



Dipartimento di scienze economiche,
aziendali, matematiche e statistiche
“Bruno de Finetti”

Research Paper Series, N. 3, 2016

Un'introduzione ai sistemi dinamici stocastici con variabili latenti

Attilio Wedlin
DEAMS – Università di Trieste



UNIVERSITÀ
DEGLI STUDI DI TRIESTE

Research Paper Series

Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche “Bruno de Finetti”

Piazzale Europa 1

34127, Trieste

Tel.: +390405587927

Fax: +390405587005

<http://www.deams.units.it>

EUT Edizioni Università di Trieste

Via E.Weiss, 21 - 34128 Trieste

Tel. +390405586183

Fax +390405586185

<http://eut.units.it>

eut@units.it

ISBN: 978-88-8303-753-5



9 788883 037535 >

Un'introduzione ai sistemi dinamici stocastici con variabili latenti

Attilio Wedlin
DEAMS – Università di Trieste

ABSTRACT¹

Questa nota si propone di introdurre ai processi stocastici parzialmente osservabili o sistemi dinamici stocastici con variabili latenti. Tali modelli sono ampiamente utilizzati nelle più diverse discipline, dall'economia alla biologia. L'esposizione è mirata ad un rapido apprendimento degli aspetti essenziali rinviando i dettagli dimostrativi ad un secondo tempo.

KEYWORDS: Modelli dinamici stocastici, processi di Markov, variabili latenti, variabili osservabili, problemi di filtraggio, filtro di Kalman, filtro di Wonham.

¹ Corresponding author: Attilio Wedlin, DEAMS, Università di Trieste, attilio.wedlin@deams.units.it

1. Introduzione

Ci siamo proposti di introdurre il lettore (tipicamente uno studente universitario dei corsi di laurea in Matematica, Fisica, Ingegneria, Statistica, Economia,.....) all'argomento dei sistemi dinamici stocastici con variabili latenti che trova oramai ampie applicazioni in tutti i campi sopra citati e in altri ancora. Si richiede al lettore di aver seguito almeno un corso di Calcolo delle probabilità e processi stocastici; le conoscenze richieste si possono comunque acquisire, per esempio, sui testi di G. Grimmett e D. Stirzaker o di A.N. Shiriyayev indicati in Riferimenti bibliografici.

Si è scelto, per alleggerire la trattazione, di dimostrare solo alcuni dei risultati presentati e di rinviare per le dimostrazioni e gli approfondimenti ad altre fonti. Si è preferito insistere su alcune esemplificazioni allo scopo di fornire chiari punti di riferimento per un primo avvicinamento al tema.

2. Sulla nozione di sistema dinamico stocastico

Dal punto di vista matematico un "sistema dinamico" è rappresentato da una coppia di oggetti (S, γ) ove $S = \{s\}$ denota l'insieme delle possibili determinazioni, o stati, di un sistema (fisico, chimico, biologico, economico,) e γ indica un'applicazione, o trasformazione, di S in se stesso. Il più delle volte gli stati s sono numeri reali e le trasformazioni γ sono funzioni continue e limitate delle variabili di stato.

Sistema dinamico deterministico a tempo discreto

Se $s_0 \in S$ è lo stato iniziale al tempo $t = 0$, al tempo $t = 1$ lo stato sarà $s_1 = \gamma(s_0)$; al tempo $t = 2$ lo stato sarà $s_2 = \gamma(s_1) = \gamma[\gamma(s_0)]$, che si indica con $\gamma^2(s_0)$, e così via. L'interesse principale riguarda il comportamento del sistema al crescere del tempo, cioè lo studio dell'evoluzione, o traiettoria, di $s_n = \gamma(s_{n-1}) = \gamma^n(s_0)$ al crescere di n . Si osservi che, **noto lo stato iniziale s_0 del sistema**, la sua evoluzione futura è una traiettoria di tipo deterministico, senza nessun elemento di incertezza.

Un semplice esempio nel caso $S = R$ è fornito dalla trasformazione $\gamma(s) = a.s$ con $a \in R$; la traiettoria del sistema è descritta dalla $s_n = \gamma(s_{n-1}) = a.s_{n-1} = a^n .s_0$ e il tipo di andamento, al crescere di n , risulta determinato dal valore del coefficiente a e in particolare dal fatto che sia $|a| < 1$ oppure $|a| \geq 1$.

Sistema dinamico con stato iniziale incerto

Se invece lo stato iniziale non è noto con precisione non lo saranno neanche gli stati successivi e l'incertezza interesserà l'intera traiettoria. Occorre allora rappresentare lo spazio degli stati mediante uno spazio di probabilità (S, A, P) ove A indica una σ – algebra di sottoinsiemi di S e P indica una misura di probabilità su A che esprime la nostra conoscenza parziale della situazione concernente lo stato iniziale S_0 . La trasformazione γ deve allora essere assunta A - misurabile per cui è $\gamma^{-1}(A) = \{s \in S : \gamma(s) \in A\} \in A, \forall A \in A$.

Indicato lo stato iniziale non noto, e quindi aleatorio, del sistema con S_0 gli stati successivi $S_1 = \gamma(S_0), S_2 = \gamma(S_1) = \gamma^2(S_0), \dots$ saranno tutti aleatori e costituiscono il processo stocastico (a valori reali se $S = R$) $\{S_n; n \geq 0\}$ che esprime l'evoluzione del sistema nel tempo (o la sua traiettoria aleatoria).

La maggior parte dei sistemi dinamici studiati riguarda quelli per i quali il processo stocastico $\{S_n; n \geq 0\}$ è **stazionario in senso stretto**, cioè quelli le cui distribuzioni congiunte finite – dimensionali sono invarianti rispetto a traslazioni rigide del parametro operativo; scelta cioè arbitrariamente una sequenza A_1, A_2, \dots, A_m di insiemi di A dev'essere:

$$\Pr \left\{ \bigcap_{j=1}^m [S_{n_j} \in A_j] \right\} = \Pr \left\{ \bigcap_{j=1}^m [S_{n_j+k} \in A_j] \right\}$$

per ogni intero $m \geq 1$, ogni sequenza (n_1, \dots, n_m) e ogni intero k per cui $n_j + k \in Z^+$.

Si dimostra che sussiste una forte connessione tra la condizione di stazionarietà in senso stretto per $\{S_n; n \geq 0\}$ e la proprietà della misura P descritta dalla condizione

$P(A) = P(\gamma^{-1}A)$, per ogni $A \in A$. In letteratura l'ultima condizione per P e' detta "invarianza di P rispetto a γ ".

Un insieme $A \in A$ e' detto "invariante rispetto a γ " se soddisfa la $A = \gamma^{-1}A$; si prova che tali insiemi formano una sotto σ – algebra \mathfrak{I} di A . Infine la misura P , invariante rispetto a γ , e' detta "ergodica" se per ogni evento A invariante, cioè tale che $A \in \mathfrak{I}$, e' $P(A) = 1$ oppure $P(A) = 0$.

Possiamo enunciare a questo punto il celebre

Teorema ergodico di G. Birkhoff: se S_0 e' lo stato iniziale aleatorio del sistema e la probabilità P e' invariante rispetto γ , considerato il processo stocastico a valori reali $\{S_n = \gamma^n(S_0); n \geq 1\}$, sussiste, al divergere di n , la

$$\frac{1}{n} \sum_{h=0}^{n-1} \gamma^h(S_0) \xrightarrow{qc} E(S_0 | \mathfrak{I})$$

e se la misura di probabilità è ergodica allora $E(S_0|\mathfrak{F}) = E(S_0)$. Inoltre, se f è una funzione integrabile definita in S , allora

$$\frac{1}{n} \sum_{h=0}^{n-1} f[\gamma^h(S_0)] \xrightarrow{qc} E[f(S_0)|\mathfrak{F}]$$

e se P è ergodica si ha anche $E[f(S_0)|\mathfrak{F}] = E[f(S_0)]$.

Sostanzialmente il teorema ergodico è una **legge forte dei grandi numeri** per processi stazionari; osserviamo che il teorema di B. de Finetti per **processi scambiabili** (un caso particolare di stazionarietà), dimostrato nel 1928, costituisce un caso particolare del teorema di Birkhoff ed i corrispondenti eventi invarianti di \mathfrak{F} sono quelli il cui valore logico non varia rispetto a permutazioni finite e arbitrarie degli argomenti.

