad c_2 = c_1 + s; \quad r, s \in R, \Rightarrow$$

$$\Rightarrow \text{Cov}(\alpha, \beta / H) = u_1 u_2 \cdot \frac{r}{b} \cdot \frac{s}{d}.$$

It can be proved that the correlation between the parameters implies a correlation between the effects process $\{Y_j; j \geq 1\}$ and the Poisson count process $\{N(t); t \geq 0\}$; more precisely we have

$$\text{Cov}(N_k, Y_{kj} / H) = -\frac{d^2}{c(c+s)} \text{Cov}(\alpha, \beta / H).$$

The opinion which is expressed by means of the preceding assumptions corresponds to a compound-Poisson scheme constituted by a *partially-exchangeable process*, that is by the exchangeable processes $\{Y_j; j \geq 1\}$ and $\{N_k; k \geq 0\}$ which are mutually correlated. By the Bayes theorem:

$$g(\alpha, \beta / H \wedge K) \propto g(\alpha, \beta / H) \cdot l(\alpha, \beta / K) \propto$$

$$\propto \left[\sum_i u_i \cdot \Gamma(\alpha; a_i, b_i) \cdot \Gamma(\beta; c_i, d_i) \right] \cdot \left[(\alpha \cdot \beta)^n \cdot e^{-m\alpha - y \cdot \beta} \right] \propto$$

$$\propto \left[\sum_i u_i \cdot (\alpha^{a_i-1} \cdot e^{-\alpha b_i}) \cdot (\beta^{c_i-1} \cdot e^{-\beta d_i}) \right] \cdot \left[(\alpha \cdot \beta)^n \cdot e^{-m\alpha - y \cdot \beta} \right] \propto$$

$$\propto \sum_i u_i' \cdot (\alpha^{a_i+n-1} \cdot e^{-\alpha(b_i+m)}) \cdot (\beta^{c_i+n-1} \cdot e^{-\beta(d_i+y)})$$

ALCUNI RIFERIMENTI BIBLIOGRAFICI

B. de Finetti, (1970). **TEORIA DELLE PROBABILITA**. Einaudi ed.

W. Feller, (1971). **AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS**. J. Wiley, II ed.

O. Kallenberg, (1997). **FOUNDATIONS OF MODERN PROBABILITY**
Springer.

J.Lamperti, (1977). **STOCHASTIC PROCESSES**. Springer-Verlag.

P. Lévy, (1965). **PROCESSUS STOCASTIQUES**. Gauthier-Villars.

K. Sato, (1999). **LEVY PROCESSES AND INFINITELY DIVISIBLE DISTRIBUTIONS**. Cambridge University Press.

A.N. Shiryaev (1999). **ESSENTIALS OF STOCHASTIC FINANCE**.
World Scientific.

D. Applebaum, “Lévy processes – From Probability to Finance and Quantum Groups”

M. Keller – Ressel, “An intuitive introduction to operator semi-groups”

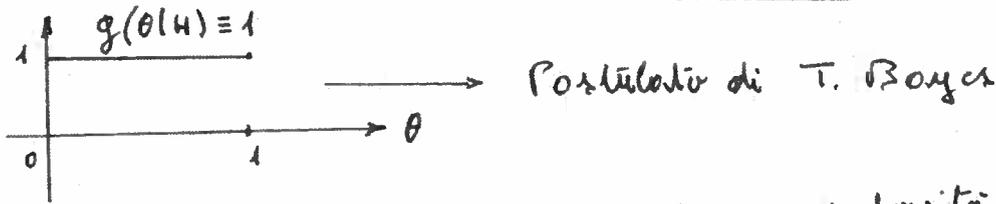
Thomas Bayes (1702 - 1761)

 $\{E_i\}, i = 1, 2, \dots$ Estensioni successive di probaie con membership elemento ($\Theta = \omega + t$)

Θ $\left\{ \begin{array}{l} \text{noto} \xrightarrow{\theta^*} \{E_i\} \text{ bernoulliano con } P(E_i) = \theta^*, \forall i \\ \text{non noto} \rightarrow \{E_i\} \text{ scambiate con } P(E_i | \Theta = \theta) = \theta, \forall i \end{array} \right.$

Se $K = \{S_{p+q} = \sum_{i=1}^{p+q} |E_i| = p\}$ qual'è $P(E_{p+q+1} | S_{p+q} = p)$?

$H = \phi: P(E_{p+q+1} | S_{p+q} = p) = \frac{p+1}{p+q+2} \neq \frac{p}{p+q} = \hat{\theta}$



Teorema di Bayes per densità di prob.

