

UNIVERSITÁ DEGLI STUDI DI PADOVA

Facoltà di Scienze Statistiche

Corso di Laurea Specialistica in Scienze Statistiche Economiche,
Finanziarie e Aziendali

Tesi di Laurea

**METODI MONTE CARLO SEQUENZIALI
PER LA TEORIA DEL FILTRAGGIO
UN MODELLO PER MEDIA E VARIANZA STOCASTICA**

Relatore Prof. MARCO FERRANTE

Laureanda NADIA FRIGO

ANNO ACCADEMICO 2005-2006

Indice

Elenco delle figure	iii
Introduzione	v
1 Modelli state space	1
1.1 Specificazione dei modelli state space	2
1.2 Problemi associati ai modelli state space	3
1.2.1 Previsione	3
1.2.2 Filtraggio	5
1.2.3 Smoothing	6
2 Metodi per il filtraggio e lo smoothing	9
2.1 Metodi esatti	9
2.1.1 Modelli markoviani nascosti	10
2.1.2 Filtro di Kalman	10
2.2 Metodi approssimati	13
2.2.1 Filtro di Kalman esteso (FKE)	13
2.2.2 Altri metodi approssimati	13
2.3 Metodi Monte Carlo	14
2.3.1 Campionamento Monte Carlo perfetto	14
2.3.2 Algoritmi Markov Chain Monte Carlo	15
2.3.3 Campionamento per importanza	18
2.3.4 Filtro bootstrap	20

3	Modello per media e varianza stocastica	25
3.1	Filtro finito-dimensionale	25
3.2	Modello	27
3.2.1	Caso $\lambda \in [0, \frac{1}{2})$	28
3.2.2	Caso generale	33
4	Filtraggio nel modello per media e varianza	37
4.1	Modello	37
4.2	Simulazione	38
4.2.1	Convergenza	42
4.2.2	Stabilitá	44
5	Convergenza di metodi Monte Carlo sequenziali	55
5.1	Notazione e terminologia	55
5.1.1	Processi di Markov e nuclei di transizione	56
5.1.2	Modello considerato	57
5.2	Filtro Ottimo	58
5.3	Particle filter	59
5.3.1	Algoritmo	59
5.3.2	Meccanismi di ramificazione	62
5.4	Teoremi sulla convergenza	66
5.4.1	Considerazioni preliminari	67
5.4.2	Applicazione al Filtro Ottimo	69
6	Convergenza nel modello per media e varianza	73
6.1	Proprietá di Feller	74
6.1.1	Modello state space non lineare	74
6.2	Condizioni sulla funzione di verosimiglianza	76
A	Funzioni di Bessel e distribuzione GIG	79
A.1	Funzione di Bessel modificata del secondo tipo	85
A.2	Le funzioni D_ν e R_ν	88
A.3	Proprietá statistiche della distribuzione GIG	88
	Bibliografia	96

Elenco delle figure

2.1	Schema del filtro bootstrap	23
4.1	Processi Simulati	39
4.2	Istogramma e distribuzione esatta di $X_n Y_{0:n}$	40
4.3	Istogramma e distribuzione esatta di $Z_n Y_{0:n}$	42
4.4	Distribuzione empirica congiunta e curve di livello $n = 100$. .	43
4.5	Distribuzione empirica congiunta e curve di livello $n = 150$. .	44
4.6	Istogramma e distribuzione esatta di $X_n Y_{0:n}$	45
4.7	Istogramma e distribuzione esatta di $Z_n Y_{0:n}$	46
4.8	Distribuzione empirica congiunta e curve di livello $n = 100$. .	47
4.9	Distribuzione empirica congiunta e curve di livello $n = 150$. .	47
4.10	Istogramma e distribuzione esatta di $X_n Y_{0:n}$	48
4.11	Istogramma e distribuzione esatta di $Z_n Y_{0:n}$	49
4.12	Distribuzione empirica congiunta e curve di livello $n = 90$. .	50
4.13	Distribuzione empirica congiunta e curve di livello $n = 150$. .	51
4.14	Istogramma e distribuzione esatta di $X_n Y_{0:n}$	52
4.15	Istogramma e distribuzione esatta di $Z_n Y_{0:n}$	53
4.16	Distribuzione empirica congiunta e curve di livello $n = 100$. .	54
4.17	Distribuzione empirica congiunta e curve di livello $n = 150$. .	54
5.1	Alberi binari di ramificazione	64
A.1	Grafico della funzione di Bessel del primo tipo $J_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$	80
A.2	Grafico della funzione di Bessel del primo tipo $J_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$	81
A.3	Grafico della funzione di Bessel del secondo tipo $Y_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$	82

A.4	Grafico della funzione di Bessel del secondo tipo $Y_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$	83
A.5	Grafico della funzione di Bessel modificata del primo tipo $I_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$	84
A.6	Grafico della funzione di Bessel modificata del primo tipo $I_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$	85
A.7	Grafico della funzione di Bessel modificata del secondo tipo $K_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$	86
A.8	Grafico della funzione di Bessel modificata del secondo tipo $K_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$	87
A.9	Grafico di $R_\lambda(\cdot)$ per diversi valori di λ	91
A.10	Grafico di $R(x)$ per diversi valori di x	92
A.11	Grafico di $D_\lambda(\cdot)$ per diversi valori di λ	93
A.12	Grafico di $D(x)$ per diversi valori di x	94

Introduzione

Molti problemi di analisi di dati reali richiedono la stima di quantità non note sulla base di una successione di osservazioni incomplete o poco precise. In molti casi le osservazioni si rendono disponibili una dopo l'altra, in modo sequenziale nel tempo. Spesso situazioni di questo tipo possono essere rappresentate mediante modelli *state space*. Tali modelli sono costituiti da un processo stocastico sottostante non osservabile, chiamato processo del segnale o degli stati o processo latente, avente una semplice dinamica markoviana e da un processo stocastico che rappresenta le osservazioni, legato al processo del segnale da particolari relazioni di dipendenza. Tutta l'inferenza statistica, sia quella riguardante il modello, sia quella riguardante il segnale o le osservazioni future, deve essere effettuata sulla base delle osservazioni disponibili, in quanto il processo del segnale non può essere osservato.

Il problema del filtraggio stocastico consiste nel calcolare in ogni istante di tempo le distribuzioni condizionate del segnale date le osservazioni disponibili fino a quell'istante. Due successive distribuzioni di filtraggio sono legate da una formula ricorsiva che può essere scomposta in due passi: il primo consiste nell'effettuare una previsione sulla base della struttura dinamica markoviana del processo del segnale, mentre il secondo, detto passo di aggiornamento, utilizza l'informazione aggiuntiva data dalla nuova osservazione per correggere la previsione mediante una applicazione della formula di Bayes. Il calcolo esatto delle soluzioni di queste equazioni ricorsive è possibile soltanto in pochi casi particolari. In generale è quindi necessario ricorrere a metodi approssimati. A tale scopo, un interesse crescente è stato rivolto recentemente allo sviluppo e allo studio di metodi Monte Carlo sequenziali. Tali metodi, chiamati anche *particle filters* o filtri Monte Carlo ricorsivi, sono algoritmi iterativi che generano e aggiornano ricorsivamente un insieme di valori simulati (detti *particles*) allo scopo di fornire una approssi-

mazione per le distribuzioni condizionate di interesse. I particles evolvono nel tempo cambiando posizione e generando un numero di discendenti, secondo una dinamica in due passi che ‘segue da vicino’ l’evoluzione descritta dalla formula ricorsiva esatta. A differenza degli altri metodi approssimati, i metodi Monte Carlo sequenziali sono molto flessibili, facili da implementare e applicabili in contesti molto generali. Ad esempio essi vengono applicati, oltre che nell’analisi di serie storiche e in problemi di trattamento del segnale, nel riconoscimento di discorsi, nella visione artificiale, nel rilevamento degli oggetti, in problemi di controllo, in matematica finanziaria, in econometria, in geofisica, in biologia molecolare e in genetica.

Nel Capitolo 1 si definir  in modo formale un modello state space generale e si presenteranno delle formule ricorsive esatte per il problema di filtraggio.

Nel Capitolo 2 si illustreranno metodi ricorsivi esatti e approssimati per risolvere le formule ricorsive presentate nel Capitolo 1.

Nel Capitolo 3 si presenter  un problema di filtraggio per un processo stocastico parzialmente osservabile che fornisce una modellizzazione per la media e la varianza di osservazioni normalmente distribuite e si prover  l’esistenza di un filtro finito dimensionale per questo modello.

Nel Capitolo 4 verranno riportati alcuni risultati di tipo empirico ottenuti implementando l’algoritmo Monte Carlo sequenziale (particle filter) descritto nel Capitolo 2 al modello state space presentato nel Capitolo 3.

Nel Capitolo 5 si forniranno delle condizioni sufficienti per la convergenza dell’algoritmo particle filter alle distribuzioni di filtraggio e di previsione in ogni istante di tempo fissato.

Nel Capitolo 6 si applicheranno i risultati sulla convergenza ottenuti nel Capitolo 5 al modello per media e varianza stocastica presentato nel Capitolo 3.

Nell’Appendice A si presenteranno alcune propriet  statistiche della distribuzione gaussiana inversa generalizzata che verranno utilizzate per determinare il filtro finito dimensionale per il modello descritto nel Capitolo 3. Inoltre si forniranno la definizione e alcune considerazioni sulla funzione di Bessel che compare come fattore di normalizzazione nella densit  di una variabile con distribuzione gaussiana inversa generalizzata.

Capitolo 1

Modelli state space

Un modello state space in generale assume che esista un processo sottostante non osservabile, chiamato *processo degli stati* (o *processo latente* o *segnale*), avente una semplice dinamica markoviana e che le osservazioni di questo processo siano incomplete, affette da rumore oppure non disponibili.

Questi modelli furono introdotti per la prima volta nei primi anni Sessanta in ingegneria da Kalman e Bucy (Kalman, 1960, Kalman e Bucy, 1961) e da Baum con quello che attualmente é chiamato algoritmo *Expectation-Maximization* (EM) (Baum e Petrie, 1966). Giá da allora questi metodi furono ampiamente usati in problemi di controllo e per il riconoscimento verbale, ma soltanto in seguito fu riconosciuta la loro importanza dal punto di vista statistico. Ciò puó sorprendere poiché concetti di variabili latenti e di stima ricorsiva erano presenti in statistica giá da lungo tempo, e casi particolari del filtro di Kalman si possono trovare nella letteratura statistica giá prima del 1960 (si veda Plackett, 1950 e Lauritzen, 1981). I modelli state space apparvero nella letteratura delle serie temporali negli anni Settanta (Akaike, 1974, Harrison e Stevens, 1976) e si affermarono durante gli anni Ottanta.