Sistema dinamico stocastico

In quanto detto finora l'incertezza nel sistema dinamico (S, γ) riguardava soltanto lo stato iniziale e la probabilità P esprimeva una conoscenza parziale su di esso. Può intervenire però una seconda fonte di incertezza: **la trasformazione γ anziché essere nota con certezza può essere un elemento non noto di un insieme Γ di applicazioni di S in S** . Un'eventuale conoscenza parziale può essere espressa da una misura di probabilità Q definita su una qualche σ -algebra di sottoinsiemi di Γ . I primi a suggerire questa generalizzazione della nozione di sistema dinamico furono S. Ulam e J. von Neumann in un articolo del 1945 dal titolo "Random Ergodic Theorems".

Definiremo allora un "sistema dinamico stocastico" come una terna di elementi (S, Γ, Q) dei quali il primo è lo spazio degli stati $S = \{s\}$, il secondo è un insieme di applicazioni di S in S , il terzo è una misura di probabilità su una σ -algebra di sottoinsiemi di Γ .

Se il sistema è a tempo discreto e lo stato di partenza al tempo $t = 0$ è noto essere s_0 , la prima transizione $s_0 \rightarrow S_1 = \gamma_1(s_0)$ è effettuata dall'applicazione $\gamma_1 \in \Gamma$ scelta in base alla probabilità Q : il risultato S_1 è aleatorio. La seconda transizione $S_1 \rightarrow S_2 = \gamma_2(S_1)$ è effettuata da un'applicazione $\gamma_2 \in \Gamma$, eventualmente diversa da γ_1 , scelta sempre in base a Q e **indipendentemente dalla prima scelta**. Chiaramente S_2 è aleatorio. E così via: la sequenza $\{S_1, S_2, \dots\}$ costituisce un processo stocastico che ha valori reali se $S = R$. È importante sottolineare che le successive applicazioni di Γ sono **mutuamente indipendenti ed hanno la stessa distribuzione Q** .

Se anche lo stato iniziale è ignoto le fonti di incertezza sono due e si dovrà far intervenire anche la probabilità P oltre alla Q . Il sistema dinamico dovrebbe venire quindi rappresentato come una coppia di spazi di probabilità $\{(S, A, P), (\Gamma, B, Q)\}$ in cui le

trasformazioni di Γ agiscono sullo spazio degli stati S . Si dimostra che il processo stocastico $\{S_0, S_1, \dots\}$ è una catena di Markov.

Un primo esempio di sistema dinamico stocastico a tempo discreto è fornito dall'equazione alle differenze stocastica

$$X_n = a.X_{n-1} + \varepsilon_n, \quad X_0 = x_0,$$

ove a e x_0 sono assunte essere costanti note e ove $\{\varepsilon_n; n \geq 1\}$ è un processo stocastico Gaussiano, con variabili i.i.d. e caratterizzato dai primi due momenti $E(\varepsilon_n) \equiv 0$ e $Var(\varepsilon_n) \equiv \sigma_\varepsilon^2$. Tale modello stocastico è noto in letteratura come modello “autoregressivo del primo ordine” (AR(1) in breve). Ovviamente in questo esempio gli stati di S , da X_1 in poi, sono aleatori e costituiscono il processo stocastico generato dal modello; inoltre la distribuzione di probabilità Q coincide con le distribuzioni del processo $\{\varepsilon_n; n \geq 1\}$.

Un secondo esempio di sistema dinamico stocastico, questa volta a tempo continuo, è fornito dall'equazione differenziale stocastica

$$dX(t) = -a.X(t)dt + dW(t), \quad X(0) \approx N(0, \sigma_0^2),$$

ove $a > 0$ e ove $\{W(t); t \geq 0\}$ è un processo di Wiener standardizzato, cioè con $Var[W(t)] = t$; si assume inoltre che sia $X(0) \perp\!\!\!\perp W(t)$ per ogni $t \geq 0$. Tale modello stocastico è una versione della “equazione di Langevin” ed il processo soluzione $\{X(t)\}$ è detto “processo di Ornstein – Uhlenbeck” (se $\sigma_0^2 = 1/2a$, nel qual caso $\{X(t)\}$ è stazionario).

In entrambi gli esempi le trasformazioni di Γ sono stocastiche e descritte dalle suddette equazioni, mentre le misure di probabilità, che indichiamo con Q_1 e Q_2 , sono determinate dai processi di rumore $\{\varepsilon_n\}$ e $\{W(t)\}$. L'obiettivo principale è lo studio dell'evoluzione delle variabili di stato

$$X_n = a^n .x_0 + \sum_{h=0}^{n-1} a^h .\varepsilon_{n-h} \quad \text{e} \quad X(t) = X(0).e^{-a.t} + \int_0^t e^{-a.(t-s)} dW(s)$$

che risultano dalla risoluzione delle suddette equazioni stocastiche. Entrambi i processi stocastici $\{X_n\}$ e $\{X(t)\}$ sono processi di Markov a tempo discreto e, rispettivamente, continuo. Le corrispondenti densità condizionate di transizione sono espresse rispettivamente dalle:

$$p(x_n, x_{n+1}) \approx N(a.x_n; \sigma_\varepsilon^2) \quad \text{e} \quad p(t-s, x, y) \approx N\left[x.e^{-a.(t-s)} ; (2a)^{-1} . [1 - e^{-2a.(t-s)}]\right].$$

Il collegamento formale tra queste densità di transizione e le probabilità Q su Γ è espresso dalle seguenti uguaglianze:

$$\Pr[X_{n+1} \in A | X_n = x_n] = \int_A p(x_n, x_{n+1}) dx_{n+1} = Q_1[\gamma \in \Gamma : \gamma(x_n) \in A] \text{ e}$$

$$\Pr\{X(t) \in A | X(s) = x\} = \int_A p(t-s, x, y) dy = Q_2[\gamma \in \Gamma : \gamma(x) \in A].$$

A questo proposito un importante problema generale riguarda la possibilità di rappresentare una catena di Markov con insieme degli stati S e avente fissate probabilità o densità di transizione mediante una composizione di trasformazioni aleatorie appartenenti ad una fissata famiglia Γ del tipo presentato in precedenza: in altri termini si tratta di individuare una misura di probabilità Q su una σ -algebra B di Γ tale da realizzare le suddette uguaglianze. In generale, quando una tale composizione esiste essa non è unica.

E' noto il seguente risultato:

Proposizione 1: se l'insieme degli stati S di una fissata catena di Markov a tempo discreto è un sottoinsieme di uno spazio metrico, separabile e completo (Polish space) allora per ogni famiglia di funzioni di transizione $P(x, A) = \int_A p(x, y) dy$, con $x \in S$ e $A \in \mathcal{A}$, della catena esiste una misura di probabilità Q su B per la quale è $P(x, A) = Q\{\gamma \in \Gamma : \gamma(x) \in A\}$.

Per un primo studio dei sistemi dinamici stocastici si consiglia la monografia introduttiva di R. Bhattacharya e M. Majumdar "Random dynamical systems: a review" del 2004 e reperibile in rete. Per approfondimenti ulteriori si possono vedere, in sequenza crescente di complessità, i testi di A.N. Shiriyayev, di Y. Kifer e di L. Arnold citati in Riferimenti bibliografici. In particolare, il testo di L. Arnold generalizza la trattazione di Y. Kifer sostituendo alla scelta casuale delle applicazioni di Γ in condizioni di indipendenza e uguale distribuzione quella, piu' generale, di scelta in condizione di stazionarietà'.

3. Sistemi dinamici con variabili latenti

Si e' visto che le equazioni stocastiche alle differenze finite e differenziali rappresentano modelli dinamici stocastici. In molte applicazioni le variabili di stato X non sono osservabili e si possono osservare soltanto loro trasformazioni aleatorie Y : si parla allora di **variabili di stato latenti**. Si afferma anche che il processo stocastico vettoriale $\{(X_t, Y_t); t \in T\}$ e' **parzialmente osservabile**.

Quindi il modello probabilistico, oltre all'equazione dinamica delle variabili di stato, contiene una seconda equazione che definisce le variabili osservabili Y_t in termini delle variabili latenti X_t e di un eventuale rumore aleatorio o perturbazione aleatoria non osservabile.

Il primo semplice esempio di sistema dinamico con variabili latenti che presentiamo non e' pero' di questo tipo: esso e' costituito da una sequenza di variabili di stato X_n , non osservabili, costituente una catena di Markov con un numero finito M di stati $\{1,2,\dots,\dots,M\}$, caratterizzata da una distribuzione iniziale ν e da una matrice di transizione $P = [p_{ij}]$, e da un secondo processo osservabile Y_n costituito da variabili condizionatamente indipendenti rispetto al processo $\{X_n\}$. Se si assume che le variabili Y_n abbiano le stesse determinazioni delle variabili di stato X_n , per completare la specificazione del modello occorre fissare ancora la matrice quadrata delle probabilita' condizionate $Q = [q_{jk}]$ ove si definisce $q_{jk} = \Pr[Y_n = k | X_n = j]$.