Dimostrazione: $g(\theta|H \wedge K) \propto g(\theta|H) \cdot l(\theta|K)$

ove $\left\{ \begin{array}{l} g(\theta|H) \sim \text{Beta}(\alpha = \beta = 1) \rightarrow g(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ l(\theta|K) = f P(K|\theta) = \theta^p (1-\theta)^q \end{array} \right.$

$g(\theta|H \wedge K) \propto g(\theta|H) \cdot l(\theta|K) = \theta^p (1-\theta)^q$

Beta(p+1, q+1)

$E(\theta|H \wedge K) = \int_0^1 \theta \cdot g(\theta|H \wedge K) d\theta = \frac{p+1}{p+q+2}$

$\frac{\alpha}{\alpha + \beta}$

3) Analisi statistica bayesiana su proporzioni

Prendiamo in considerazione il problema dell'inferenza statistica sulla **proporzione incognita Θ degli "individui" di una collettività aventi una fissata caratteristica di interesse**: può trattarsi di lampadine elettriche prodotte in un dato stabilimento alcune delle quali presentano un certo difetto di fabbricazione o di esseri umani ricoverati in una data struttura ospedaliera alcuni dei quali hanno il gruppo sanguigno di tipo RH negativo o altro.

Allo scopo di costruire uno schema generale **gli "individui" verranno assimilati a palline contenute in un'urna, alcune di colore bianco e altre di colore nero, e si supporrà di indagare sulla proporzione di bianche** procedendo ad esperimenti consistenti in estrazioni successive dall'urna di una pallina alla volta, con reimbussolamento della pallina estratta. In questo modo **la composizione dell'urna (cioè la percentuale di palline bianche) rimane la stessa dopo ogni estrazione**. Indicheremo con $\{X_t ; t \geq 1\}$ il processo di osservazione, ove X_t rappresenta l'indicatore dell'evento $E_t =$ "il risultato della t-esima estrazione è una pallina bianca". Supporremo, in generale, di non conoscere il numero totale delle palline nell'urna e di poter effettuare un fissato numero m di estrazioni.

Il modello statistico accolto è quindi costituito dall'ipotesi di scambiabilità del processo di osservazione $\{X_t ; t \geq 1\}$, che consente di ritenere le X_t condizionatamente i.i.d., e dalle scelte della comune distribuzione condizionata $P(X_t = x / \Theta = \theta) = \theta^x \cdot (1 - \theta)^{1-x}$ (detta distribuzione di Bernoulli) e della distribuzione iniziale di tipo Beta con densità $g_0(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \theta^{\alpha-1} \cdot (1 - \theta)^{\beta-1}$ per la proporzione incognita Θ . Tale modello è indicato in letteratura come "schema Bernoulli - Beta".

Immaginando di effettuare m estrazioni dall'urna e di ottenere s palline bianche ed $m - s$ nere, disposte in un qualche ordine, la corrispondente verosimiglianza per la proporzione incognita Θ sarebbe data dalla

$$\ell(\theta / x_1, \dots, x_m) = \prod_{j=1}^m \ell(\theta / x_j) = \theta^s \cdot (1 - \theta)^{m-s}.$$

Formalmente, la verosimiglianza coincide con la probabilità condizionata dell'evento $K = \{ \sum_{i=1}^m X_i = s \}$ implicato dalla sequenza x_1, \dots, x_m dei risultati dell'osservazione; il condizionamento fa riferimento all'ipotesi generica $\Theta = \theta$.

Incominciando con l'**approccio inferenziale indiretto**, l'applicazione del teorema di Bayes alla densità iniziale è espressa dalla trasformazione:

$$g(\theta / K) \propto g_0(\theta) \cdot [\theta^s \cdot (1 - \theta)^{m-s}] \propto \frac{\Gamma(\alpha + \beta + m)}{\Gamma(\alpha + s) \Gamma(\beta + m - s)} \theta^{\alpha+s-1} \cdot (1 - \theta)^{\beta+m-s-1},$$

che fornisce come distribuzione finale per Θ ancora una Beta con parametri

$$\alpha^* = \alpha + s \quad e \quad \beta^* = \beta + m - s.$$

In base alla densità finale $g(\theta / K)$ la valutazione della probabilità di ottenere una pallina bianca alla $(m+1)$ -esima estrazione è determinabile secondo la

$$P(E_{m+1} / K) = E(\Theta / K) = \frac{\alpha^*}{(\alpha^* + \beta^*)} = \frac{\alpha + s}{\alpha + \beta + m} = \frac{(\alpha + \beta) \cdot \frac{\alpha}{\alpha + \beta} + m \cdot \frac{s}{m}}{(\alpha + \beta) + m}.$$

Aggiungiamo che le probabilità finali $P\{\sum_{t=m+1}^{m+n} X_t = n/K\}$ risultano determinate dalla relazione:

$$P\{\sum_{t=m+1}^{m+n} X_t = n/K\} = E(\Theta^n / K) = [(\alpha^*)_n / (\alpha^* + \beta^*)_n],$$

ove $(\alpha^*)_n = \alpha^* \cdot (\alpha^* + 1) \cdot (\alpha^* + 2) \cdot \dots \cdot (\alpha^* + n - 1)$ è il simbolo di Pochhammer .