Negli anni Novanta essi sono diventati un centro di interesse. Ciò é dovuto a un campo di applicabilitá ancora piú ampio, che include per esempio la biologia molecolare e la genetica, e alla possibilitá di trattare molte situazioni di non linearitá e non gaussianitá con moderni metodi Monte Carlo.

In questo capitolo si dará una introduzione ai modelli state space e markoviano nascosti e saranno presentate delle ricorsioni utili per sviluppare metodi ricorsivi per questi modelli.

1.1 Specificazione dei modelli state space

Un *modello state space generale* consiste in un processo non osservabile $\{X_n\}_{n \in \mathbb{N}}$, chiamato *processo degli stati* e in un processo $\{Y_n\}_{n \in \mathbb{N}^+}$ che rappresenta la serie temporale delle osservazioni derivanti da $\{X_n\}_{n \in \mathbb{N}}$. Si assumono le seguenti proprietà:

- (M1) $\{X_n\}_{n \in \mathbb{N}}$ é un processo di Markov;
- (M2) Condizionatamente a X_0, X_1, \dots, X_N le osservazioni Y_1, Y_2, \dots, Y_N sono indipendenti e per ogni $n = 1, \dots, N$ si ha che Y_n dipende solo da X_n .

Si suppone inoltre che le variabili X_n e Y_n assumano valori negli spazi misurabili $(\mathcal{X}, \mathcal{F})$ e $(\mathcal{Y}, \mathcal{G})$ rispettivamente. Si indicheranno con μ e ν le rispettive misure di riferimento su tali spazi. Definiamo:

- $a_0(\cdot)$ la funzione di densità secondo cui é distribuita la variabile X_0
- $a_n(x_{n-1}, x)$ per $n \geq 1$ la densità di transizione al tempo n dallo stato x_{n-1} allo stato x
- $b_n(x_n, y)$ per $n \geq 1$ la funzione di densità della variabile Y_n quando il processo latente si trova in x_n .

Schematicamente

$$\begin{aligned} X_0 &\sim a_0(x)d\mu(x) \\ X_n|X_{n-1} = x_{n-1} &\sim a_n(x_{n-1}, x)d\mu(x) \\ Y_n|X_n = x_n &\sim b_n(x_n, y)d\nu(x). \end{aligned}$$

Un modello state space generale risulta completamente specificato non appena si definiscono le funzioni di densità $a_0(x)$, $a_n(x_{n-1}, x)$ e $b_n(x_n, y)$ per $n \geq 1$ e può essere scritto in forma molto flessibile come

$$X_n = g_n(X_{n-1}, V_n) \tag{1.1}$$

$$Y_n = h_n(X_n, W_n) \tag{1.2}$$

con g_n e h_n funzioni arbitrarie.

Quando la legge di distribuzione del processo di Markov latente é discreta, il modello prende il nome di *modello markoviano nascosto*, quando tale legge é continua il modello viene chiamato *state space*. Gli sviluppi di questi due modelli sono avvenuti prevalentemente in maniera separata, sebbene

le ricorsioni e gli algoritmi di base siano strettamente collegati. In questo lavoro si userá il termine modello state space per indicare entrambi i casi; se necessario tale distinzione verrá evidenziata adoperando la terminologia opportuna.

1.2 Problemi associati ai modelli state space

Tramite l'applicazione di un modello state space si possono affrontare due tipi di problemi.

Il primo riguarda la stima dei possibili valori assunti dal processo latente nei vari istanti, in base ad un certo numero di osservazioni. In questo caso si suppongono note le densitá a_n e b_n e si stima la funzione di densitá condizionata di X_n date le osservazioni y_1, \dots, y_s . In base alla relazione temporale esistente tra s e n si distinguono tre tipi di problemi: *smoothing* (se $s > n$), *filtering* (se $s = n$) e *previsione* (se $s < n$).

Il secondo problema é legato alla stima dei parametri non noti introdotti nelle densitá che specificano il modello. Il problema dell'inferenza su parametri non noti in a_n e b_n rappresenta forse il problema piú interessante dal punto di vista statistico, ma essere in possesso di algoritmi veloci ed efficienti per risolvere il primo tipo di problema é un prerequisito indispensabile per calcolare, ad esempio, la funzione di verosimiglianza.

Nei seguenti paragrafi si presenteranno in maniera formale i problemi di filtraggio, smoothing e previsione in termini di ricorsioni. Tramite le densitá di filtraggio é possibile esprimere la funzione di densitá congiunta delle osservazioni. Tale funzione é alla base della stima dei parametri non noti tramite la funzione di verosimiglianza.

In seguito useremo la seguente notazione: per $s < n$ si definisce $y_{s:n} = (y_s, y_{s+1}, \dots, y_n)$ e analogamente per $x_{s:n}$. La densitá condizionata di X_s dato $Y_{1:n} = y_{1:n}$ verrá indicata con $f_{s|n}(x_s|y_{1:n})$.

Lo scopo é quello di stimare in modo ricorsivo nel tempo la densitá a posteriori $p(x_{0:n}|y_{1:n})$ e le funzioni ad essa associate (incluse le densitá marginali $f_{s|n}(x_s|y_{1:n})$ per $s \leq n$, cioé le densitá di filtraggio e di smoothing).

1.2.1 Previsione

Il problema di previsione riguarda la stima della densitá di X_n date le osservazioni $y_{1:s}$ con $0 < s < n$. Si chiama *densitá di previsione in k passi*, con $k \geq 1$, la densitá di X_{n+k} dato $y_{1:n}$.

Il lemma seguente definisce una formula ricorsiva che permette di calcolare le densità di previsione a partire dalla densità di filtraggio $f_{n|n}(x_n|y_{1:n})$, sfruttando la markovianità del processo latente.

Lemma 1.2.1 *Le densità di previsione per gli stati possono essere calcolate dalle densità di filtraggio secondo la seguente ricorsione in $k \geq 1$*

$$f_{n+k|n}(x_{n+k}|y_{1:n}) = \int a_{n+k}(x, x_{n+k}) f_{n+k-1|n}(x|y_{1:n}) d\mu(x). \quad (1.3)$$

Dimostrazione. Condizionando a $X_{n+k-1} = x_{n+k-1}$ si ha

$$f_{n+k|n}(x_{n+k}|y_{1:n}) = \int p(x_{n+k}|y_{1:n}, x_{n+k-1}) p(x_{n+k-1}|y_{1:n}) d\mu(x_{n+k-1}).$$

Dato $X_{n+k-1} = x_{n+k-1}$, X_{n+k} e $Y_{1:n}$ sono indipendenti e quindi segue immediatamente

$$f_{n+k|n}(x_{n+k}|y_{1:n}) = \int a_{n+k}(x, x_{n+k}) f_{n+k-1|n}(x|y_{1:n}) d\mu(x).$$

□

Le densità di previsione permettono di calcolare anche la densità di una osservazione futura, Y_{n+k} , date le osservazioni $y_{1:n}$, come si dimostra nel seguente lemma.

Lemma 1.2.2 *Le densità di previsione per le osservazioni possono essere calcolate dalle densità di previsione per gli stati secondo l'equazione*

$$p(y_{n+k}|y_{1:n}) = \int b_{n+k}(y_{n+k}, x) f_{n+k|n}(x|y_{1:n}) d\mu(x). \quad (1.4)$$

Dimostrazione. Condizionando rispetto all'evento $X_{n+k} = x_{n+k}$ si ha

$$p(y_{n+k}|y_{1:n}) = \int p(y_{n+k}|y_{1:n}, x_{n+k}) p(x_{n+k}|y_{1:n}) d\mu(x_{n+k}).$$

Dato $X_{n+k} = x_{n+k}$, Y_{n+k} e $Y_{1:n}$ sono indipendenti e quindi

$$p(y_{n+k}|y_{1:n}) = \int p(y_{n+k}|x_{n+k}) p(x_{n+k}|y_{1:n}) d\mu(x_{n+k})$$

$$= \int b_{n+k}(y_{n+k}, x) f_{n+k|n}(x|y_{1:n}) d\mu(x).$$

□

Grazie alle formule ricorsive viste nei Lemmi 1.2.1 e 1.2.2 é possibile calcolare in modo ricorsivo la funzione di densitá congiunta di $Y_{1:N}$:

$$\begin{aligned} p(y_{1:N}) &= \prod_{n=1}^N p(y_n|y_{1:n-1}) = \\ &= \int \dots \int a_0(x_0) \prod_{n=1}^N a_n(x_{n-1}, x_n) b_n(y_n, x_n) \prod_{n=0}^N d\mu(x_n), \end{aligned}$$

intendendo $p(y_1|y_0) \equiv p(y_1)$.

1.2.2 Filtraggio

Il problema del filtraggio riguarda la stima della densitá condizionata di X_n date le osservazioni fino al tempo n , $y_{1:n}$, detta appunto *densitá di filtraggio*. Il seguente teorema definisce una formula ricorsiva in avanti che permette il calcolo delle densitá di filtraggio a partire dalla densitá iniziale $a_0(\cdot)$.

Teorema 1.2.3 *Le densitá di filtraggio possono essere calcolate secondo la seguente formula ricorsiva in avanti in n , per $n > 0$*

$$\begin{aligned} f_{n+1|n+1}(x_{n+1}|y_{1:n+1}) &= \frac{b_{n+1}(y_{n+1}, x_{n+1}) f_{n+1|n}(x_{n+1}|y_{1:n})}{p(y_{n+1}|y_{1:n})} \\ &= \frac{b_{n+1}(y_{n+1}, x_{n+1}) \int a_{n+1}(x, x_{n+1}) f_{n|n}(x|y_{1:n}) d\mu(x)}{p(y_{n+1}|y_{1:n})}, \end{aligned} \quad (1.5)$$

dove la base della ricorsione é $f_{0|0}(\cdot) = a_0(\cdot)$.