Se indichiamo con X_1^n la sequenza (X_1, \dots, X_n) e con Y_1^n la sequenza (Y_1, \dots, Y_n) le precedenti assunzioni possono essere espresse al modo seguente:

- 1) $\Pr\{X_{n+1} = x_{n+1} | (X_1^n = x_1^n) \cap (Y_1^n = y_1^n)\} = \Pr\{X_{n+1} = x_{n+1} | X_n = x_n\} \in P$,
- 2) $\Pr\{Y_n = y_n | (X_1^n = x_1^n) \cap (Y_1^{n-1} = y_1^{n-1})\} = \Pr\{Y_n = y_n | X_n = x_n\} \in Q$.

La terna (ν, P, Q) caratterizza completamente il processo bivariato $\{(X_n, Y_n); n \geq 1\}$; per esempio determiniamo la seguente probabilita' congiunta utilizzando le 1) e 2) oltre che la distribuzione ν :

$$\Pr\{(X_1^n = x_1^n) \cap (Y_1^n = y_1^n)\} = \Pr(x_1^n) \Pr(y_1^n | x_1^n) = \left[\nu(x_1) \cdot \prod_{j=2}^n p(x_j | x_{j-1}) \right] \left[\prod_{i=1}^n q(y_i | x_i) \right].$$

Le variabili osservabili Y possono avere determinazioni differenti dagli stati $\{1,2,\dots,\dots,M\}$: se indichiamo con $\{a,b,\dots,\dots,z\}$ le determinazioni delle variabili osservabili la matrice Q delle probabilita' condizionate $p(y_n | x_n)$ avra' M righe e M colonne. Inoltre le determinazioni osservabili possono costituire un insieme continuo, per esempio l'intervallo $[\alpha, \beta]$ dell'asse reale; in tal caso la matrice Q sara' sostituita da M distribuzioni condizionate $F_{Y/X}(y|j)$, $j \in \{1,2,\dots,\dots,M\}$, $y \in [\alpha, \beta]$ oppure dalle corrispondenti densita' nel caso che queste esistano.

Un secondo esempio e' costituito dal modello AR(1) per le variabili latenti X completato da un'equazione lineare stocastica per le variabili osservabili Y :

$$3) \quad X_n = a.X_{n-1} + \varepsilon_n, \quad X_0 = x_0, \quad \varepsilon_n \approx NWN(0, \sigma_\varepsilon^2);$$

$$4) \quad Y_n = b.X_n + \xi_n, \quad Y_0 = 0, \quad \xi_n \approx NWN(0, \sigma_\xi^2), \quad \xi \coprod \varepsilon;$$

si assume infine che i coefficienti a, b e le varianze $\sigma_\varepsilon^2, \sigma_\xi^2$ siano parametri noti.

Dalle ipotesi fatte si ricava facilmente che il processo vettoriale (X_n, Y_n) e' Markoviano omogeneo e Gaussiano; le probabilita' condizionate di transizione per il processo $\{X_n\}$ sono espresse dalle $P\{X_n \in A | X_{n-1} = x\} = \Lambda(x, A) = \int_A f_\varepsilon(z - a.x) dz$ mentre le probabilita'

condizionate per il processo $\{Y_n\}$ sono date dalle

$$P\{Y_n \in B | X_n = x\} = \Psi(B|x) = \int_B f_\xi(y - b.x) dy. \text{ Naturalmente le due funzioni integrande}$$

f_ε ed f_ξ sono le corrispondenti densita' condizionate e sono entrambe di tipo Gaussiano.

Denotando con F_n^X e F_n^Y le σ -algebre generate dai n.a. (X_1, \dots, X_n) e (Y_1, \dots, Y_n) e indicando con $\nu(dx_0)$ la distribuzione di probabilita' della condizione iniziale X_0 si hanno le:

$$5) \quad P\{X_n \in A, Y_n \in B | F_{n-1}^X, F_{n-1}^Y\} = P\{X_n \in A, Y_n \in B | X_{n-1}, Y_{n-1}\} = \int_A \Lambda(X_{n-1}, dz) \cdot \Psi(B|z),$$

$$6) \quad P\left\{\bigcap_{j=1}^n (Y_j \in B_j) | F_n^X\right\} = \prod_{j=1}^n \Psi(B_j | X_j),$$

$$7) \quad P\left\{\bigcap_{j=0}^n [(X_j \in A_j) \cap (Y_j \in B_j)]\right\} = \int_{A_0} \dots \int_{A_n} \nu(dx_0) \cdot |0 \in B_0| \cdot \prod_{j=1}^n [\Lambda(x_{j-1}, dx_j) \cdot \Psi(B_j | x_j)]$$

.

Evidentemente le precedenti espressioni poggiano sulle ipotesi assunte in precedenza; ricordiamole brevemente:

- a) Markovianita' dei processi (X_n, Y_n) e (X_n) ,
- b) Mutua indipendenza stocastica condizionata delle variabili osservabili Y_n nota la traiettoria delle variabili di stato o la σ -algebra F_∞^X ,
- c) La terna $\{\nu(dx_0), \Lambda(x, dz), \Psi(dy|x)\}$ specifica l'intera struttura probabilistica del modello dinamico stocastico sopra descritto.

Precisiamo che la linearita' del modello non influisce sulle precedenti condizioni; infatti nulla cambia sostanzialmente se si assumono al posto delle precedenti le seguenti equazioni:

$$8) \quad X_n = g(X_{n-1}) + \varepsilon_n, \quad X_0 \approx \nu(x_0) \approx N(m_0, \sigma_0^2), \quad \varepsilon_n \approx NWN(0, \sigma_\varepsilon^2), \quad \varepsilon_n \coprod X_0$$

$$9) \quad Y_n = h(X_n) + \xi_n, \quad Y_0 = 0, \quad \xi_n \approx NWN(0, \sigma_\xi^2), \quad \xi \prod \mathcal{E},$$

ove le funzioni $g(\cdot)$ e $h(\cdot)$, non necessariamente lineari, sono supposte assegnate; si ha ancora

$$P\{X_n \in A | X_{n-1} = x\} = \Lambda(x, A) = \int_A f_\varepsilon [z - g(x)] dz \text{ e}$$

$$P\{Y_n \in B | X_n = x\} = \Psi(B|x) = \int_B f_\xi [y - h(x)] dy.$$

4. Problemi di stima per variabili latenti: modelli a tempo discreto.

Mostriamo ora come avviene l'apprendimento sulle variabili di stato X_n attraverso gli incrementi di informazione basati sull'osservabilità delle variabili Y_n . In particolare ci occuperemo del cosiddetto "problema di filtraggio" (filtering problem) riguardante la determinazione della distribuzione di probabilità condizionata $F_n(x|Y_1^n)$ delle variabili di stato X_n rispetto alla sequenza Y_1^n delle variabili osservabili, o alla σ -algebra F_n^Y da esse generata. Equivalentemente ci si può proporre la determinazione delle funzioni di regressione $E[f(X_n)|Y_1^n]$ per funzioni arbitrarie, purchè limitate e misurabili, delle variabili di stato; com'è noto $E[f(X_n)|Y_1^n]$ costituisce lo stimatore ottimale dei minimi quadrati per il numero aleatorio non osservabile $f(X_n)$. Si ha evidentemente per $f(X_n) = |X_n \leq x|$ l'uguaglianza $F_n(x|Y_1^n) = P\{X_n \leq x | Y_1^n\} = E[|X_n \leq x| Y_1^n]$.

Inoltre, se $f(X_n) = X_n$ allora $E[f(X_n)|Y_1^n] = E(X_n|Y_1^n)$ fornisce lo stimatore ottimale per X_n .

Allo scopo di prendere in considerazione un modello abbastanza generale, ma non troppo complicato, supponiamo che i due processi X_n e Y_n siano definiti dalle equazioni precedenti 8) e 9) ma che i processi di rumore, ancora costituiti da variabili i.i.d. e non correlati tra loro, abbiano densità di probabilità f_ε ed f_ξ non necessariamente di tipo Gaussiano.

Si dimostra allora che la distribuzione $F_n(x|Y_1^n)$, che supponiamo abbia densità $\pi_n(x)$, soddisfa la seguente relazione ricorrente

$$(*) \quad \pi_n(x) = \frac{\int_{\mathcal{R}} f_{\xi}[Y_n - h(x)] \int_{\mathcal{R}} f_{\varepsilon}[x - g(s)] \pi_{n-1}(s) ds}{\int_{\mathcal{R}} f_{\xi}[Y_n - h(x)] \int_{\mathcal{R}} f_{\varepsilon}[x - g(s)] \pi_{n-1}(s) ds dx}$$

ove $\pi_0(x) = \nu_0(x)$.