Poiché la legge temporale iniziale del processo scambiabile di indicatori di eventi $\{X_t; t \geq 1\}$ è rappresentabile mediante la successione, al variare di n , delle probabilità $\{(\alpha)_n / (\alpha + \beta)_n\}$ e la legge finale è rappresentata dalla analoga successione $\{(\alpha^*)_n / (\alpha^* + \beta^*)_n\}$ viene spontaneo chiedersi quale trasformazione porta dalla prima alla seconda e tale questione costituisce l'essenza del **procedimento bayesiano diretto**.

Indicando, per semplicità, le probabilità $P\{\sum_{t=1}^n X_t = n\}$ col simbolo $\omega(n, n)$ e le probabilità $P\{\sum_{t=1}^n X_t = r\}$, $r \leq n$, col simbolo $\omega(n, r)$, si può facilmente dimostrare che sussiste la seguente relazione tra esse:

$$\omega(n, r) = \binom{n}{r} \cdot (-1)^{n-r} \cdot \Delta^{n-r} \omega(r, r),$$

ove con $\Delta^{n-r} \omega(r, r)$ si è indicata la differenza di ordine $n - r$ per le $\omega(r, r)$ avente l'espressione generale

$$\Delta^{n-r} \omega(r, r) = \sum_{j=0}^{n-r} \binom{n-r}{j} \cdot (-1)^{n-r-j} \omega(r+j, r+j).$$

Si ottiene allora che in corrispondenza alle $\omega(r, r) = [(\alpha)_r / (\alpha + \beta)_r]$ sussistono le :

$$\omega(n, r) = \binom{n}{r} \cdot \frac{(\alpha)_r \cdot (\beta)_{n-r}}{(\alpha + \beta)_n}.$$

Considerato ora un generico evento $A = (S_k = k)$, riguardante k eventi distinti tra loro e dagli m eventi osservati, si ottiene:

$$P\{A / K\} = P(K \wedge A) / P(K) = \{[\omega(m+k, s+k) / \binom{m+k}{s+k}]\} / \{[\omega(m, s) / \binom{m}{s}]\} =$$

$$= \left[\frac{(\alpha)_{s+k} \cdot (\beta)_{m-s}}{(\alpha + \beta)_{m+k}} \right] / \left[\frac{(\alpha)_s \cdot (\beta)_{m-s}}{(\alpha + \beta)_m} \right] = \frac{(\alpha)_{s+k} \cdot (\alpha + \beta)_m}{(\alpha)_s \cdot (\alpha + \beta)_{m+k}} = \frac{(\alpha + s)_k}{(\alpha + \beta + m)_k}$$

e, chiaramente, l'ultimo rapporto coincide con il momento k-esimo di una densità di tipo Beta con parametri $\alpha+s$ e $\beta+m-s$.

Si è verificata così l'equivalenza del procedimento inferenziale diretto con quello indiretto in quanto quest'ultimo avrebbe fornito la suddetta probabilità

$P\{A / K\}$ mediante la

$$P\{A / K\} = \int_0^1 \theta^k \cdot g(\theta / K) d\theta = \frac{(\alpha + s)_k}{(\alpha + \beta + m)_k} .$$

E' bene ripetere che entrambi i procedimenti inferenziali descrivono il passaggio da una legge temporale iniziale $\Lambda(H)$ per il processo di osservazione scambiabile

$\{X_t ; t \geq 1\}$ ad una nuova legge temporale $\Lambda(H \wedge K)$, che denomineremo "finale", per il processo residuo $\{X_t ; t \geq m+1\}$ costituito dalle variabili osservabili non ancora osservate.

Nel problema inferenziale finora considerato, che potremmo denominare "apprendimento statistico su proporzioni", si è specificata una legge temporale iniziale $\Lambda(H)$ costituita dalla successione delle probabilità $\{\omega(n, n), n \geq 1\}$; la sua scelta deve fondarsi sullo stato di informazione iniziale H che può essere più o meno ampio. Nell'esempio da noi considerato si è scelta la successione $\omega(n, n) = (\alpha)_n / (\alpha + \beta)_n$ dei momenti di una densità Beta (α, β) , con parametri fissati.