Dimostrazione. Applicando il teorema di Bayes

$$f_{n+1|n+1}(x_{n+1}|y_{1:n+1}) = p(x_{n+1}|y_{n+1}, y_{1:n}) = \frac{p(y_{n+1}|x_{n+1}, y_{1:n}) p(x_{n+1}|y_{1:n})}{p(y_{n+1}|y_{1:n})}.$$

Dato $X_{n+1} = x_{n+1}$, Y_{n+1} e $Y_{1:n}$ sono indipendenti, quindi

$$\frac{p(y_{n+1}|x_{n+1}, y_{1:n}) p(x_{n+1}|y_{1:n})}{p(y_{n+1}|y_{1:n})} = \frac{p(y_{n+1}|x_{n+1}) p(x_{n+1}|y_{1:n})}{p(y_{n+1}|y_{1:n})}.$$

Per concludere, basta osservare che il primo fattore é proprio $b_{n+1}(y_{n+1}, x_{n+1})$ e sostituire il secondo fattore, che coincide con la densit  di previsione ad un passo, con la formula data dal Lemma 1.2.1.

□

Si noti che il denominatore in (1.5) é soltanto un fattore di normalizzazione. La dimostrazione del Teorema 1.2.3 mette in evidenza come la procedura ricorsiva di filtraggio alterni due fasi: la prima ha lo scopo di calcolare la densit  di previsione ad un passo (fase di *previsione*), nella seconda la densit  di previsione viene utilizzata assieme all'osservazione corrente per ottenere la densit  di filtraggio all'istante successivo (fase di *aggiornamento*). In modo schematico:

$$\text{previsione} \quad f_{n|n-1}(x_n|y_{1:n-1}) = \int f_{n-1|n-1}(x|y_{1:n-1})a_n(x, x_n)d\mu(x) \quad (1.6)$$

$$\text{aggiornamento} \quad f_{n|n}(x_n|y_{1:n}) = \frac{b_n(y_n, x_n)f_{n|n-1}(x_n|y_{1:n-1})}{\int b_n(y_n, x)f_{n|n-1}(x|y_{1:n-1})d\mu(x)}. \quad (1.7)$$

1.2.3 Smoothing

Il problema dello smoothing consiste nello stimare, fissato un istante temporale N , la densit  di X_n , con $n < N$, condizionatamente al vettore delle osservazioni $y_{1:N}$. Consideriamo la catena di Markov X_0, X_1, \dots, X_N dato $y_{1:N}$, le cui densit  di transizione sono definite dal seguente lemma.

Lemma 1.2.4 *Condizionatamente a $y_{1:N}$, (X_0, X_1, \dots, X_N) é una catena di Markov non omogenea con*

1. *densit  di transizione in avanti*

$$\begin{aligned} p(x_n|x_{n-1}, y_{1:N}) &= p(x_n|x_{n-1}, y_{n:N}) \\ &= \frac{a_n(x_{n-1}, x_n)b_n(y_n, x_n)p(y_{n+1:N}|x_n)}{p(y_{n:N}|x_{n-1})} \end{aligned}$$

dove

$$p(y_{n:N}|x_{n-1}) = \int a_n(x_{n-1}, x_n)b_n(y_n, x_n)p(y_{n+1:N}|x_n)d\mu(x_n); \quad (1.8)$$

2. densità di transizione all'indietro

$$p(x_n|x_{n+1}, y_{1:N}) = p(x_n|x_{n+1}, y_{1:n}) = \frac{a_{n+1}(x_n, x_{n+1})f_{n|n}(x_n|y_{1:n})}{f_{n+1|n}(x_{n+1}|y_{1:n})}. \quad (1.9)$$

Dimostrazione.

Punto 1. Per la notazione introdotta si ha che

$$p(x_n|x_{n-1}, y_{1:N}) = p(x_n|x_{n-1}, y_{1:n-1}, y_{n:N}),$$

inoltre X_n e $Y_{1:n-1}$ sono indipendenti dati $X_{n-1} = x_{n-1}$ e $Y_{n:N} = y_{n:N}$. Applicando il teorema di Bayes si ricava

$$\begin{aligned} p(x_n|x_{n-1}, y_{1:n-1}, y_{n:N}) &= p(x_n|x_{n-1}y_{n:N}) = \frac{p(x_n, x_{n-1}, y_{n:N})}{p(x_{n-1}, y_{n:N})} \\ &= \frac{p(x_n, y_{n:N}|x_{n-1})p(x_{n-1})}{p(y_{n:N}|x_{n-1})p(x_{n-1})} \\ &= \frac{p(y_{n:N}|x_{n-1}, x_n)p(x_n|x_{n-1})}{p(y_{n:N}|x_{n-1})} = \frac{p(y_{n:N}|x_{n-1}, x_n)a_n(x_{n-1}, x_n)}{p(y_{n:N}|x_{n-1})}. \end{aligned}$$

Dato $X_n = x_n$, $Y_{n:N}$ e X_{n-1} sono indipendenti, quindi

$$\begin{aligned} p(y_{n:N}|x_{n-1}, x_n) &= p(y_{n:N}|x_n) = p(y_n, y_{n+1:N}|x_n) \\ &= p(y_{n+1:N}|x_n, y_n)p(y_n|x_n) = p(y_{n+1:N}|x_n, y_n)b_n(y_n, x_n). \end{aligned}$$

Inoltre, dato $X_n = x_n$, anche $Y_{n+1:N}$ e Y_n sono indipendenti e per tale motivo si elimina la dipendenza da y_n nel primo fattore. Ritornando alla formula ricavata applicando il teorema di Bayes e sostituendo nel primo fattore quanto appena osservato si ha

$$\begin{aligned} p(x_n|y_{1:N}, x_{n-1}) &= \frac{p(y_{n:N}|x_{n-1}, x_n)a_n(x_{n-1}, x_n)}{p(y_{n:N}|x_{n-1})} \\ &= \frac{p(y_{n+1:N}|x_n)b_n(y_n, x_n)a_n(x_{n-1}, x_n)}{p(y_{n:N}|x_{n-1})}. \end{aligned}$$

Condizionando a $X_n = x_n$,

$$\begin{aligned} p(y_{n:N}|x_{n-1}) &= \int p(y_{n:N}|x_{n-1}, x_n)p(x_{n-1}|x_n)d\mu(x_n) \\ &= \int b_n(y_n, x_n)p(y_{n+1:N}|x_n)a_n(x_{n-1}, x_n)d\mu(x_n). \end{aligned}$$

Punto 2. Scomponendo il vettore delle osservazioni e osservando che dato $X_{n+1} = x_{n+1}$ e $Y_{1:n} = y_{1:n}$, X_n e $Y_{n+1:n}$ sono indipendenti, si può scrivere

$$p(x_n|x_{n+1}, y_{1:N}) = p(x_n|x_{n+1}, y_{1:n}, y_{n+1:N}) = p(x_n|x_{n+1}, y_{1:n}). \quad (1.10)$$

Applicando il Teorema di Bayes si ottiene

$$\begin{aligned} p(x_n|x_{n+1}, y_{1:N}) &= \frac{p(x_{n+1}|x_n, y_{1:n})p(x_n|y_{1:n})}{p(x_{n+1}|y_{1:n})} = \frac{p(x_{n+1}|x_n)p(x_n|y_{1:n})}{p(x_{n+1}|y_{1:n})} \\ &= \frac{a_{n+1}(x_n, x_{n+1})f_{n|n}(x_n|y_{1:n})}{f_{n+1|n}(x_{n+1}|y_{1:n})}. \end{aligned} \quad (1.11)$$

□

Grazie alla (1.9), é possibile calcolare le distribuzioni condizionate di $X_{s:n}$ dato $Y_{1:N} = y_{1:N}$ per ogni $0 \leq s \leq n \leq N$, poiché vale

$$p(x_{s:n}|y_{1:N}) = f_{n|N}(x_n|y_{1:N}) \prod_{r=s}^{n-1} p(x_r|x_{r+1}, y_{1:N}).$$

Teorema 1.2.5 *Le densità di smoothing $f_{n|N}(\cdot|y_{1:N})$ possono essere calcolate mediante la seguente ricorsione in $n > 0$,*

$$f_{n|N}(x_n|y_{1:N}) = f_{n|n}(x_n|y_{1:n}) \int \frac{a_{n+1}(x_n, x)}{f_{n+1|n}(x|y_{1:n})} f_{n+1|N}(x|y_{1:N}) d\mu(x),$$

con base della ricorsione $f_{N|N}(x_N|y_{1:N})$.

Dimostrazione. Condizionando a $X_{n+1} = x_{n+1}$ e ricordando la (1.10), si ha

$$f_{n|N}(x_n|y_{1:N}) = p(x_n|y_{1:N}) = \int p(x_n|y_{1:N}, x_{n+1})p(x_{n+1}|y_{1:N})d\mu(x_{n+1}).$$

Per la (1.11) si ottiene

$$\begin{aligned} &= \int \frac{a_{n+1}(x_n, x_{n+1})f_{n|n}(x_n|y_{1:n})}{f_{n+1|n}(x_{n+1}|y_{1:n})} f_{n+1|N}(x_{n+1}|y_{1:N})d\mu(x_{n+1}) \\ &= f_{n|n}(x_n|y_{1:n}) \int \frac{a_{n+1}(x_n, x)}{f_{n+1|n}(x|y_{1:n})} f_{n+1|N}(x|y_{1:N})d\mu(x). \end{aligned}$$

□

Capitolo 2

Metodi per il filtraggio e lo smoothing

Le ricorsioni presentate nel capitolo precedente sono soltanto apparentemente semplici. In generale infatti non é possibile calcolare il fattore di normalizzazione $p(y_{1:n})$ e le marginali della densitá $p(x_{0:n}|y_{1:n})$ (densitá di filtraggio, smoothing e previsione). Questo problema nasce dalla difficoltá di calcolare complessi integrali di dimensione elevata.

In questo capitolo verranno considerati dei metodi che sono stati sviluppati per superare questo problema. Verranno dapprima illustrati dei casi particolari in cui sono possibili calcoli esatti e successivamente si accennerá a metodi approssimati standard, indicando alcuni loro limiti. Si dará una introduzione ai metodi Monte Carlo, dando particolare enfasi alla descrizione del piú semplice metodo Monte Carlo sequenziale.