Per la dimostrazione si veda, per esempio, P. Chigansky – Nonlinear filtering al capitolo 3, paragrafi 2 e 3. Noi ci limitiamo a dare un cenno di dimostrazione supponendo che sia noto l'evento $Y_0^n = y_0^n$ e indicando con la lettera f tutte le densità coinvolte nel discorso: il loro significato sarà desumibile dagli argomenti indicati per ciascuna di esse. Si ha:

$$\begin{aligned} f(x_n / y_0^n) &= \frac{f(y_0^n / x_n) \cdot f(x_n)}{f(y_0^n)} = \frac{f(y_0^{n-1}, y_n / x_n) \cdot f(x_n)}{f(y_0^{n-1}, y_n)} = \frac{f(y_n / y_0^{n-1}, x_n) \cdot f(y_0^{n-1} / x_n) \cdot f(x_n)}{f(y_n / y_0^{n-1}) \cdot f(y_0^{n-1})} = \\ &= \frac{f(y_n / x_n) \cdot f(x_n / y_0^{n-1})}{f(y_n / y_0^{n-1})} = \frac{f(y_n / x_n) \cdot \int_{\mathcal{R}} f(x_n / x_{n-1}) \cdot f(x_{n-1} / y_0^{n-1}) dx_{n-1}}{\int_{\mathcal{R}} f(y_n / x_n) \cdot \int_{\mathcal{R}} f(x_n / x_{n-1}) \cdot f(x_{n-1} / y_0^{n-1}) dx_{n-1} dx_n} \end{aligned}$$

e il lettore riconoscerà facilmente la corrispondenza dell'espressione ottenuta con quella del secondo membro della (*).

Si ottiene un'equazione più semplice della (*) al modo seguente: ponendo

$$\sigma(x_n) = f(y_n / x_n) \cdot \int_{\mathcal{R}} f(x_n / x_{n-1}) \cdot f(x_{n-1} / Y_0^{n-1}) dx_{n-1},$$

per cui dalla precedente espressione è

$$f(x_n / y_0^n) = \frac{\sigma(x_n)}{\int_{\mathcal{R}} \sigma(x_n) dx_n},$$

si prova che la densità di misura non normalizzata $\sigma(x_n)$ soddisfa l'equazione differenziale lineare e ricorsiva, detta "equazione di M. Zakai",

$$\begin{aligned} \sigma(x_n) &= f(Y_n | x_n) \cdot \int_{\mathcal{R}} f(x_n / x_{n-1}) \cdot \sigma(x_{n-1}) dx_{n-1}, \\ \sigma(x_0) &= \nu(x_0). \end{aligned}$$

Trovandone la soluzione, eventualmente per via numerica, la densità di probabilità $f(x_n | Y_0^n)$ si trova con la semplice normalizzazione della $\sigma(x_n)$.

Per il caso in cui il sistema dinamico è descritto dalle precedenti equazioni 8) e 9) la corrispondente equazione lineare di Zakai per la densità di misura non normalizzata $\sigma(x_n)$ ha l'espressione

$$\sigma(x_n) = f_\xi[Y_n - h(x_n)] \cdot \int_R f_\varepsilon[x_n - g(x_{n-1})] \cdot \sigma(x_{n-1}) dx_{n-1},$$

$$\sigma(x_0) = \nu(x_0),$$

ove f_ξ ed f_ε hanno forma funzionale Gaussiana.

In generale le suddette equazioni stocastiche sono risolubili, e quindi utilizzabili nei problemi concreti, solo in pochi casi. In quanto segue ne indicheremo due soli: il primo caso è quello in cui le funzioni $g(x)$ e $h(x)$ sono lineari e tutte le distribuzioni sono Gaussiane (modello lineare di Gauss – Markov). Il secondo caso, non direttamente riconducibile al precedente, è quello in cui le variabili latenti X_n hanno un numero finito di stati; la catena Markoviana $\{X_n\}$ non sarà quindi definita da una equazione alle differenze del tipo 8) ma direttamente dalla sua distribuzione iniziale e dalla matrice delle probabilità condizionate di transizione.

Incominceremo appunto da quest'ultima situazione.

Filtro di W.M. Wonham per catene di Markov a tempo discreto

Indicheremo con $\{X_n; n \geq 0\}$ una catena di Markov caratterizzata dalla matrice $\Lambda = [\lambda_{ij}]$ delle probabilità condizionate di transizione $\lambda_{ij} = P(X_n = x_j | X_{n-1} = x_i)$ e dalla distribuzione iniziale $p_0(i) = P(X_0 = x_i)$, $i = 1, \dots, d$. Supporremo che le variabili della catena non siano osservabili, mentre siano osservabili le variabili $Y_n = X_n + \varepsilon_n$ ove le perturbazioni aleatorie ε_n , non osservabili, sono assunte i.i.d., dotate della medesima densità di probabilità $f_\varepsilon(\cdot)$, e indipendenti da X_0 .

E' noto che in assenza di osservazioni sulle variabili Y_n , cioè nello stato di informazione iniziale, le distribuzioni delle X_n , cioè i vettori p_n con elementi $p_n(i)$, sono determinate dalle relazioni

$$10) \quad p_n = \Lambda^T \cdot p_{n-1} = (\Lambda^T)^n \cdot p_0.$$

E' anche noto che il problema di stima ottimale dei minimi quadrati delle variabili della catena X_n , $n \geq 1$, basata sull'osservabilità delle Y_n consiste nella determinazione delle

funzioni di regressione $E[X_n | Y_0, Y_1, \dots, Y_n] = E[X_n | F_n^Y]$, ove F_n^Y denota la σ -algebra generata dalla sequenza osservabile Y_0, Y_1, \dots, Y_n .

Se π_n denota il vettore delle probabilità $\pi_n(i) = P[X_n = x_i | F_n^Y] = E[X_n = x_i | F_n^Y]$ si ha

$$E[X_n | Y_0, Y_1, \dots, Y_n] = E[X_n | F_n^Y] = \sum_{i=1}^d x_i \cdot \pi_n(i) .$$

Sussiste per le probabilità $\pi_n(i)$ il seguente risultato

Teorema di W.M. Wonham: nelle suddette ipotesi si hanno le relazioni

$$11) \quad \begin{aligned} \pi_0(i) &= P[X_0 = x_i | Y_0] = \frac{f_\varepsilon(Y_0 - x_i) \cdot p_0(i)}{\sum_i f_\varepsilon(Y_0 - x_i) \cdot p_0(i)} , \\ \pi_n(i) &= P[X_n = x_i | F_n^Y] = \frac{f_\varepsilon(Y_n - x_i) \cdot \sum_h \lambda_{hi} \cdot \pi_{n-1}(h)}{\sum_i f_\varepsilon(Y_n - x_i) \cdot \sum_h \lambda_{hi} \cdot \pi_{n-1}(h)} . \end{aligned}$$

Si prova facilmente che il vettore delle $\pi_n(i)$, $i = 1, \dots, d$, ha l'espressione seguente:

$$12) \quad \pi_n = [\pi_n(1), \dots, \pi_n(d)]^T = \frac{\text{diag}[f_\varepsilon(Y_n - x_i)] \cdot \Lambda^T \cdot \pi_{n-1}}{[f_\varepsilon(Y_n - x_1), \dots, f_\varepsilon(Y_n - x_d)] \cdot \Lambda^T \cdot \pi_{n-1}} .$$

Per la dimostrazione si veda per esempio P. Chigansky – Introduction to stochastic processes (p. 56).

Filtro di Kalman

Se le funzioni $g(\cdot)$ e $h(\cdot)$ nelle 8) e 9) sono lineari il problema di filtraggio risulta molto più semplice. Negli anni 60 del secolo scorso sono stati introdotti due procedimenti ricorsivi di stima oggi molto noti: il “filtro di Kalman” per modelli lineari dinamici a tempo discreto, quali quello espresso dalle 3) e 4), ed il “filtro di Kalman – Bucy” per modelli lineari dinamici a tempo continuo. Noi illustreremo brevemente il filtro di Kalman per il modello

$$13) \quad X_n = g_n \cdot X_{n-1} + \varepsilon_n , \quad X_0 \approx \nu_0(x) \approx N(\mu_0, \sigma_0^2), \quad \varepsilon_n \approx NWN(0, \sigma_\varepsilon^2), \quad \varepsilon_n \perp\!\!\!\perp X_0$$

$$14) \quad Y_n = h_n \cdot X_n + \xi_n , \quad Y_0 = 0, \quad \xi_n \approx NWN(0, \sigma_\xi^2), \quad \xi_n \perp\!\!\!\perp \varepsilon ,$$

che è leggermente piu' generale di quello espresso dalle 3), 4) per il fatto che i coefficienti noti e costanti a e b sono sostituiti dalle successioni numeriche note $\{g_n\}$ e $\{h_n\}$ e per l'aleatorietà di X_0 .