La particolare densità Beta con parametri $\alpha = \beta = 1$, uniforme sull'intervallo $(0, 1)$, è detta "non informativa" in quanto assegna ad ogni sub-intervallo di $(0, 1)$ avente lunghezza λ una stessa probabilità, appunto λ , indipendentemente dalla sua collocazione all'interno dell'intervallo $(0, 1)$. In tal modo, si afferma, non si esprimerebbe alcuna opinione sul valore del parametro incognito Θ o, equivalentemente, sulla probabilità subordinata $P\{X_i = 1 / \Theta = \theta\} = \theta$. Alla densità Beta uniforme corrisponde la successione dei momenti $\omega(n, n) = (n + 1)^{-1}$ ed è facile verificare che risulta anche $\omega(n, m) = (n + 1)^{-1}$, per ogni $m \leq n$. Per quanto si è già visto, in presenza dell'incremento di informazione K , l'applicazione del teorema di Bayes a tale densità porta ad una densità finale Beta $(s+1, m-s+1)$ il cui valor medio è

$$E(\Theta / K) = \frac{s+1}{m+2} \cong \frac{s}{m},$$

praticamente uguale alla stima di massima verosimiglianza, s/m , del parametro Θ .

4) Analisi statistica bayesiana su processi di conteggio

Consideriamo ora **processi di osservazione concernenti sequenze di eventi E_j che si verificano in epoche aleatorie**. L'analisi probabilistica della sequenza di eventi $\{E_j; j \geq 1\}$ può essere condotta da due distinti punti di vista tra loro complementari: si può considerare

- 1) il processo stocastico di conteggio a parametro continuo $\{N(t); t \geq 0\}$, ove $N(t)$ è il numero aleatorio di eventi arrivati nell'intervallo di tempo $(0, t]$, oppure
- 2) il processo stocastico a parametro discreto dei tempi di attesa tra arrivi successivi $\{T_j; j \geq 1\}$, ove T_j misura l'ampiezza dell'intervallo aleatorio di tempo che intercorre tra l'arrivo di E_{j-1} e di E_j .

Nei successivi procedimenti inferenziali bayesiani assumeremo che il processo di osservazione sia costituito da un processo di conteggio di tipo **poissoniano** con intensità Θ non nota del quale sono osservabili gli incrementi N_h riguardanti intervalli di tempo unitari successivi; per ogni possibile valore θ dell'intensità non nota Θ , assumeremo $P\{N_h = n / \Theta = \theta\} = e^{-\theta} \cdot \theta^n / n!$ e i n.a. N_h essere i.i.d.

Per i tempi di attesa T_j tra arrivi successivi assumeremo che per ogni possibile valore θ dell'intensità non nota Θ , essi siano i.i.d. con la comune densità di probabilità condizionata $f(t / \theta) = \theta \cdot \exp(-\theta t)$.

In entrambi i casi avremo dunque **processi di osservazione a parametro discreto e scambiabili** in quanto i n.a. osservabili N_h e, rispettivamente, T_j sono assunti essere i.i.d. subordinatamente ad ogni ipotesi $\Theta = \theta$. I modelli statistici parametrici che discendono dalle precedenti ipotesi e che riguardano, per esempio, m' e, rispettivamente, m'' osservazioni sono i seguenti:

a) per il processo di conteggio:
$$P\left\{\bigcap_{h=1}^{m'} (N_h = n_h) / \theta\right\} = \frac{\theta^{\sum_h n_h} \cdot e^{-m' \cdot \theta}}{\prod_h (n_h!)},$$

b) per il processo dei tempi di attesa:
$$f(t_1, \dots, t_{m''} / \theta) = \theta^{m''} \cdot e^{-\theta \cdot \sum_j t_j},$$

che si riferiscono agli eventi $K_1 = \bigcap_{h=1}^{m'} (N_h = n_h)$ e, rispettivamente, $K_2 = \bigcap_{j=1}^{m''} (T_j = t_j)$.

E' facile ricavare le stime di massima verosimiglianza per θ corrispondenti agli eventi osservabili K_1 e K_2 ; si ottengono precisamente i valori $\hat{\theta} = \frac{\sum n_h}{m'}$ e, rispettivamente, $\hat{\theta} = \frac{m''}{\sum t_h}$.