2.1 Metodi esatti

Come é ben noto in letteratura, il problema del filtraggio e quello dello smoothing ammettono una soluzione analitica solo per pochi modelli. In particolare questo vale per modelli state space lineari e gaussiani, la cui soluzione é legata al filtro di Kalman, e per catene di Markov nascoste con un numero finito di stati.

2.1.1 Modelli markoviani nascosti

Nel caso di catene di Markov nascoste con un numero finito di stati (sia esso M), il calcolo delle ricorsioni di filtraggio e di smoothing é semplice perché gli integrali assumono la forma di sommatorie finite. Per il filtraggio, indicati con x^1, x^2, \dots, x^M i possibili valori assunti dagli stati del sistema, si ha che

$$\begin{aligned} f_{n+1|n}(x_{n+1}|y_{1:n}) &= \sum_{i=1}^M a_{n+1}(x_n^i, x_{n+1}) f_{n|n}(x_n^i|y_{1:n}), \\ f_{n+1|n+1}(x_{n+1}|y_{1:n+1}) &= \frac{b_{n+1}(y_{n+1}, x_{n+1}) f_{n+1|n}(x_{n+1}|y_{1:n})}{p(y_{n+1}|y_{1:n})}, \\ p(y_{n+1}|y_{1:n}) &= \sum_{i=1}^M b_{n+1}(y_{n+1}, x_{n+1}^i) f_{n+1|n}(x_{n+1}^i|y_{1:n}). \end{aligned}$$

Ad ogni istante temporale N , per calcolare $p(X_{n+1} = x^j|y_{1:n})$ si deve affrontare un numero di operazioni dell'ordine di $O(M)$. Poiché tale quantità deve essere calcolata per ogni $j = 1, \dots, M$, si avrà una complessità computazionale totale dell'ordine $O(NM^2)$. Data la linearità in N , questo é possibile anche per serie temporali molto lunghe. Considerazioni analoghe possono essere fatte per le ricorsioni di smoothing.

2.1.2 Filtro di Kalman

I modelli lineari gaussiani costituiscono una tipologia di modelli state space molto importante in quanto esiste un metodo analitico per risolvere direttamente le ricorsioni presentate in precedenza. In questi modelli i rumori sono gaussiani così come la distribuzione iniziale a_0 . Sotto queste ipotesi, tutte le $f_{n|s}$ sono densità gaussiane e quindi é sufficiente calcolare le medie condizionate e le matrici di covarianza condizionata di X_n dato $y_{1:s}$, che saranno indicate con $m_{n|s} = E[X_n|y_{1:s}]$ e $R_{n|s} = Cov[X_n|y_{1:s}]$, rispettivamente. Il modello state space lineare gaussiano é definito da

$$X_n = G_n X_{n-1} + V_n, \quad V_n \sim N(0, \Sigma_n) \quad (2.1)$$

$$Y_n = H_n X_n + W_n, \quad W_n \sim N(0, \Omega_n) \quad (2.2)$$

dove V_n e W_n , per ogni n , sono variabili casuali normali indipendenti, e $G_n, H_n, \Sigma_n, \Omega_n$ sono matrici note che specificano un particolare modello lineare gaussiano.

Si verifica che

$$\begin{aligned} m_{n|n-1} &= E[X_n|y_{1:n-1}] = E[G_n X_{n-1} + V_n|y_{1:n-1}] \\ &= G_n E[X_{n-1}|y_{1:n-1}] = G_n m_{n-1|n-1}, \end{aligned}$$

$$\begin{aligned} R_{n|n-1} &= Cov[X_n|y_{1:n-1}] = Cov[G_n X_{n-1} + V_n|y_{1:n-1}] \\ &= G_n Cov[X_{n-1}|y_{1:n-1}] G_n' + \Sigma_n = G_n R_{n-1|n-1} G_n' + \Sigma_n. \end{aligned}$$

Applicando le ricorsioni di filtraggio si ha che

$$\begin{aligned} R_{n|n} &= (H_n' \Omega_n^{-1} H_n + R_{n|n-1}^{-1})^{-1} \\ m_{n|n} &= m_{n|n-1} + R_{n|n} H_n' \Omega_n^{-1} (y_n - H_n m_{n|n-1}). \end{aligned}$$

Inoltre

$$\begin{aligned} R_{n|n}^{-1} R_{n|n-1} H_n' &= (H_n' \Omega_n^{-1} H_n + R_{n|n-1}^{-1}) R_{n|n-1} H_n' \\ &= H_n' \Omega_n^{-1} H_n R_{n|n-1} H_n' + H_n' \\ &= H_n' \Omega_n^{-1} (\Omega_n + H_n R_{n|n-1} H_n') \\ &= H_n' \Omega_n^{-1} M_n, \end{aligned}$$

dove $M_n = \Omega_n + H_n R_{n|n-1} H_n'$. Sostituendo $R_{n|n} = R_{n|n-1} H_n' M_n^{-1} \Omega_n H_n'$, si ha

$$\begin{aligned} m_{n|n} &= m_{n|n-1} + R_{n|n} H_n' \Omega_n^{-1} (y_n - H_n m_{n|n-1}) \\ &= m_{n|n-1} + R_{n|n-1} H_n' M_n^{-1} \Omega_n H_n' H_n' \Omega_n^{-1} (y_n - H_n m_{n|n-1}) \\ &= m_{n|n-1} + R_{n|n-1} H_n' M_n^{-1} (y_n - H_n m_{n|n-1}). \end{aligned}$$

Sostituendo $R_{n|n} H_n' \Omega_n^{-1} = R_{n|n-1} H_n' M_n^{-1}$, si ha

$$\begin{aligned} R_{n|n} &= (H_n' \Omega_n^{-1} H_n + R_{n|n-1}^{-1})^{-1} \\ R_{n|n}^{-1} &= H_n' \Omega_n^{-1} H_n + R_{n|n-1}^{-1} \\ R_{n|n}^{-1} &= (H_n' \Omega_n^{-1} H_n R_{n|n-1} + 1) R_{n|n-1}^{-1} \\ R_{n|n} &= R_{n|n-1} (H_n' \Omega_n^{-1} H_n R_{n|n-1} + 1)^{-1} \end{aligned}$$

$$\begin{aligned} R_{n|n} (H_n' \Omega_n^{-1} H_n R_{n|n-1} + 1) &= R_{n|n-1} \\ R_{n|n} H_n' \Omega_n^{-1} H_n R_{n|n-1} + R_{n|n} &= R_{n|n-1} \\ R_{n|n-1} H_n' M_n^{-1} H_n R_{n|n-1} + R_{n|n} &= R_{n|n-1} \end{aligned}$$

$$R_{n|n} = R_{n|n-1} - R_{n|n-1}H_n' M_n^{-1} H_n R_{n|n-1}.$$

Riassumendo,

$$M_n = \Omega_n + H_n R_{n|n-1} H_n'$$

$$R_{n|n} = R_{n|n-1} - R_{n|n-1} H_n' M_n^{-1} H_n R_{n|n-1}$$

$$m_{n|n} = m_{n|n-1} + R_{n|n-1} H_n' M_n^{-1} (y_n - H_n m_{n|n-1}). \quad (2.3)$$

Tali equazioni costituiscono il filtro di Kalman. Si noti come la (2.3) abbia una forma intuitiva: la media della densità di filtraggio é uguale alla media di previsione piú un elemento di correzione che dipende da quanto la nuova osservazione differisce dalla previsione. Analogamente, utilizzando la (1.9), si può dimostrare che la distribuzione condizionata di X_n dati x_{n+1} e $y_{1:N}$ é gaussiana con vettore delle medie \bar{m}_n e matrice di covarianza \bar{R}_n , dove

$$\begin{aligned} \bar{m}_n &= m_{n|n} + \bar{R}_n^{-1} G_{n+1}' \Sigma_{n+1}^{-1} (x_{n+1} - m_{n+1|n+1}) \\ &= m_{n|n} + R_{n|n} G_{n+1}' R_{n+1|n}^{-1} (x_{n+1} - m_{n+1|n+1}), \\ \bar{R}_n &= (G_{n+1}' \Sigma_{n+1}^{-1} G_{n+1} + R_{n+1|n}^{-1})^{-1} \\ &= R_{n|n} - R_{n|n} G_{n+1}' R_{n+1|n}^{-1} G_{n+1} + R_{n+1|n}. \end{aligned}$$

Da qui si ottiene il *Kalman smoother*:

$$\begin{aligned} m_{n|N} &= m_{n|n} + S_n (m_{n+1|N} - m_{n+1|n}), \\ R_{n|N} &= R_{n|n} - S_n (R_{n+1|n} - R_{n+1|N}) S_n'. \end{aligned}$$

dove

$$S_n = R_{n|n} G_{n+1}' R_{n+1|n}^{-1}$$

Nel Capitolo 3 vedremo un altro esempio in cui é possibile definire delle ricorsioni di dimensione finita per il calcolo delle distribuzioni di filtraggio. Si tratterá di trovare un modello e una famiglia \mathcal{M} di densità sullo spazio degli stati tale che ogni elemento di \mathcal{M} sia descritto da un numero finito di parametri e tale che, per ogni $y_{1:n+1}$, $f_{n|n}(\cdot | y_{1:n}) \in \mathcal{M}$ implichi $f_{n+1|n+1}(\cdot | y_{1:n+1}) \in \mathcal{M}$. In questa situazione sará sufficiente aggiornare i parametri ad ogni passo.

I casi in cui sono possibili calcoli esatti restano tuttavia molto limitati.

2.2 Metodi approssimati

In molte situazioni di interesse, le condizioni su cui si basano i metodi esatti risultano troppo restrittive. Molto spesso i dati reali sono non gaussiani e i modelli non lineari. Il filtro di Kalman non può essere utilizzato e risulta quindi necessario ricorrere a metodi approssimati. In questa sezione verranno considerati il filtro di Kalman esteso e i metodi Monte Carlo.