La forma Gaussiana delle distribuzioni di X_0 e dei due processi di rumore ε_n e ξ_n fa si' che tutte le distribuzioni che considereremo saranno ancora Gaussiane per cui bastera' determinare i soli loro momenti del primo e secondo ordine.

Il procedimento iterativo consiste di una sequenza di passi ciascuno dei quali consta di due fasi dette di "previsione" e di "aggiornamento" e descritte sommariamente dallo schema seguente:

$$f(x_{n-1}|F_{n-1}^Y) \longrightarrow f(x_n|F_{n-1}^Y) \longrightarrow f(x_n|F_n^Y).$$

Partendo dalla distribuzione $\nu_0(x)$ di X_0 e supponendo di aver determinato, passo dopo passo, la densita' normale $f(x_{n-1}|F_{n-1}^Y)$ dello stato X_{n-1} avvalendosi della osservabilita' delle variabili Y_1, Y_2, \dots, Y_{n-1} , delle quali F_{n-1}^Y indica la σ -algebra generata, si passa a determinare $f(x_n|F_{n-1}^Y)$: questa prima trasformazione costituisce la fase di previsione. La seconda fase, di carattere induttivo, parte dalla distribuzione $f(x_n|F_{n-1}^Y)$ e, utilizzando l'informazione derivante dalla variabile osservabile Y_n , arriva alla distribuzione "aggiornata" $f(x_n|F_n^Y)$.

Indicando con \hat{X}_{n-1} e σ_{n-1}^2 il valor medio e la varianza di $f(x_{n-1}|F_{n-1}^Y)$ e con Y_1^{n-1} il vettore delle prime n-1 osservazioni (Y_1, \dots, Y_{n-1}), i parametri della $f(x_n|F_{n-1}^Y)$ si determinano utilizzando l'equazione 13) e le altre ipotesi del modello al modo seguente:

$$E(X_n|Y_1^{n-1}) = g_n \cdot E(X_{n-1}|Y_1^{n-1}) + E(\varepsilon_n|Y_1^{n-1}) = g_n \cdot \hat{X}_{n-1},$$

$$Var(X_n|Y_1^{n-1}) = g_n^2 \cdot Var(X_{n-1}|Y_1^{n-1}) + Var(\varepsilon_n|Y_1^{n-1}) = g_n^2 \cdot \sigma_{n-1}^2 + \sigma_\varepsilon^2.$$

La fase di aggiornamento puo' essere implementata in vari modi, ma noi sceglieremo di impiegare il teorema di Bayes secondo il quale dev'essere

$$f(x_n|Y_1^n) \propto f(x_n|Y_1^{n-1}) \cdot f(y_n|x_n, Y_1^{n-1})$$

ove il secondo fattore $f(y_n|x_n, Y_1^{n-1})$ rappresenta la verosimiglianza di X_n relativa all'osservazione $Y_n = y_n$. Formalmente, questa verosimiglianza ha la forma funzionale di una densita' normale con parametri determinati dall'equazione 14) al modo seguente:

$$E\left(Y_n | x_n, Y_1^{n-1}\right) = h_n \cdot E\left(X_n | Y_1^{n-1}\right) + E\left(\xi_n | Y_1^{n-1}\right) = h_n \cdot g_n \cdot \hat{X}_{n-1} ,$$

$$Var\left(Y_n | x_n, Y_1^{n-1}\right) = h_n^2 \cdot Var\left(X_n | Y_1^{n-1}\right) + Var\left(\xi_n | Y_1^{n-1}\right) = h_n^2 \cdot \left(g_n^2 \cdot \sigma_{n-1}^2 + \sigma_\xi^2\right) + \sigma_\xi^2 .$$

Eseguendo il prodotto $f(x_n | Y_1^{n-1}) \cdot f(y_n | x_n, Y_1^{n-1})$ e procedendo alla necessaria normalizzazione si trovano i parametri della densita' $f(x_n | Y_1^n)$:

$$\hat{X}_n = E\left(X_n | Y_1^n\right) = E\left(X_n | Y_1^{n-1}\right) + \frac{h_n \cdot Var\left(X_n | Y_1^{n-1}\right)}{h_n^2 \cdot Var\left(X_n | Y_1^{n-1}\right) + \sigma_\xi^2} \cdot \left[y_n - h_n \cdot E\left(X_n | Y_1^{n-1}\right)\right] ,$$

$$\sigma_n^2 = Var\left(X_n | Y_1^n\right) = \frac{\sigma_\xi^2 \cdot Var\left(X_n | Y_1^{n-1}\right)}{h_n^2 \cdot Var\left(X_n | Y_1^{n-1}\right) + \sigma_\xi^2} .$$

5. Problemi di stima per variabili latenti: modelli a tempo continuo.

Considereremo ora un semplice modello lineare e stocastico costituito dalle equazioni differenziali seguenti:

$$15) \quad dX(t) = -a \cdot X(t)dt + dW(t), \quad X(0) \approx N(0, \sigma_0^2), \quad W(t) \perp\!\!\!\perp X(0) ,$$

$$16) \quad dY(t) = X(t)dt + dV(t), \quad Y(0) = 0, \quad V(t) \perp\!\!\!\perp W(t) \perp\!\!\!\perp X_0 ,$$

ove la prima equazione e' gia' nota al lettore (si veda a pag. 3); la seconda afferma che le osservazioni sono costituite dalle variabili di stato alle quali si aggiunge un rumore aleatorio rappresentato da un secondo processo standardizzato di Wiener $V(t)$, indipendente da $W(t)$.

Più precisamente, l'equazione 16) suggerisce l'equivalente equazione $\dot{Y}(t) = X(t) + \dot{V}(t)$ ove il secondo addendo e' un disturbo Gaussiano formalmente coincidente con la derivata prima del processo di Wiener $V(t)$. Più' corretto e' interpretare l'equazione 16) come l'equazione integrale

$$Y(t) = Y(0) + \int_0^t X(s)ds + V(t) = \int_0^t X(s)ds + V(t)$$

ove l'integrale stocastico va inteso "in media quadratica".

Se $\sigma_0^2 = 1/2a$ si dimostra che le variabili di stato $X(t)$ formano un processo stazionario detto processo di Ornstein – Uhlenbeck e talvolta l'attuale modello con variabili latenti viene denominato in letteratura "osservazioni di un processo di Ornstein – Uhlenbeck affette da errori accidentali di misura".

Filtro di Kalman - Bucy

Un problema di stima analogo a quello posto per il modello precedente a tempo discreto consiste nella individuazione dello stimatore dei minimi quadrati per $X(t)$ basato sull'osservabilità del processo $Y(s)$ nell'intervallo $[0, t]$. Un famoso risultato del 1961, il teorema di Kalman – Bucy, mostra che lo stimatore ottimale $\hat{X}(t)$ e' dato dalla soluzione dell'equazione differenziale stocastica

$$17) \quad d\hat{X}(t) = -a.X(t)dt + P(t). \left[dY(t) - \hat{X}(t)dt \right], \quad \hat{X}(0) = 0,$$

ove la funzione deterministica $P(t) = E \left[X(t) - \hat{X}(t) \right]^2$, varianza dell'errore di stima, e' data dalla soluzione dell'equazione differenziale non lineare e deterministica, detta equazione di Riccati,

$$18) \quad \dot{P}(t) = 1 - P^2(t) - 2.a.P(t), \quad P(0) = \sigma_0^2 .$$

Per una semplice presentazione e dimostrazione del filtro di Kalman – Bucy suggeriamo al lettore il testo di M.H.A. Davis (paragrafo 4.4) citato in Riferimenti.

Filtro di Wonham per catene di Markov a tempo continuo con un numero finito di stati

Indicheremo con $\{X(t); t \geq 0\}$ una catena di Markov caratterizzata dalla matrice $(d \times d)$ delle intensità di transizione $G = [g_{ij}]$ e dalla distribuzione iniziale espressa dal vettore p_0 . Riesce utile per il seguito la rappresentazione della catena $X(t)$ data dalla $X(t) = \sum_{i=1}^d x_i . I_t(i)$ ove $I_t(i) = |X(t) = x_i|$. Il vettore I_t , con componenti $I_t(i)$, assume valori nell'insieme finito dei vettori unitari e ortogonali $\{e_1, \dots, e_d\}$ dello spazio euclideo R_d . Riesce $p_0 = E(I_0)$ e naturalmente $p_t = E(I_t)$.