Se, nell'approccio bayesiano dell'inferenza statistica su Θ per entrambi i modelli statistici assumiamo che la distribuzione iniziale $G_0(\theta)$ sia di tipo Gamma (α, β) con densità $g_0(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$, avente parametri numericamente fissati, allora a tali modelli parametrici corrispondono le seguenti leggi temporali iniziali $\Lambda(H)$:

$$a) P \left\{ \bigcap_{h=1}^{m'} (N_h = n_h) \right\} = \frac{\Gamma(\alpha + \sum n_h)}{\Gamma(\alpha) \cdot [\prod (n_h!)]} \cdot \frac{\beta^\alpha}{(\beta + m')^{\alpha + \sum n_h}}, \quad m' \geq 1;$$

$$b) f(t_1, \dots, t_{m''}) = \frac{\Gamma(\alpha + m'')}{\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{(\beta + \sum t_j)^{\alpha + m''}}, \quad m'' \geq 1.$$

Inferenza statistica sul processo $\{N_h; h \geq 1\}$

Descriveremo per primo il procedimento Bayesiano, che chiameremo **indiretto**, consistente nell'adeguamento della distribuzione iniziale $G_0(\theta)$ del parametro incognito all'incremento di informazione osservato (o osservabile) $K_1 = \bigcap_{h=1}^{m'} (N_h = n_h)$ mediante l'applicazione del teorema di Bayes. Questa fornisce la densità finale:

$$g(\theta / K_1) \propto g_0(\theta) \cdot \ell(\theta / K_1) \propto \theta^{(\alpha + \sum n_h) - 1} e^{-\theta(\beta + m')}$$

che corrisponde ad una distribuzione Gamma (α^*, β^*) con parametri $\alpha^* = \alpha + \sum n_h$ e $\beta^* = \beta + m'$.

E' detta "stima Bayesiana per Θ " un qualunque indice di posizione di questa distribuzione; scegliendo come stima bayesiana il valor medio della distribuzione finale e indicando per comodità $\sum_{h=1}^{m'} n_h = n$ si ha:

$$E(\Theta / K_1) = \frac{\alpha + n}{\beta + m'} = \frac{\beta \cdot \frac{\alpha}{\beta} + m' \cdot \frac{n}{m'}}{\beta + m'}$$

e dall'ultimo rapporto si osserva che $E(\Theta / K_1)$ è espresso da una combinazione lineare convessa del valor medio iniziale $E(\Theta) = \alpha / \beta$ e della stima di massima verosimiglianza n / m' .

Al crescere del numero m' dei periodi di osservazione, cioè al crescere dell'informazione campionaria, la stima bayesiana tende a coincidere con quella di massima verosimiglianza.

Se fossimo interessati a fare una qualunque previsione sui valori delle variabili non ancora osservate N_h , $h > m'$, potremmo ricorrere al teorema di rappresentazione per processi scambiabili al modo seguente.

Limitando la previsione alla sequenza $N_{m'+1}, \dots, N_{m'+k}$ e indicando con (v_1, \dots, v_k) un insieme arbitrario dei loro possibili valori si ha:

$$\begin{aligned} P\left\{\bigcap_{i=1}^k (N_{m'+i} = v_i) / K_1\right\} &= \int_{\{\theta\}} \left[\prod_{i=1}^k P[(N_{m'+i} = v_i) / \theta] \right] \cdot g(\theta / K_1) d\theta = \\ &= \frac{\Gamma(\alpha + \sum n_h + \sum v_i)}{\Gamma(\alpha + \sum n_h) \cdot [\prod (v_i!)]} \cdot \frac{(\beta + m')^{\alpha + \sum n_h}}{(\beta + m' + k)^{\alpha + \sum n_h + \sum v_i}} \end{aligned}$$

Il procedimento inferenziale Bayesiano, che chiameremo diretto, diretto conduce a questa stessa espressione della distribuzione previsionale per le variabili $N_{m'+1}, \dots, N_{m'+k}$ mediante l'applicazione del teorema delle probabilità composte alle distribuzioni della legge temporale iniziale $\Lambda(H)$

$$P\left\{\bigcap_{h=1}^m (N_h = n_h) / H\right\} = \frac{\Gamma(\alpha + \sum n_h)}{\Gamma(\alpha) \cdot [\prod (n_h!)]} \cdot \frac{\beta^\alpha}{(\beta + m')^{\alpha + \sum n_h}}, \quad m' \geq 1.$$

Infatti si ha, come facilmente si verifica :

$$P \left\{ \bigcap_{i=1}^k (N_{m+i} = \nu_i) / H \cap K_1 \right\} = \frac{P \left\{ K_1 \cap \left[\bigcap_{i=1}^k (N_{m+i} = \nu_i) \right] / H \right\}}{P \{ K_1 / H \}} =$$

$$= \frac{\Gamma(\alpha + \sum n_h + \sum \nu_i)}{\Gamma(\alpha + \sum n_h) \cdot \left[\prod (\nu_i!) \right]} \cdot \frac{(\beta + m')^{\alpha + \sum n_h}}{(\beta + m' + k)^{\alpha + \sum n_h + \sum \nu_i}}.$$

Inferenza statistica sul processo $\{ T_j; j \geq 1 \}$.