2.2.1 Filtro di Kalman esteso (FKE)

Nel caso in cui il modello descritto dalle (1.1), (1.2) non possa essere riscritto nella forma (2.1), (2.2) perché le funzioni g_n e h_n non sono lineari, allora una descrizione sufficiente della non linearità è data da una approssimazione locale. Il filtro di Kalman esteso si basa su questa approssimazione. Se le funzioni g_n e h_n sono note e sufficientemente regolari, il metodo consiste nell'applicare il filtro di Kalman standard alle versioni linearizzate di (1.1), (1.2):

$$\begin{aligned} X_n &= g_n(m_{n-1|n-1}, 0) + \frac{\partial}{\partial x} g_n(m_{n-1|n-1}, 0)(X_{n-1} - m_{n-1|n-1}) \\ &\quad + \frac{\partial}{\partial v} g_n(m_{n-1|n-1}, 0)V_n \\ Y_n &= h_n(m_{n|n-1}, 0) + \frac{\partial}{\partial x} h_n(m_{n|n-1}, 0)(X_n - m_{n|n-1}) \\ &\quad + \frac{\partial}{\partial v} h_n(m_{n|n-1}, 0)W_n. \end{aligned}$$

Il FKE non fornisce informazioni sulla forma delle funzioni di densità di previsione e di filtraggio, generalmente si assume che siano gaussiane. Tuttavia, se la vera densità è non gaussiana, l'errore di approssimazione può risultare molto grande (come è stato recentemente provato empiricamente).

2.2.2 Altri metodi approssimati

Dato che i casi in cui le ricorsioni di filtraggio e di smoothing possono essere calcolate in modo esatto sono rari e dato che il FKE non fornisce informazioni circa la forma delle funzioni di densità condizionata, risulta necessario ricorrere ad altri metodi approssimati.

Kitagawa (1987) ha sviluppato delle approssimazioni numeriche. Tuttavia l'integrazione numerica è difficile se le dimensioni sono elevate. Inoltre, per

una buona scelta dei nodi di integrazione sarebbe necessario conoscere in anticipo la forma delle funzioni di densità di filtraggio.

Un approccio per modelli lineari con rumori non gaussiani é quello in cui sono stati utilizzati modelli basati su misture di gaussiane. Si ha quindi un filtro di Kalman lineare per ogni distribuzione componente la mistura. Poiché il numero di componenti in $f_{n|n}$ cresce esponenzialmente con n , é necessario ridurre questo numero mediante approssimazioni. Di solito si sostituisce una mistura di distribuzioni normali con una singola distribuzione normale con momento primo e secondo uguali a quello della mistura. Gli algoritmi sviluppati seguendo questo approccio sono molteplici, tuttavia la restrizione a modelli state space lineari resta rigida. Per questo motivo, l'attenzione é stata posta sui metodi Monte Carlo.

2.3 Metodi Monte Carlo

In questo paragrafo verranno presentati dei metodi per calcolare distribuzioni a posteriori basati su simulazioni Monte Carlo. Si osserverá che, se si possiedono un gran numero di valori campionati dalle distribuzioni di interesse, non é difficile ottenere una approssimazione degli integrali presenti nelle ricorsioni presentate nel Capitolo 1. Purtroppo non sempre é possibile campionare direttamente dalle distribuzioni di interesse. Per questo bisognerà ricorrere a metodi Monte Carlo alternativi, come il campionamento per importanza. Rendendo ricorsiva questa tecnica, si ottiene il metodo di campionamento sequenziale per importanza. Tuttavia, al crescere di n , questo metodo fornisce risultati poco affidabili: sará necessario aggiungere un ulteriore passo di selezione.

2.3.1 Campionamento Monte Carlo perfetto

Supponiamo sia possibile generare M valori simulati $x_{0:n}^{(1)}, x_{0:n}^{(2)}, \dots, x_{0:n}^{(M)}$ indipendenti e identicamente distribuiti, chiamati anche *particles*, secondo la distribuzione condizionata di $X_{0:n}$ dato $y_{1:n}$. Una stima della distribuzione di probabilità a posteriori di $X_{0:n}$ dato $y_{1:n}$ é data dalla distribuzione di probabilità empirica $P_M(\cdot, y_{1:n})$ definita dalla

$$P_M(dx_{0:n}, y_{1:n}) = \frac{1}{M} \sum_{i=1}^M \delta_{x_{0:n}^{(i)}}(dx_{0:n}),$$

dove $\delta_{x_{0:n}}^{(i)}(\cdot)$ denota la distribuzione di probabilità degenerare in $x_{0:n}^{(i)}$. Per ogni funzione φ_n integrabile rispetto alla distribuzione condizionata di $X_{0:n}$ dato $y_{1:n}$, ponendo

$$I(\varphi_n) = E[\varphi_n(X_{0:n}) | Y_{1:n} = y_{1:n}]$$

una stima per $I(\varphi_n)$ é data da

$$I_M(\varphi_n) = \int \varphi_n(x_{0:n}) P_M(dx_{0:n} | y_{1:n}) = \frac{1}{M} \sum_{i=1}^M \varphi_n(x_{0:n}^{(i)})$$

Questa stima é non distorta e, indicando con $\sigma_{\varphi_n}^2 = V[\varphi_n(X_{0:n}) | Y_{1:n} = y_{1:n}] = E[\varphi_n(X_{0:n})^2 | Y_{1:n} = y_{1:n}] - I^2(\varphi_n)$ la matrice di covarianza a posteriori di $\varphi_n(X_{0:n})$, se $\sigma_{\varphi_n}^2 < +\infty$, allora la varianza di $I_M(\varphi_n) = \sigma_{\varphi_n}^2/M$. Per la legge forte dei grandi numeri

$$I_M(\varphi_n) \xrightarrow{q.c.} I(\varphi_n) \quad M \rightarrow +\infty$$

dove $\xrightarrow{q.c.}$ indica la convergenza quasi certa. Inoltre

$$\sqrt{M}[I_M(\varphi_n) - I(\varphi_n)] \xrightarrow{d} \mathcal{N}(0, \sigma_{\varphi_n}^2) \quad M \rightarrow +\infty$$

dove \xrightarrow{d} denota la convergenza in distribuzione. L'utilizzo di questo metodo Monte Carlo perfetto offre importanti vantaggi.

Dal campione casuale $(X_{0:n}^{(1)}, \dots, X_{0:n}^{(M)})$ si puó stimare facilmente qualsiasi quantità $I(\varphi_n)$ e la velocità di convergenza al vero valore di questo stimatore non dipende dalla dimensione del dominio di integrazione.

Sfortunatamente, di solito non é possibile campionare in modo efficiente dalla densità a posteriori $p(x_{0:n} | y_{1:n})$ ad ogni istante n , in quanto essa é la densità di una variabile casuale multivariata, non corrisponde usualmente a modelli standard ed é nota solo a meno di un fattore di proporzionalità. In statistica applicata i metodi Markov Chain Monte Carlo (MCMC) sono un approccio frequente per campionare da distribuzioni di probabilità complesse. Nel seguente paragrafo daremo una introduzione a tali metodi.

2.3.2 Algoritmi Markov Chain Monte Carlo

Gli algoritmi di tipo Markov Chain Monte Carlo (MCMC) forniscono uno strumento molto potente per calcolare integrali. L'obiettivo é di stimare in

maniera Monte Carlo l'integrale

$$E_{\pi_n}[\varphi_n(X)] = \int \varphi_n(x)\pi_n(x)dx \approx \frac{1}{M} \sum_{n=1}^M \varphi_n(x^{(i)})$$

dove π_n rappresenta la distribuzione condizionata di $X_{0:n}$ dato $y_{1:n}$ e $x^{(i)}$ é la realizzazione di una variabile casuale $X \sim \pi_n$.

Lo scopo di MCMC é quello di costruire una catena ergodica (e reversibile) rispetto alla legge π_n . In pratica si genera una catena di Markov da una distribuzione a priori, si scartano le prime osservazioni (che risentono della distribuzione iniziale) e si considerano le successive come realizzazioni identicamente distribuite secondo la distribuzione stazionaria della catena π_n , che é la distribuzione di interesse.

La condizione di reversibilit a rispetto alla distribuzione $\pi_n(\cdot)$ ci dice che $\pi_n(x)q(y|x) = \pi_n(y)q(x|y)$ dove q é la densit a di transizione.

Se una catena é reversibile rispetto a π_n , allora quest'ultima é stazionaria. Infatti:

$$\sum_x \pi_n(x)q(y|x) = \sum_x \pi_n(y)q(x|y) = \pi_n(y) \underbrace{\sum_x q(x|y)}_1 = \pi_n(y)$$

Costruiamo un algoritmo che utilizzi la condizione di reversibilit a (che implica la stazionariet a) per generare una catena che converga a π_n .

Supponiamo che $q(y|x)$ sia la densit a di transizione della catena di Markov. In generale sar a:

$$\pi_n(x)q(y|x) > \pi_n(y)q(x|y)$$

oppure

$$\pi_n(x)q(y|x) < \pi_n(y)q(x|y)$$

Per bilanciare le due situazioni costruisco una funzione di (x, y) per cui valga:

$$\pi_n(x)q(y|x)\alpha(x, y) = \pi_n(y)q(x|y)\alpha(y, x)$$

L'algoritmo Metropolis Hastings (1970) é l'algoritmo pi u generale. Esso consiste nel generare un valore candidato y da una **distribuzione proposta** q e poi accettarlo con probabilit a $\alpha \in [0, 1]$ detta **probabilit a di accettazione**

$$\alpha(x, y) = \begin{cases} \frac{\pi_n(y)q(x|y)}{\pi_n(x)q(y|x)} & \text{se } \pi_n(x)q(y|x) \neq 0 \\ 1 & \text{altrimenti} \end{cases}$$

Le fasi dell'algoritmo sono le seguenti:

1. si sceglie il valore iniziale $x^{(0)}$ e si pone $n = 0$
2. si simula un valore y da $q(\cdot|x^{(n)})$, si simula, indipendentemente dalla simulazione di y , un valore u da $U(0, 1)$
3. se $u \leq \alpha(x_n, y)$ allora $x^{(n+1)} = y$. In questo caso il valore viene accettato, altrimenti $x^{(n+1)} = x^{(n)}$ (la catena non si muove)
4. incremento n e torno al punto 2

Con una scelta opportuna della distribuzione proposta q , la catena é ergodica e valgono i teoremi ergodici. La distribuzione q puó essere scelta facile da simulare e la distribuzione di interesse π_n si puó conoscere anche a meno del fattore di normalizzazione. Infatti essa interviene in $\alpha(x, y)$ tramite il rapporto $\frac{\pi_n(y)}{\pi_n(x)}$. Casi particolari dell'algorithmo Metropolis Hastings si ottengono variando la scelta della distribuzione proposta $q(y|x)$.