E' noto che nello stato di informazione iniziale i vettori p_t sono determinati dall'equazione differenziale vettoriale $\frac{\partial}{\partial t} p_t = G^T . p_t$ la cui soluzione è espressa dalla

$$19) \quad p_t = \exp\{(t-s).G^T\}.p_s = \exp\{t.G^T\}.p_0, \quad \forall s \in [0, \infty).$$

Assumeremo che le variabili $X(t)$ non siano osservabili mentre invece risultino osservabili le variabili $Y(t) = \int_0^t g(X_s)ds + W(t)$, ove $W(t)$ denota un processo di Wiener standardizzato, indipendente dal processo $X(t)$, e $g(\cdot)$ una funzione misurabile e limitata, e indicheremo con π_t i vettori definiti dalle $\pi_t = E[I_t | F_t^Y]$, ove $F_t^Y = \sigma(Y_s; s \leq t)$. Enunceremo ora un importante risultato riguardante l'evoluzione dei vettori π_t e denominato filtro di Wonham e Shiriyayev:

Teorema: Il vettore $\pi_t = E[I_t | F_t^Y]$ soddisfa la seguente equazione differenziale stocastica

$$20) \quad d\pi_t = G^T \cdot \pi_t dt + \{diag[\pi_t(i)] - \pi_t \cdot \pi_t^T\} \cdot g \cdot [dY_t - x^T \cdot \pi_t dt]$$

ove g indica il vettore colonna avente le componenti $g(x_i)$, $i = 1, \dots, d$.

La soluzione della suddetta equazione differenziale stocastica si ottiene solitamente al modo seguente: introducendo il processo stocastico $\Phi_t = \exp\left\{\int_0^t g(x_s)dY_s - \frac{1}{2}\int_0^t g^2(x_s)ds\right\}$ e definendo la probabilità P^* , equivalente alla P , secondo la $dP^* = \Phi_t dP$ si ha

$$21) \quad \pi_t = E[I_t | F_t^Y] = \frac{E^*[I_t \cdot \Phi_t | F_t^Y]}{E^*[\Phi_t | F_t^Y]} = \frac{\rho_t}{\|\rho_t\|},$$

ove E^* è relativa alla probabilità P^* .

Si dimostra che il vettore ρ_t è soluzione della seguente equazione lineare stocastica di Zakai:

$$22) \quad d\rho_t = G^T \cdot \rho_t dt + \{diag[g(x_i)]\} \cdot \rho_t dY_t, \quad \rho_0 = p_0.$$

Trovato ρ_t , il vettore π_t si determina normalizzando ρ_t , cioè dividendo le sue componenti per $\sum_i \rho_t(i)$.

Per una differente linea dimostrativa si può vedere P. Chigansky – Non linear filtering a p. 117.

Per il caso $d = 2$, cioè per catene di Markov simmetriche con due soli stati, $x_1 = 1$ e

$x_2 = 0$, $p_0 = (1/2, 1/2)$ e matrice delle intensità $G = \begin{bmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{bmatrix}$ si ottiene per

$\pi_t(1) = \Pr[X(t) = 1 | F_t^Y]$ l'equazione differenziale stocastica seguente:

$$d\pi_t(1) = -2\lambda \left[\pi_t(1) - \frac{1}{2} \right] dt + [g(x_1) - g(x_2)] \cdot \pi_t(1) \cdot [1 - \pi_t(1)] \cdot \{ dY_t - [g(x_1) \cdot \pi_t(1) + g(x_2) \cdot (1 - \pi_t(1))] dt \}$$

che tiene conto della $\pi_t(1) + \pi_t(2) = 1$. Se $g(\cdot)$ è la funzione identica l'equazione precedente si semplifica diventando

$$d\pi_t(1) = \lambda \cdot [1 - 2 \cdot \pi_t(1)] dt + \pi_t(1) \cdot [1 - \pi_t(1)] \cdot [dY_t - \pi_t(1) dt];$$

si tratta pur sempre di un'equazione differenziale stocastica non lineare di tutto rispetto.

Infine osserviamo che in taluni problemi di teoria delle comunicazioni gli stimatori dei minimi quadrati per le variabili di stato non forniscono la risposta più appropriata alle aspettative: supponiamo che le variabili di stato possano assumere un numero finito di valori possibili, per esempio 0 oppure 1. Se le variabili osservabili forniscono le trasformazioni disturbate degli stati, e quindi una sequenza di numeri reali, può risultare importante individuare la sequenza di zeri e uni più probabile sulla base delle osservazioni e non una sequenza di stime dei minimi quadrati che sono necessariamente numeri reali nell'intervallo $[0, 1]$. Tale obiettivo è noto in letteratura come "massimizzazione della probabilità condizionata" ed è indicato brevemente con la sigla "sequenza-MAP" (Maximizing A Posterior probability). Un procedimento di stima che adotta questo criterio è l'algoritmo di A.J. Viterbi che può essere considerato un caso particolare del procedimento di "programmazione dinamica". Per informazioni ulteriori si veda per esempio A.M. Fraser in Riferimenti.

6. Alcuni approfondimenti sui problemi di stima

In questo paragrafo assumeremo che tutte le distribuzioni della terna rappresentativa $\{\nu(dx_0), \Lambda(x, dx'), \Psi(dy|x)\}$ ammettano una densità di probabilità rispetto alla misura di Lebesgue per cui scriveremo

$$\nu(dx_0) = \nu(x_0) dx_0, \quad \Lambda(x, dx') = f(x'|x) dx', \quad \Psi(dy|x) = f(y|x) dy.$$

L'uso della lettera f per entrambe le distribuzioni condizionate non dovrebbe ingenerare confusioni se si fa attenzione agli argomenti delle stesse.

Oltre al problema di filtraggio (filtering) che già conosciamo e che concerne la determinazione e la propagazione della densità di probabilità condizionata $f(x_n | Y_0^n)$ al variare di n , ci occuperemo del problema di interpolazione (smoothing) che concerne la determinazione e l'evoluzione della densità di probabilità condizionata $f(x_k | Y_0^n)$, con $k < n$, al variare di k con n fissato.

Note le tre densità di probabilità $\nu(x_0), f(x'|x)$ e $f(y|x)$ della terna rappresentativa, per ogni $x \in R$, la densità congiunta per la sequenza di variabili $X_0, Y_0, \dots, X_n, Y_n$ è determinata dalla

$$f(x_0, y_0, \dots, x_n, y_n) = f(x_0^n, y_0^n) = f(y_0^n / x_0^n) \cdot f(x_0^n) = \left[\prod_{k=0}^n f(y_k / x_k) \right] \cdot \left[\nu(x_0) \cdot \prod_{k=1}^n f(x_k | x_{k-1}) \right] = \\ = \nu(x_0) \cdot f(y_0 | x_0) \cdot \prod_{k=1}^n \left[f(y_k | x_k) \cdot f(x_k | x_{k-1}) \right] = \nu(x_0) \cdot f(y_0 | x_0) \cdot \prod_{k=1}^n f(x_k, y_k | x_{k-1})$$

Naturalmente, la densità congiunta delle sole variabili osservabili si ottiene dalla

$$f(y_0, \dots, y_n) = \int_{R^{n+1}} f(x_0^n, y_0^n) dx_0^n$$

e tale espressione può formalmente rappresentare anche la funzione di verosimiglianza degli eventuali elementi incogniti nel modello dinamico rispetto alla sequenza di osservazioni y_0^n .

Dalle due precedenti densità congiunte si possono facilmente determinare le densità condizionate $f(x_0^n / y_0^n)$, $f(x_k / y_0^n)$ e $f(x_n / y_0^n)$ che rivestono un ruolo centrale per i problemi di filtraggio e di interpolazione. Precisamente, le prime due risolvono un problema di “interpolazione congiunta” e, rispettivamente, di “interpolazione marginale”; l’ultima il problema di filtraggio e di essa abbiamo trattato a proposito dell’equazione ricorsiva 13).

Con riferimento al problema di interpolazione assumeremo fissato l’intero n e la sequenza osservata $(y_0, y_1, \dots, y_n)^T = y_0^n$ e ci porremo il problema di determinare le densità di probabilità $f(x_k / y_0^n)$ al variare di k tra 0 ed $n-1$. In lingua inglese tale problema è detto “fixed interval smoothing”. Si vedrà che anche in questo caso esiste un procedimento iterativo noto come “forward – backward recursion”.