Ora l'incremento di informazione sul processo degli arrivi sarà costituito dall'evento $K_2 = \bigcap_{j=1}^{m''} (T_j = t_j)$ riguardante l'osservazione di m'' tempi di attesa tra arrivi successivi T_j : in tal caso si osservano m'' eventi per un tempo totale pari a $\sum_{j=1}^{m''} t_j = t$.

La corrispondente funzione di verosimiglianza è espressa dalla

$$\ell(\theta / K_2) \propto f(t_1, \dots, t_{m''} / \theta) = \theta^{m''} \cdot \exp \left(-\theta \cdot \sum_{j=1}^{m''} t_j \right)$$

cosicché la statistica sufficiente è costituita dalla coppia (m'', t) .

L'adeguamento della distribuzione iniziale $G(\theta)$ all'incremento di informazione K_2 si ottiene dalla:

$$g(\theta / K_2) \propto g_0(\theta) \cdot \ell(\theta / K_2) \propto \theta^{\alpha + m'' - 1} \cdot e^{-\theta(\beta + t)},$$

in base alla quale si ha che la densità finale è di tipo Gamma $(\alpha^{**}, \beta^{**})$ ove $\alpha^{**} = \alpha + m''$ e $\beta^{**} = \beta + t$. Assumendo quale stima bayesiana per θ il valor medio della densità finale è :

$$E(\Theta / K_2) = \frac{\alpha + m''}{\beta + t} = \frac{\beta \cdot \frac{\alpha}{\beta} + t \cdot \frac{m''}{t}}{\beta + t},$$

ove l'ultimo rapporto è ancora costituito da una combinazione lineare del valor medio iniziale α/β e della stima di massima verosimiglianza m''/t .

Al crescere del tempo totale di osservazione $t = \sum_{j=1}^{m''} t_j$ la stima bayesiana tende a coincidere con m'' / t .

Per costruire una previsione sui successivi k tempi tra arrivi successivi non ancora osservati $T_{m''+1}, \dots, T_{m''+k}$ il procedimento è oramai noto: la densità previsionale congiunta per le suddette variabili si ottiene applicando il teorema di rappresentazione al modo seguente:

$$f(\tau_1, \dots, \tau_k / K_2) = \int_{\{\theta\}} \left[\prod_{i=1}^k f(\tau_i / \theta) \right] \cdot g(\theta / K_2) d\theta = \frac{\Gamma(\alpha + m'' + k)}{\Gamma(\alpha + m'')} \cdot \frac{(\beta + t)^{\alpha + m''}}{(\beta + t + \sum \tau_i)^{\alpha + m'' + k}}.$$

Il procedimento inferenziale **diretto** conduce a questa stessa espressione mediante l'applicazione del **teorema delle probabilità composte alle distribuzioni della legge temporale iniziale $\Lambda(H)$**

$$f(t_1, \dots, t_{m''} / H) = \frac{\Gamma(\alpha + m'')}{\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{(\beta + \sum t_j)^{\alpha + m''}}, \quad m'' \geq 1.$$

Si ha dunque con facili calcoli:

$$f(\tau_1, \dots, \tau_k / H \cap K_2) = \frac{f(t_1, \dots, t_{m''}; \tau_1, \dots, \tau_k / H)}{f(t_1, \dots, t_{m''} / H)} = \frac{\Gamma(\alpha + m'' + k)}{\Gamma(\alpha + m'')} \cdot \frac{(\beta + t)^{\alpha + m''}}{(\beta + t + \sum \tau_i)^{\alpha + m'' + k}}.$$

Si osservi infine che, con riferimento al processo di osservazione residuo $\{N_h; h \geq m'+1\}$ e secondo l'**impostazione classica**, dopo l'utilizzazione dell'informazione statistica K_1 il processo è considerato essere un processo di Poisson con intensità $\hat{\theta} = \frac{\sum n_h}{m'}$. Nell'**impostazione bayesiana**, dopo l'utilizzazione di K_1 , il processo $\{N_h; h \geq m'+1\}$ è invece una miscela di processi poissoniani caratterizzata dalla funzione peso $g(\theta / K_1)$.

Ovviamente, la stessa considerazione riguarda anche il processo dei tempi di attesa residuo $\{T_j; j \geq m''+1\}$, la stima $\hat{\theta} = \frac{m''}{t}$ e la distribuzione finale $g(\theta / K_2)$.

5) Analisi Bayesiana su processi Gaussiani di misurazione

Consideriamo il problema di stima di una grandezza fisica incognita Θ , per esempio una temperatura o una lunghezza o altro, che può venire misurata un numero indefinito di volte con uno strumento di misura che introduce soltanto errori accidentali.