ESEMPI:

- Algorithmo Metropolis $q(x|y) = q(y|x)$
- Algorithmo Random-Walk Metropolis $q(y|x) \propto h(|x - y|)$
- Algorithmo Independence Sampler $q(y|x) = q(y)$
- Algorithmo Metropolis-Hastings a blocchi (utilizzato quando la dimensione dei parametri é molto elevata). Invece di aggiornare l'intero vettore x , é spesso piú conveniente dividere x in componenti di dimensioni minori e aggiornarle una alla volta. Caso particolare di questo metodo é l' **Algorithmo Gibbs sampler**. La distribuzione proposta per aggiornare l' i -esima componente di x é:

$$q(x^{(i)}|x^{(-i)}) \propto \pi_n(x^{(i)}|x^{(-i)})$$

In questo caso la probabilitá di accettazione é pari a 1, cioé i candidati generati dall'algorithmo vengono sempre accettati.

Sebbene questi algoritmi permettano di campionare in modo semplice anche da distribuzioni complicate, i metodi MCMC non sono adatti a problemi di stima ricorsiva. Per questo motivo sono stati sviluppati metodi alternativi.

2.3.3 Campionamento per importanza

Campionamento per importanza

Una soluzione alternativa classica consiste nell'utilizzare il metodo di *campionamento per importanza*. Si introduce una funzione di densità arbitraria, detta *densità proposta* $q(x_{0:n}|y_{1:n})$, dalla quale sia possibile campionare facilmente. Si supponga di voler stimare $I(\varphi_n)$. Purché il supporto di $p(x_{0:n}|y_{1:n})$ sia contenuto nel supporto di $q(x_{0:n}|y_{1:n})$, vale

$$I(\varphi_n) = \frac{\int \varphi_n(x_{0:n})\omega^*(x_{0:n})q(x_{0:n}|y_{1:n})d\mu_{n+1}(x_{0:n})}{\int \omega^*(x_{0:n})q(x_{0:n}|y_{1:n})d\mu_{n+1}(x_{0:n})}, \quad (2.4)$$

dove $\omega^*(x_{0:n})$, detto *peso d'importanza*, é dato da

$$\omega^*(x_{0:n}) = \frac{p(x_{0:n}|y_{1:n})}{q(x_{0:n}|y_{1:n})}.$$

Quindi, se é possibile simulare M particles indipendenti e identicamente distribuiti $x_{0:n}^{(1)}, x_{0:n}^{(2)}, \dots, x_{0:n}^{(M)}$ secondo la densità proposta, una possibile stima Monte Carlo per $I(\varphi_n)$ é

$$\hat{I}_M(\varphi_n) = \frac{\frac{1}{M} \sum_{i=1}^M \varphi_n(x_{0:n}^{(i)})\omega^*(x_{0:n}^{(i)})}{\frac{1}{M} \sum_{j=1}^M \omega^*(x_{0:n}^{(j)})} = \sum_{i=1}^M \varphi_n(x_{0:n}^{(i)})\tilde{\omega}_n^{(i)}, \quad (2.5)$$

dove i pesi d'importanza normalizzati $\tilde{\omega}_n^{(i)}$ sono dati dalla

$$\tilde{\omega}_n^{(i)} = \frac{\omega^*(x_{0:n}^{(i)})}{\sum_{j=1}^M \omega^*(x_{0:n}^{(j)})}. \quad (2.6)$$

Si noti che

$$\omega^*(x_{0:n}^{(i)}) = \frac{p(x_{0:n}^{(i)}|y_{1:n})}{q(x_{0:n}^{(i)}|y_{1:n})} = \frac{p(y_{1:n}|x_{0:n}^{(i)})p(x_{0:n}^{(i)})}{p(y_{1:n})q(x_{0:n}^{(i)}|y_{1:n})}$$

e che generalmente non é possibile calcolare la costante di normalizzazione $p(y_{1:n})$. Per questo motivo, in realtà la funzione $\omega^*(x_{0:n})$ viene sostituita da

$$\omega(x_{0:n}) = \frac{p(y_{1:n}|x_{0:n})p(x_{0:n})}{q(x_{0:n}^{(i)}|y_{1:n})} \propto \omega^*(x_{0:n}).$$

Questa sostituzione non modifica i valori delle espressioni (2.4)-(2.6), in quanto la costante di proporzionalità compare sia al numeratore che al denominatore.

Per M finito, $\hat{I}_M(\varphi_n)$ é uno stimatore distorto ma consistente; infatti si può applicare la legge forte dei grandi numeri, cioè $\hat{I}_M(\varphi_n) \xrightarrow{q.c.} I(\varphi_n)$ per $M \rightarrow +\infty$. Inoltre si può ottenere un teorema limite centrale con velocità di convergenza indipendente dal dominio di integrazione.

Il campionamento per importanza é un metodo di integrazione Monte Carlo generale; tuttavia nella sua forma piú semplice non é adatto a problemi di stima ricorsiva. Infatti é necessario ottenere tutti i dati $y_{1:n}$ prima di stimare $p(x_{0:n}|y_{1:n})$ e, in generale, ogni volta che una nuova osservazione y_{n+1} diventa disponibile, occorre ricalcolare i pesi di importanza su tutta la sequenza di stati. La complessità di questa operazione é crescente al crescere di n . Nel prossimo paragrafo verrà proposta una strategia per superare questo problema.

Campionamento sequenziale per importanza

Il metodo di campionamento per importanza può essere modificato in modo da rendere possibile il calcolo di una stima $P_M(dx_{0:n}|y_{1:n})$ della distribuzione di $X_{0:n}|Y_{1:n}$ senza modificare le traiettorie simulate passate $(x_{0:n-1}^{(1)}, \dots, x_{0:n-1}^{(M)})$. Perché ciò sia possibile, la densità proposta $q(x_{0:n}|y_{1:n})$ all'istante n deve essere scelta in modo tale da ammettere come densità marginale la densità proposta all'istante $n-1$, $q(x_{0:n-1}|y_{1:n-1})$, cioè

$$q(x_{0:n}|y_{1:n}) = q(x_{0:n-1}|y_{1:n-1})q(x_n|x_{0:n-1}, y_{1:n}).$$

Iterando

$$q(x_{0:n}|y_{1:n}) = q(x_0) \prod_{j=1}^n q(x_j|x_{0:j-1}|y_{1:j-1}). \quad (2.7)$$

Una densità proposta che soddisfi a questa proprietà permette di calcolare ricorsivamente i pesi di importanza (2.6). Si ha che

$$\tilde{\omega}_n \propto \omega^*(x_{0:n}) = \frac{p(x_{0:n}|y_{1:n})}{q(x_{0:n}|y_{1:n})}$$

Sostituendo la definizione della densità proposta

$$= \frac{p(x_{0:n}|y_{1:n})}{q(x_{0:n-1}|y_{1:n-1})q(x_n|x_{0:n-1}, y_{1:n})}$$

Utilizzando il teorema di Bayes e il fatto che Y_n e $(X_{0:n-1}, Y_{1:n-1})$ sono indipendenti condizionatamente a $X_n = x_n$ e che X_n e $(X_{0:n-2}, Y_{1:n-1})$ sono

indipendenti condizionatamente a $X_{n-1} = x_{n-1}$, si ha che

$$= \frac{\frac{p(y_n|x_n)p(x_n|x_{n-1})p(x_{0:n-1}|y_{1:n-1})}{p(y_n|y_{1:n-1})}}{q(x_{0:n-1}|y_{1:n-1})q(x_n|x_{0:n-1}, y_{1:n})}.$$

Ora

$$\frac{p(x_{0:n-1}|y_{1:n-1})}{q(x_{0:n-1}|y_{1:n-1})} = \omega^*(x_{0:n-1}) \propto \tilde{\omega}_{n-1}.$$

Quindi

$$\tilde{\omega}_n \propto \tilde{\omega}_{n-1} \frac{p(y_n|x_n)p(x_n|x_{n-1})}{q(x_n|x_{0:n-1}, y_{1:n})}.$$

Questo metodo viene chiamato *campionamento sequenziale per importanza* (SIS). Un caso particolare molto importante si ottiene adottando come densità proposta la densità di $X_{0:n}$, detta **densità a priori**:

$$q(x_{0:n}|y_{1:n}) = p(x_{0:n}) = a_0(x_0) \prod_{j=1}^n a_j(x_{j-1}, x_j).$$

In questo caso i pesi di importanza diventano

$$\tilde{\omega}_n \propto \tilde{\omega}_{n-1} \frac{b_n(x_n, y_n)a_n(x_{n-1}, x_n)}{a_n(x_{n-1}, x_n)} = \tilde{\omega}_{n-1} b_n(x_n, y_n).$$

Anche se il metodo é piú generale, nel seguito utilizzeremo come densità proposta la densità a priori. Sfortunatamente, al crescere di n il metodo SIS é solitamente inefficiente. Per questo nella prossima sezione vedremo un metodo che permette di risolvere questo problema.

2.3.4 Filtro bootstrap

Il metodo SIS presenta uno spiacevole problema al crescere di n : la distribuzione dei pesi di importanza $\tilde{\omega}_n^{(i)}$ diventa sempre piú asimmetrica. In pratica, dopo alcuni passi, soltanto una particella ha un peso di importanza non nullo. L'algoritmo non riesce quindi a rappresentare adeguatamente le distribuzioni a posteriori di interesse. Per evitare questo fenomeno degenerativo occorre aggiungere un ulteriore passo di selezione.