Evidentemente, per $k < n$ si ha:

$$f(x_k / y_0^n) = \int_R \dots \int_R f(x_0^n / y_0^n) dx_0 dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n, \text{ essendo}$$

$$f(x_0^n / y_0^n) = \frac{f(x_0^n, y_0^n)}{f(y_0^n)}, \text{ nell'ipotesi che il denominatore sia positivo.}$$

Proposizione 2 : per $f(x_k / y_0^n)$ sussiste la seguente utile rappresentazione

$$f(x_k / y_0^n) = \frac{1}{f(y_0^n)} \cdot f(x_k, y_0^k) \cdot f(y_{k+1}^n | x_k),$$

ove

$$f(x_k, y_0^k) = \int_R \dots \int_R [v(x_0) \cdot f(y_0 | x_0) \cdot \prod_{j=1}^k f(y_j | x_j) \cdot f(x_j | x_{j-1})] dx_0 \dots dx_k \text{ e}$$

$$f(y_{k+1}^n | x_k) = \int_R \dots \int_R [\prod_{j=k+1}^n f(x_j | x_{j-1}) \cdot f(y_j | x_j)] dx_{k+1} \dots dx_n .$$

La verifica delle suddette espressioni è semplice.

L'utilità della precedente Proposizione consiste nella possibilità di avvalersi di relazioni ricorrenti per le due densità congiunte $f(x_k, y_0^k)$ e $f(y_{k+1}^n | x_k)$ espresse nella prossima proposizione.

Proposizione 3 : sussistono le seguenti relazioni ricorrenti

$$f(x_k, y_0^k) = f(y_k | x_k) \cdot \int_R f(x_k | x_{k-1}) \cdot f(x_{k-1} | y_0^{k-1}) dx_{k-1}, \quad f(x_0, y_0) = v(x_0) \cdot f(y_0 | x_0): \quad k = 1, 2, \dots$$

$$f(y_{k+1}^n | x_k) = \int_R f(y_{k+1} | x_{k+1}) \cdot f(x_{k+1} | x_k) \cdot f(y_{k+2}^n | x_{k+1}) dx_{k+1}, \quad f(y_n | x_n) = 1: \quad k = n - 2, n - 3, \dots .$$

Anche la verifica di queste relazioni è semplice.

Il procedimento ricorsivo denominato “forward – backward recursion” consiste nella determinazione delle due sequenze $\{f(x_k | y_0^k); k = 0, 1, \dots, n\}$ e

$\{f(y_{k+1}^n | x_k); k = n - 1, \dots, 0\}$ a partire dalla sequenza osservata y_0^n , dalla terna $\{v(x_0), f(x^1 | x), f(y | x); x \in R\}$ e dalle relazioni della Proposizione 3 ed infine nella

determinazione delle densità interpolanti $f(x_k | y_0^n) = \frac{1}{f(y_0^n)} \cdot [f(x_k, y_0^k) \cdot f(y_{k+1}^n | x_k)]$,

$k = 1, 2, \dots .$

Osserviamo che nella letteratura riguardante i modelli dinamici Markoviani con variabili latenti (Hidden Markov Models) le densità $f(x_k, y_0^k)$ e $f(y_{k+1}^n | x_k)$ sono indicate spesso con

i simboli $\alpha_k(x_k; y_0^k)$ e $\beta_{k/n}(x_k; y_{k+1}^n)$ e denominate “forward kernel” e “backward function” allo scopo di sottolineare la corrispondenza con le probabilità congiunte

$$\alpha_k(x_k; y_0^k) = P \left[\left(\bigcap_{j=0}^k Y_j = y_j \right) \cap (X_k = x_k) \right] \text{ e } \beta_{k/n}(y_{k+1}^n / x_k) = P \left[\left(\bigcap_{i=1}^{n-k} Y_{k+i} = y_{k+i} \right) \mid (X_k = x_k) \right]$$

che compaiono nei primi lavori degli anni 60 dell’altro secolo concernenti modelli Markoviani con spazio degli stati finiti.

Un’interessante relazione tra la densità interpolante $f(x_k | y_0^n)$, $k < n$, e quella della Proposizione 1 relativa al problema di filtraggio $f(x_k | y_0^k)$ è espressa dalla seguente

Proposizione 4 : sussiste la seguente relazione $f(x_k | y_0^n) = \left[\frac{f(y_{k+1}^n | x_k)}{f(y_{k+1}^n | y_0^k)} \right] \cdot f(x_k | y_0^k)$.

Dimostrazione: il secondo membro può essere riscritto come

$$\frac{f(y_{k+1}^n | x_k) \cdot f(x_k | y_0^k)}{f(y_{k+1}^n | y_0^k)} = \frac{f(x_k, y_0^n)}{f(y_0^k) \cdot f(y_{k+1}^n | y_0^k)} = \frac{f(x_k, y_0^n)}{f(y_0^n)} = f(x_k | y_0^n) .$$

Oltre ai problemi di filtraggio e di interpolazione si pongono anche problemi di **previsione**: note le osservazioni y_0^n e la terna $\{v(x_0), f(x'|x), f(y|x); x \in R\}$ può interessare la determinazione della densità di probabilità $f(x_{n+1} | y_0^n)$ o più in generale la densità $f(x_{n+m} | y_0^n)$ con $m > 1$ o infine la densità congiunta $f(x_{n+1}, x_{n+2}, \dots, x_{n+m} | y_0^n)$. Ci limitiamo ad osservare che, nota la densità $f(x_n | y_0^n)$, quella previsionale $f(x_{n+1} | y_0^n)$ si determina al modo seguente:

$$f(x_{n+1} | y_0^n) = \int_R f(x_{n+1} | x_n) \cdot f(x_n | y_0^n) dx_n$$

e lasciamo al lettore di determinare le altre densità previsionali menzionate.

APPENDICE A

Cenni sulle catene di Markov

1) Catene a tempo discreto

Indicheremo con $\{X_n; n \geq 0\}$ una catena di Markov caratterizzata dalla matrice $\Lambda = [\lambda_{ij}]$ delle probabilità condizionate di transizione $\lambda_{ij} = P(X_n = x_j | X_{n-1} = x_i)$ e dalla distribuzione iniziale $p_0(i) = P(X_0 = x_i)$, $i = 1, \dots, d$. Supporremo che le variabili della catena non siano osservabili, mentre siano osservabili le variabili $Y_n = X_n + \varepsilon_n$ ove le perturbazioni aleatorie ε_n sono assunte i.i.d., dotate della medesima densità di probabilità $f(\varepsilon)$, e indipendenti da X_0 .

E' noto che in assenza di osservazioni sulle variabili Y_n , cioè nello stato di informazione iniziale, le distribuzioni delle X_n , cioè i vettori p_n con elementi $p_n(i)$, sono determinate dalle relazioni

$$(1) \quad p_n = \Lambda^T \cdot p_{n-1} = (\Lambda^T)^n \cdot p_0 .$$

E' anche noto che il problema di stima ottimale dei minimi quadrati delle variabili della catena X_n , $n \geq 1$, basata sull'osservabilità delle Y_n consiste nella determinazione delle funzioni di regressione $E[X_n | Y_0, Y_1, \dots, Y_n] = E[X_n | F_n^Y]$, ove F_n^Y denota la σ -algebra generata dalla sequenza osservabile Y_0, Y_1, \dots, Y_n .

Se π_n denota il vettore delle probabilità $\pi_n(i) = P[X_n = x_i | F_n^Y] = E[X_n = x_i | F_n^Y]$ si ha

$$E[X_n | Y_0, Y_1, \dots, Y_n] = E[X_n | F_n^Y] = \sum_{i=1}^d x_i \cdot \pi_n(i) .$$

Sussiste per le probabilità $\pi_n(i)$ il seguente risultato

Teorema di W.M. Wonham:

$$(2) \quad \begin{aligned} \pi_0(i) &= P[X_0 = x_i | Y_0] = \frac{f(Y_0 - x_i) \cdot p_0(i)}{\sum_i f(Y_0 - x_i) \cdot p_0(i)} , \\ \pi_n(i) &= P[X_n = x_i | F_n^Y] = \frac{f(Y_n - x_i) \cdot \sum_h \lambda_{hi} \cdot \pi_{n-1}(h)}{\sum_i f(Y_n - x_i) \cdot \sum_h \lambda_{hi} \cdot \pi_{n-1}(h)} . \end{aligned}$$

Il vettore delle $\pi_n(i)$ ha l'espressione seguente:

$$(3) \quad \pi_n = [\pi_n(1), \dots, \pi_n(d)]^T = \frac{\text{diag}[f(Y_n - x_i)] \cdot \Lambda^T \cdot \pi_{n-1}}{[f(Y_n - x_1), \dots, f(Y_n - x_d)] \cdot \Lambda^T \cdot \pi_{n-1}} .$$