Ricordiamo che sono detti **accidentali** quegli errori di misura, non eliminabili, dovuti a moltissimi fattori indipendenti tra loro, ciascuno dei quali è ritenuto trascurabile, ma che nel loro insieme possono dar luogo a scostamenti sensibili del valore misurato dal valore effettivo della grandezza esaminata. In Teoria degli errori, applicando il teorema centrale limite, si stabilisce che gli errori accidentali di misura abbiano una distribuzione di probabilità di tipo **gaussiano** con valor medio nullo e varianza pari al reciproco della precisione dello strumento di misura.

Assumeremo che le ripetute misurazioni di Θ avvengano nelle stesse condizioni generali cosicché sia plausibile assumere **scambiabile** il processo di osservazione

$\{X_t; t \geq 1\}$ o, equivalentemente, che i risultati delle misurazioni X_t siano **condizionatamente indipendenti** e abbiano la stessa distribuzione condizionata rispetto ad ogni possibile valore θ di Θ . Per i motivi su esposti assumeremo anche che sia $F_{X_t}(x/\theta) \square N(\theta, \sigma^2)$ ove σ^2 è il reciproco della precisione nota dello strumento di misura.

Il modello statistico parametrico che si è assunto risulta dunque costituito da una famiglia di densità campionarie $\{F_{1, \dots, n}(x/\theta); n \geq 1\}$ di tipo normale, ciascuna delle quali è caratterizzata dal vettore medio $\theta \cdot \mathbf{1}$ e dalla matrice di dispersione $\sigma^2 \cdot I_n$.

Ne discende che la funzione di verosimiglianza di Θ relativa ai risultati

$\mathbf{x} = (x_1, \dots, x_n)'$ di n misurazioni successive è espressa dalla:

$$l(\theta / x_1, \dots, x_n) \propto f(x_1, \dots, x_n / \theta) = \prod_{i=1}^n f(x_i / \theta) \propto \exp\left\{-n \cdot (\theta - \bar{x})^2 / 2 \cdot \sigma^2\right\},$$

che fornisce la statistica sufficiente (n, \bar{x}) e la stima di massima verosimiglianza

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Da un punto di vista bayesiano occorre completare il modello statistico assunto esprimendo lo stato di informazione iniziale mediante una distribuzione probabilistica su Θ ; **supponiamo allora di scegliere una distribuzione normale $g_0(\theta)$ con parametri μ_0 e σ_0^2 fissati. E' facile verificare che la legge temporale iniziale $\Lambda(H)$ per il processo $\{X_t; t \geq 1\}$ è costituita da distribuzioni congiunte di tipo normale caratterizzate dal comune valor medio μ_0 , dalla comune varianza $\sigma_0^2 + \sigma^2$ e dalla covarianza σ_0^2 comune a tutte le coppie di variabili osservabili tra loro diverse.**

Se l'incremento di informazione è costituito dall'evento $K = \bigcap_{t=1}^m (X_t = x_t)$ l'inferenza bayesiana su Θ si realizza, **nel procedimento indiretto**, con l'adeguamento della distribuzione $g_0(\theta)$ mediante l'applicazione del teorema di Bayes:

$$g(\theta / K) \propto g_0(\theta) \cdot l(\theta / x_1, \dots, x_m) \propto \exp\{-(\theta - \mu_m)^2 / 2\sigma_m^2\},$$

ove

$$\mu_m = \frac{m\sigma^{-2}\bar{x} + \sigma_0^{-2}\mu_0}{m\sigma^{-2} + \sigma_0^{-2}}; \quad \sigma_m^2 = \frac{1}{m\sigma^{-2} + \sigma_0^{-2}}.$$

La densità finale $g(\theta / K)$ è dunque **Gaussiana con parametri μ_m e σ_m^2 e, come già osservato in precedenza, il valor medio finale μ_m di Θ è una combinazione lineare convessa del valor medio iniziale μ_0 e della stima di massima verosimiglianza \bar{x} ; tale combinazione può essere interpretata come una stima bayesiana per Θ nello stato di informazione finale (rispetto a K) e tende a coincidere con la stima di massima verosimiglianza al crescere del numero n delle misurazioni eseguite.**

Si noti anche che la varianza della densità finale σ_m^2 decresce in modo monotono al crescere di m e non dipende dai risultati delle misurazioni: è bene sottolineare che quest'ultima caratteristica non ha carattere generale, ma dipende dal particolare modello statistico scelto e precisamente dalla normalità delle distribuzioni campionarie $f(x_1, \dots, x_n / \theta)$ e della distribuzione iniziale $g_0(\theta)$.