L'idea di base del filtro bootstrap consiste nell'eliminare i particles che hanno pesi di importanza bassi e moltiplicare quelli con pesi elevati. Si

sostituisce cioè alla distribuzione empirica pesata

$$\hat{P}_M(dx_{0:n}|y_{1:n}) = \sum_{j=1}^M \tilde{\omega}_n^{(j)} \delta_{x_{0:n}^{(j)}}(dx_{0:n})$$

la distribuzione non pesata

$$P_M(dx_{0:n}|y_{1:n}) = \frac{1}{M} \sum_{j=1}^M \xi_n^{(j)} \delta_{x_{0:n}^{(j)}}(dx_{0:n}),$$

dove $\xi_n^{(j)}$ rappresenta il numero di discendenti associato al particle $x_{0:n}^{(j)}$. Si tratta di un numero intero tale che $\sum_{j=1}^M \xi_n^{(j)} = M$. Se $\xi_n^{(j)} = 0$, allora il particle $x_{0:n}^{(j)}$ viene eliminato. Gli $\xi_n^{(j)}$ sono scelti in modo che $P_M(dx_{0:n}|y_{1:n})$ approssimi $\hat{P}_M(dx_{0:n}|y_{1:n})$. Dopo il passo di selezione, i particles sopravvissuti, cioè quelli con $\xi_n^{(j)} > 0$, sono distribuiti approssimativamente secondo la distribuzione di $X_{0:n}|Y_{1:n} = y_{1:n}$. Esistono diversi meccanismi per la scelta di $\xi_n^{(j)}$. Il piú diffuso consiste nel campionare il numero di discendenti secondo una distribuzione multinomiale di parametri $M, \tilde{\omega}_n^{(1)}, \dots, \tilde{\omega}_n^{(M)}$. Questo schema produce come risultato aggiuntivo il fatto che dopo il passo di ricampionamento i particles propagati hanno pesi uniformi.

L'algoritmo puó essere specificato nel seguente modo:

1. Inizializzazione, $n = 0$

- Per $i = 1, \dots, M$ genero i particles $x_0^{(i)} \sim a_0(x)d\mu(x)$;
- pongo $n = 1$.

2. Passo di campionamento per importanza

- Per $i = 1, \dots, M$ genero i particles $\tilde{x}_n^{(i)} \sim a_n(x_{n-1}^{(i)}, x)d\mu(x)$ e pongo $\tilde{x}_{0:n}^{(i)} = (x_{0:n-1}^{(i)}, \tilde{x}_n^{(i)})$;
- per $i = 1, \dots, M$ calcolo i pesi di importanza

$$\tilde{\omega}_n^{(i)} = b_n(\tilde{x}_n^{(i)}, y_n);$$

- normalizzo i pesi di importanza.

3. Passo di selezione

- campiono con ripetizione M particles $(x_{0:n}^{(1)}, \dots, x_{0:n}^{(M)})$ dall'insieme $\{\tilde{x}_{0:n}^{(1)}, \dots, \tilde{x}_{0:n}^{(M)}\}$ secondo i pesi di importanza;
- pongo $n := n + 1$ e torno al passo 2.

L'algoritmo descritto in questo paragrafo é uno dei piú essenziali particle filter. Questo schema é applicabile ad una classe molto vasta di modelli ed é facile da implementare. Tuttavia, dal punto di vista dell'efficienza computazionale, questo algoritmo puó essere migliorato. In questa direzione si sono mossi numerosi sviluppi algoritmici che hanno permesso di migliorare significativamente le prestazioni rispetto ai metodi standard.

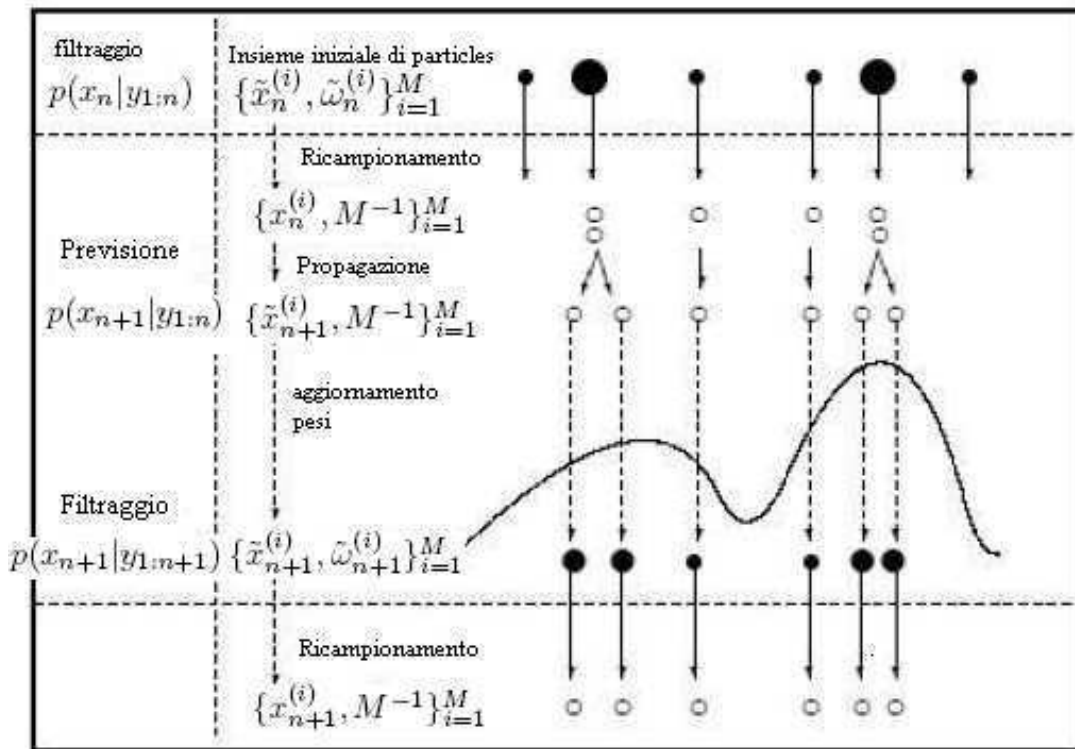


Figura 2.1: Schema del filtro bootstrap

In questa figura si considera il filtro bootstrap a partire dall'istante n con una distribuzione pesata, individuata da $\{\tilde{x}_n^{(i)}, \tilde{\omega}_n^{(i)}\}_{i=1}^M$, che fornisce una approssimazione per la distribuzione di $X_n|Y_{1:n} = y_{1:n}$. In seguito, nel passo di ricampionamento, vengono selezionati soltanto i particles piú adatti per ottenere la distribuzione non pesata, individuata da $\{\tilde{x}_n^{(i)}, M^{-1}\}_{i=1}^M$, che é ancora una approssimazione per la distribuzione di $X_n|Y_{1:n} = y_{1:n}$. Infine il passo di previsione introduce varietà, producendo la distribuzione, individuata da $\{\tilde{x}_{n+1}^{(i)}, M^{-1}\}_{i=1}^M$, che é una approssimazione per la distribuzione di $X_{n+1}|Y_{1:n} = y_{1:n}$. Per ogni particle si calcolano i pesi di importanza usando l'informazione al tempo $n + 1$. Si ottiene nuovamente una distribuzione pesata individuata da $\{\tilde{x}_{n+1}^{(i)}, \tilde{\omega}_{n+1}^{(i)}\}_{i=1}^M$, che fornisce una approssimazione per la distribuzione di $X_{n+1}|Y_{1:n+1} = y_{1:n+1}$.

Capitolo 3

Modello per media e varianza stocastica

In questo capitolo presenteremo un problema di filtraggio per un processo stocastico parzialmente osservabile $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$, che ci fornisce una modellizzazione per la media e la varianza di osservazioni normalmente distribuite. I rumori nelle equazioni sono dati da due sequenze di variabili casuali gaussiane indipendenti e da una sequenza di variabili casuali indipendenti con distribuzione Gamma. Nel Paragrafo 3.2 si proverá l'esistenza di un filtro finito dimensionale per questo modello per media e varianza stocastica, seguendo il lavoro di Ferrante, Vidoni (1999).

3.1 Filtro finito-dimensionale

Ricordiamo che il problema di filtraggio a tempo discreto può essere descritto nel seguente modo. Consideriamo un processo stocastico parzialmente osservabile $\{X_n, Y_n\}_{n \in \mathbb{N}}$, con $\{X_n\}_{n \in \mathbb{N}}$ processo stocastico non osservabile e $\{Y_n\}_{n \in \mathbb{N}}$ processo stocastico osservabile. Il processo $\{X_n, Y_n\}_{n \in \mathbb{N}}$ é usualmente definito come la soluzione di un sistema di equazioni stocastiche alle differenze:

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) + g(X_n, Y_n)W_{n+1}, & n \in \mathbb{N}, \\ Y_n &= h(X_n) + \rho(X_n)V_n, & n \in \mathbb{N} \end{aligned}$$

dove $\{W_{n+1}\}_{n \in \mathbb{N}}$ e $\{V_n\}_{n \in \mathbb{N}}$ sono sequenze di variabili casuali indipendenti, f, g, h, ρ sono funzioni reali e X_0 é una variabile casuale nota. Il problema di filtraggio consiste nel calcolare, per ogni $n \in \mathbb{N}$, la distribuzione condizionata di X_n dato $Y_{0:n} = (Y_0, \dots, Y_n)$, detta *distribuzione di filtraggio*. Chiamiamo *sistema di filtro esatto* l'algoritmo che determina queste distribuzioni condizionate. Ricordiamo che in generale un sistema di filtro esatto non é molto utile in pratica, dato che il numero delle quantità coinvolte nel calcolo delle distribuzioni di interesse aumenta con n , rendendo la derivazione analitica pressoché impossibile. Tuttavia, se siamo in grado di provare che queste distribuzioni condizionate appartengono ad una data famiglia di distribuzioni di probabilità parametrizzata da un insieme finito di parametri, il problema diventa molto piú trattabile. In questo caso diciamo che il processo stocastico parzialmente osservabile $\{X_n, Y_n\}_{n \in \mathbb{N}}$ ammette un *filtro finito dimensionale* (f.f.d. in seguito).

Definizione 3.1.1 Sia $\{P(\cdot; s) : s \in S\}$ un insieme di misure di probabilità su \mathbb{R}^d con insieme parametrico $S \subseteq \mathbb{R}^k$. Un sistema di funzioni misurabili $c_n : S \times \mathbb{R}^m \rightarrow S$ é un f.f.d. per $\{X_n, Y_n\}_{n \in \mathbb{N}}$ se, per ogni $n \in \mathbb{N}$ e $y_{0:n} = (y_0, \dots, y_n) \in \mathbb{R}^{p(n+1)}$,

$$P_{n-1}(\cdot | Y_{0:n-1} = y_{0:n-1}) = P(\cdot; s) \Rightarrow P_n(\cdot | Y_{0:n} = y_{0:n}) = P(\cdot; c_n(s, y_n)),$$

dove $P_i(\cdot | Y_{0:i} = y_{0:i})$ denota la distribuzione di X_i dato $Y_{0:i} = y_{0:i}$, per $i \in \mathbb{N}$.