Daremo ora un cenno di dimostrazione del teorema facendo riferimento alla componente generica $\pi_n(i) = P[X_n = x_i | F_t^Y] = E[X_n = x_i | F_t^Y]$ del vettore π_n . Poiché $\pi_n(i)$ è misurabile rispetto a F_t^Y esiste una funzione $\Phi_i(\cdot)$ dei numeri aleatori che generano F_t^Y tale che $\pi_n(i) = \Phi_i(Y_1, \dots, Y_n)$. Per la definizione di funzione di regressione si ha:

$$E\{[X_n = x_i | \Phi_i(Y_1, \dots, Y_n)] \cdot \Psi(Y_1, \dots, Y_n)\} = 0$$

per ogni funzione $\Psi(\cdot)$ degli n numeri aleatori Y_j . Separando Y_n dai precedenti Y_1, \dots, Y_{n-1} e ponendo $\Psi(Y_1, \dots, Y_n) = \Psi_1(Y_n) \cdot \Psi_2(Y_1, \dots, Y_{n-1})$ possiamo scrivere l'uguaglianza precedente come

$$E\{[X_n = x_i | \Phi_i(Y_1, \dots, Y_{n-1}; Y_n)] \cdot \Psi_2(Y_1, \dots, Y_{n-1}) \cdot \Psi_1(Y_n)\} = 0$$

e anche come

$$E\{[X_n = x_i | \Phi_i(Y_1, \dots, Y_{n-1}; Y_n)] \cdot \Psi_1(Y_n) / F_{n-1}^Y\} = 0$$

o infine come

$$E\{X_n = x_i | \Psi_1(Y_n) / F_{n-1}^Y\} = E\{\Phi_i(Y_1, \dots, Y_{n-1}; Y_n) \cdot \Psi_1(Y_n) / F_{n-1}^Y\} .$$

L'ultima uguaglianza si può riscrivere, con qualche passaggio che evitiamo di specificare, come

$$\int_R \Psi_1(s) \cdot f(s - x_i) ds \cdot \sum_{j=1}^d \lambda_{ji} \cdot \pi_{n-1}(j) = \sum_{h=1}^d \int_R \Phi_i(Y_1, \dots, Y_{n-1}; s) \cdot \Psi_1(s) \cdot f(s - x_h) ds \cdot \sum_{j=1}^d \lambda_{jh} \cdot \pi_{n-1}(j)$$

e per l'arbitrarietà di $\Psi_1(\cdot)$ si ottiene

$$\pi_n(i) = \Phi_i(Y_1, \dots, Y_{n-1}; s) = \frac{f(s - x_i) \cdot \sum_h \lambda_{hi} \cdot \pi_{n-1}(h)}{\sum_i f(s - x_i) \cdot \sum_h \lambda_{hi} \cdot \pi_{n-1}(h)} .$$

2) Catene a tempo continuo

Indicheremo con $\{X(t); t \geq 0\}$ una catena di Markov caratterizzata dalla matrice delle intensità di transizione $G = [g_{ij}]$ e dalla distribuzione iniziale espressa dal vettore p_0 .

Riesce utile per il seguito la rappresentazione della catena $X(t)$ data dalla $X(t) = \sum_{i=1}^d x_i \cdot I_t(i)$

ove $I_t(i) = |X(t) = x_i|$. Il vettore I_t assume valori nell'insieme finito dei vettori unitari e ortogonali $\{e_1, \dots, e_d\}$ dello spazio euclideo R_d . Riesce $p_0 = E(I_0)$ e naturalmente $p_t = E(I_t)$.

E' noto che nello stato di informazione iniziale i vettori p_t sono determinati dall'equazione differenziale vettoriale $\frac{\partial}{\partial t} p_t = G^T \cdot p_t$ la cui soluzione è espressa dalla

$$(4) \quad p_t = \exp\{(t-s) \cdot G^T\} \cdot p_s = \exp\{t \cdot G^T\} \cdot p_0, \quad \forall s \in [0, \infty).$$

Assumeremo ora che le variabili $X(t)$ non siano osservabili mentre invece risultino osservabili le variabili $Y(t) = \int_0^t g(X_s) ds + W(t)$, ove $W(t)$ denota un processo di Wiener standardizzato, indipendente dal processo $X(t)$, e $g(\cdot)$ una funzione misurabile e limitata, e indicheremo con π_t i vettori definiti dalle $\pi_t = E[I_t | F_t^Y]$, ove $F_t^Y = \sigma(Y_s; s \leq t)$. Enunceremo ora un importante risultato riguardante l'evoluzione dei vettori π_t e denominato

Filtro di M. Wonham: Il vettore $\pi_t = E[I_t | F_t^Y]$ soddisfa la seguente equazione differenziale

$$(5) \quad d\pi_t = G^T \cdot \pi_t dt + \{diag[\pi_t(i)] - \pi_t \cdot \pi_t^T\} \cdot g \cdot [dY_t - x^T \cdot \pi_t dt]$$

ove g^T indica il trasposto del vettore avente le componenti $g(x_i)$, $i = 1, \dots, d$.

La soluzione della suddetta difficile equazione differenziale stocastica si ottiene solitamente al modo seguente: introducendo il processo stocastico

$$\Phi_t = \exp\left\{\int_0^t g(x_s) dY_s - \frac{1}{2} \int_0^t g^2(x_s) ds\right\}$$

e definendo la probabilità P^* , equivalente alla P , secondo la $dP^* = \Phi_t dP$ si ha

$$(6) \quad \pi_t = E[I_t | F_t^Y] = \frac{E^*[I_t \cdot \Phi_t | F_t^Y]}{E^*[\Phi_t | F_t^Y]} = \frac{\rho_t}{\|\rho_t\|}.$$

Si dimostra che il vettore ρ_t è soluzione della seguente equazione lineare stocastica di Zakai:

$$(7) \quad d\rho_t = G^T \cdot \rho_t dt + \{diag [g(x_i)]\} \cdot \rho_t dY_t, \quad \rho_0 = p_0.$$

Trovato ρ_t , il vettore π_t si determina normalizzando ρ_t , cioè dividendo le sue componenti per $\sum_i \rho_t(i)$.

Per il caso $d = 2$, cioè per catene di Markov simmetriche con due soli stati, $x_1 = 1$ e $x_2 = 0$, $p_0 = (1/2, 1/2)$ e matrice delle intensità $G = \begin{bmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{bmatrix}$ si ottiene per

$\pi_t(1) = \Pr [X(t) = 1 | F_t^Y]$ l'equazione differenziale stocastica seguente:

$$d\pi_t(1) = -2\lambda \cdot \left[\pi_t(1) - \frac{1}{2} \right] dt + [g(x_1) - g(x_2)] \cdot \pi_t(1) \cdot [1 - \pi_t(1)] \cdot \{dY_t - [g(x_1) \cdot \pi_t(1) + g(x_2) \cdot (1 - \pi_t(1))] dt\}$$

che tiene conto della $\pi_t(1) + \pi_t(2) = 1$. Se $g(\cdot)$ è la funzione identica l'equazione precedente si semplifica sensibilmente:

$$d\pi_t(1) = \lambda \cdot [1 - 2\pi_t(1)] dt + \pi_t(1) \cdot [1 - \pi_t(1)] \cdot [dY_t - \pi_t(1) dt].$$

RIFERIMENTI

- Arnold L. – *Stochastic Differential Equations: Theory and Applications*. J. Wiley, 1974.
- Arnold L. – *Random Dynamical Systems*. Springer, 1995.
- Bhattacharya R., Majumdar M. – Random dynamical systems: a review. *Economic Theory*, 2004.
- Cappè, Moulines, Rydèn – *Inference in Hidden Markov Models*. Springer, 2009.
- Chigansky P. – Introduction to stochastic processes. Lecture Notes (On line).
- Chigansky P. – Nonlinear filtering. Lecture Notes (On line).
- Chigansky P. – Hidden Markov Models. Lecture Notes (On line).
- Davis M.H.A. – *Linear Estimation and Stochastic Control*. Chapman and Hall, 1977.
- Ephraim Y., Merhav N. – Hidden Markov Processes. *IEEE Trans. on Inf. Theory*, 2002.
- Fraser A.M. – *Hidden Markov Models and Dynamical Systems*. SIAM (2008).
- Grimmett G., Stirzaker D. – *Probability and Random Processes*. Oxford University Press, 2009.
- van Handel R. – Hidden Markov Models. Lecture Notes (On line)
- Kifer Y. – *Ergodic Theory of Random Transformations*. Birkhauser, 1986.
- Rabiner L.R. – A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*, 1989.
- Shiryayev A.N. – *Probability*. Springer-Verlag, 1984.