Poiché la densità finale di Θ differisce da quella iniziale soltanto per i valori dei due parametri, si intuisce, ma lo si verifica facilmente, che la **legge temporale finale** $\Lambda(H \wedge K)$ per il processo di misurazioni residuo $\{X_t; t \geq m+1\}$ è costituita da **distribuzioni congiunte di tipo normale** caratterizzate dal comune valor medio μ_m , dalla comune varianza $\sigma_m^2 + \sigma^2$ e dalla comune covarianza σ_m^2 per coppie di variabili X_{m+t}, X_{m+s} con $t \neq s$.

Con la precedente osservazione abbiamo sostanzialmente riassunto il **procedimento inferenziale diretto** che ricava le distribuzioni di $\Lambda(H \wedge K)$ mediante l'applicazione del teorema delle probabilità composte alle densità della famiglia $\Lambda(H)$ in accordo con la:

$$f(v_{m+1}, \dots, v_{m+k} / H \cap K) = \frac{f(x_1, \dots, x_m, v_{m+1}, \dots, v_{m+k} / H)}{f(x_1, \dots, x_m / H)},$$

avendo indicato con v_{m+1}, \dots, v_{m+k} gli argomenti della densità congiunta delle prime k variabili osservabili di $\{X_t; t \geq m+1\}$ appartenente a $\Lambda(H \wedge K)$.

Multivariate Normal Distributions

The nondegenerate r -variate normal p.d.f. is

$$f_N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-r/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

where \mathbf{x} and $\boldsymbol{\mu}$ are $r \times 1$ vectors and $\boldsymbol{\Sigma}$ is an $r \times r$ symmetric positive definite matrix. The mean and variance are

$$E(\mathbf{x}) = \boldsymbol{\mu}$$

$$V(\mathbf{x}) = \boldsymbol{\Sigma}.$$

Partition \mathbf{x} , $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ conformably:

$$\mathbf{x}' = (\mathbf{x}'_I, \mathbf{x}'_J), \boldsymbol{\mu}' = (\boldsymbol{\mu}'_I, \boldsymbol{\mu}'_J) \quad \text{and}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{II} & \boldsymbol{\Sigma}_{IJ} \\ \boldsymbol{\Sigma}_{JI} & \boldsymbol{\Sigma}_{JJ} \end{bmatrix}$$

with \mathbf{x}_I having q elements and \mathbf{x}_J having $r - q$ elements. The marginal and conditional distributions are

$$f(\mathbf{x}_I) = f_N^q(\mathbf{x}_I | \boldsymbol{\mu}_I, \boldsymbol{\Sigma}_{II})$$

$$f(\mathbf{x}_I | \mathbf{x}_J) = f_N^q(\mathbf{x}_I | \boldsymbol{\mu}_I + \boldsymbol{\Sigma}_{IJ} \boldsymbol{\Sigma}_{JJ}^{-1} (\mathbf{x}_J - \boldsymbol{\mu}_J); \boldsymbol{\Sigma}_{II} - \boldsymbol{\Sigma}_{IJ} \boldsymbol{\Sigma}_{JJ}^{-1} \boldsymbol{\Sigma}_{JI}).$$

Notation. $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for "x is normally distributed with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$."

$$\underline{\mathbf{y}} = \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\boldsymbol{\theta}} \end{bmatrix} \begin{cases} f(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}) \sim N(F\underline{\boldsymbol{\theta}}; V) \\ g(\underline{\boldsymbol{\theta}}) \sim N(\underline{\boldsymbol{\mu}}; \boldsymbol{\Sigma}) \end{cases}$$

$$\underline{\mathbf{y}} \sim N\left(\begin{bmatrix} F\underline{\boldsymbol{\mu}} \\ \underline{\boldsymbol{\mu}} \end{bmatrix}; \begin{bmatrix} F\boldsymbol{\Sigma}F' + V & F\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}F' & \boldsymbol{\Sigma} \end{bmatrix}\right)$$

\Downarrow

$$\underline{\mathbf{x}} \sim N(F\underline{\boldsymbol{\mu}}; F\boldsymbol{\Sigma}F' + V)$$

$$\underline{\boldsymbol{\theta}} | \underline{\mathbf{x}} \sim N\left(\underline{\boldsymbol{\mu}} + \boldsymbol{\Sigma}F'(F\boldsymbol{\Sigma}F' + V)^{-1}(\underline{\mathbf{x}} - F\underline{\boldsymbol{\mu}}); \boldsymbol{\Sigma} - \boldsymbol{\Sigma}F'(F\boldsymbol{\Sigma}F' + V)^{-1}F\boldsymbol{\Sigma}\right)$$