Sebbene l'esistenza di un f.f.d. sia un evento 'raro', i pochi esempi conosciuti vengono spesso usati nelle applicazioni e, per questa ragione, sono spesso di estremo interesse. Ad esempio, Ferrante e Vidoni (1998) e Vidoni (1999) hanno presentato alcuni nuovi f.f.d. a tempo discreto, coinvolgendo le distribuzioni Gamma, Beta, Dirichlet e Pareto.

Nel prossimo paragrafo verrà presentata una nuova classe di f.f.d. definita all'interno di un processo stocastico parzialmente osservabile dove sia la media che la varianza del processo osservabile sono supposte essere stocastiche. Piú precisamente, consideriamo il sistema non lineare

$$\begin{aligned} (X_0, Z_0) & \quad \text{vettore casuale dato} \\ X_{n+1} & = f_{n+1}^1(Z_n, Z_{n+1}, Y_{0:n})X_n + f_{n+1}^2(Z_n, Z_{n+1}, Y_{0:n}) + \end{aligned}$$

$$\begin{aligned}
& + f_{n+1}^3(Z_n, Z_{n+1})V_{n+1}, \quad n \geq 0, \\
Z_{n+1} & = g_{n+1}(Z_n) + W_{n+1}, \quad n \geq 0, \\
Y_n & = h_n^1(X_n, Z_n) + h_n^2(Z_n)V'_n, \quad n \geq 0,
\end{aligned} \tag{3.1}$$

con f_n^i , $i = 1, 2, 3$, g_n e h_n^j , $j = 1, 2$ funzioni misurabili e $\{V_{n+1}\}_{n \in \mathbb{N}}$, $\{V'_n\}_{n \in \mathbb{N}}$, $\{W_{n+1}\}_{n \in \mathbb{N}}$ sequenze di variabili casuali indipendenti.

Il processo tridimensionale $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$, soluzione del sistema (3.1), dove $\{X_n, Z_n\}_{n \in \mathbb{N}}$ rappresenta la parte non osservabile mentre $\{Y_n\}_{n \in \mathbb{N}}$ la parte osservabile, ci fornisce la modellizzazione della media e della varianza della distribuzione gaussiana delle osservazioni, cioè della distribuzione condizionata di Y_n dato $\{X_n, Z_n\}_{n \in \mathbb{N}}$.

Sotto opportune assunzioni sui coefficienti e sui termini del rumore, proveremo che la distribuzione di filtraggio, cioè la distribuzione di $\{X_n, Z_n\}$ dato $Y_{0:n} = (Y_0, \dots, Y_n)$ é una bivariata gaussiana-gaussiana inversa generalizzata¹ e che le funzioni che descrivono l'evoluzione dei parametri coinvolgono un f.f.d. Nell'Appendice A, in A.3, sono riportate la definizione e alcune proprietà della distribuzione gaussiana inversa generalizzata.

3.2 Modello

In questo paragrafo verranno presentati due processi stocastici parzialmente osservabili, soluzioni del sistema di equazioni stocastiche alle differenze non lineari del tipo (3.1). Questi modelli sono annidati, nel senso che il piú semplice funziona se viene considerato un insieme di parametri iniziali ristretto, mentre quello piú generale é, in qualche modo, equivalente al primo dopo un opportuno numero finito di passi. In entrambi i casi proveremo l'esistenza di un filtro finito dimensionale.

¹Sia $(X, Z) \sim N(\mu + \beta Z, Z) \times GIG(\lambda, \delta, \gamma)$, cioè la densità del vettore casuale (X, Z) é dato dal prodotto della densità condizionata di X dato Z (normale), e della densità di una variabile casuale con densità gaussiana inversa generalizzata. Chiamiamo la distribuzione di (X, Z) bivariata gaussiana-gaussiana inversa generalizzata.

3.2.1 Caso $\lambda \in [0, \frac{1}{2})$

Consideriamo il seguente sistema di equazioni stocastiche alle differenze non lineari:

$$\begin{aligned}
(X_0, Z_0) & \quad \text{vettore casuale dato} \\
X_{n+1} & = X_n + \beta \left(Z_{n+1} - \frac{Z_n}{2} \right) + \sqrt{Z_{n+1} - \frac{Z_n}{2}} V_{n+1}, \\
Z_{n+1} & = Z_n + W_{n+1}, \\
Y_n & = X_n + \sqrt{Z_n} V'_n, \quad n \in \mathbb{N}
\end{aligned} \tag{3.2}$$

con $\beta \in \mathbb{R}$ e $\{V_{n+1}\}_{n \in \mathbb{N}}$, $\{V'_n\}_{n \in \mathbb{N}}$ e $\{W_{n+1}\}_{n \in \mathbb{N}}$ tre sequenze indipendenti di variabili casuali indipendenti. Assumiamo che:

H.1 $\{V_{n+1}\}_{n \in \mathbb{N}}$ e $\{V'_n\}_{n \in \mathbb{N}}$ sono sequenze di variabili casuali gaussiane standard indipendenti

H.2 $\{W_{n+1}\}_{n \in \mathbb{N}}$ é una sequenza di variabili gamma indipendenti con parametro di forma $|\lambda_n|$ e parametro di scala $\gamma_n^2/2$, dove

$$\begin{aligned}
\lambda_0 & = \lambda - \frac{1}{2}, \quad \lambda_n = |\lambda_{n-1}| - \frac{1}{2}, \quad n \geq 1, \\
\gamma_0 & = \sqrt{\gamma^2 + \frac{\beta^2}{2}}, \quad \gamma_n = \sqrt{\gamma_{n-1}^2 + \frac{\beta^2}{2}}, \quad n \geq 1
\end{aligned} \tag{3.3}$$

Ricordiamo che la variabile casuale gamma W_{n+1} può essere vista come una variabile casuale $GIG(|\lambda_n|, 0, \gamma_n)$. Se $\lambda_n = 0$, definiamo $W_{n+1} \equiv 0$.

H.3 $(X_0, Z_0) \sim N(\mu + \beta Z_0, Z_0) \times GIG(\lambda, \delta, \gamma)$ con $\lambda \in [0, \frac{1}{2})$; (X_0, Z_0) indipendente dai processi $\{V_{n+1}\}_{n \in \mathbb{N}}$, $\{V'_n\}_{n \in \mathbb{N}}$ e $\{W_{n+1}\}_{n \in \mathbb{N}}$.

Sotto queste assunzioni, il processo stocastico parzialmente osservabile $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$ soluzione del sistema (3.2), con $\{X_n, Z_n\}_{n \in \mathbb{N}}$ parte non osservabile e $\{Y_n\}_{n \in \mathbb{N}}$ parte osservabile, ammette un filtro finito dimensionale.

Proposizione 3.2.1 *Sia $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$ l'unica soluzione del sistema (3.2) sotto le assunzioni (H.1)-(H.3). Per ogni $n \in \mathbb{N}$, la densità condizionata di (X_n, Z_n) dato $Y_{0:n} = y_{0:n}$ é un vettore casuale distribuito come una*

$N(\mu_n + \beta(Z_n/2), Z_n/2) \times GIG(\lambda_n, \delta_n, \gamma_n)$, dove $\mu_n(\cdot)$, $\delta_n(\cdot)$, $n \in \mathbb{N}$ sono definite ricorsivamente, per $n \geq 1$, da:

$$\begin{aligned} \mu_0 &= \frac{1}{2}(y_0 + \mu), & \mu_n &= \frac{1}{2}(y_n + \mu_{n-1}), \\ \delta_0 &= \sqrt{\delta^2 + \frac{1}{2}(y_0 - \mu)^2}, & \delta_n &= \sqrt{\delta_{n-1}^2 + \frac{1}{2}(y_n - \mu_{n-1})^2}. \end{aligned} \quad (3.4)$$

Dimostrazione Procediamo per induzione su n . Iniziamo provando che il risultato vale per $n = 0$, cioè mostriamo che la distribuzione condizionata di (X_0, Z_0) dato $Y_0 = y_0$ é un vettore casuale distribuito come una $N(\mu_0 + \beta(Z_0/2), Z_0/2) \times GIG(\lambda_0, \delta_0, \gamma_0)$, con $\mu_0 = \frac{1}{2}(y_0 + \mu)$ e $\delta_0 = \sqrt{\delta^2 + \frac{1}{2}(y_0 - \mu)^2}$. Da (H.3), abbiamo che (X_0, Z_0) é un vettore casuale assolutamente continuo con densità

$$\begin{aligned} p(x_0, z_0) &= \frac{1}{\sqrt{2\pi z_0}} \exp \left[-\frac{1}{2} \frac{(x_0 - \mu - \beta z_0)^2}{z_0} \right] \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} z_0^{\lambda-1} \\ &\times \exp \left[-\frac{1}{2} \left(\delta^2 \frac{1}{z_0} + \gamma^2 z_0 \right) \right]. \end{aligned}$$

Dalla (H.1) e dalla terza equazione in (3.2), si ha che la distribuzione condizionata di Y_0 dato $(X_0, Z_0) = (x_0, z_0)$ ha densità

$$p(y_0|x_0, z_0) = \frac{1}{\sqrt{2\pi z_0}} \exp \left[-\frac{1}{2} \frac{(y_0 - x_0)^2}{z_0} \right].$$

Per la regola di Bayes,

$$\begin{aligned} p(x_0, z_0|y_0) &\propto p(y_0|x_0, z_0)p(x_0, z_0) \\ &\propto z_0^{-1/2} \exp \left[-\frac{1}{2} \frac{(y_0 - x_0)^2}{z_0} \right] z_0^{-1/2} \exp \left[-\frac{1}{2} \frac{(x_0 - \mu - \beta z_0)^2}{z_0} \right] \\ &\times z_0^{\lambda-1} \exp \left[-\frac{1}{2} \left(\delta^2 \frac{1}{z_0} + \gamma^2 z_0 \right) \right] \\ &\propto \frac{1}{z_0} z_0^{\lambda-1} \exp \left[-\frac{1}{2} \left(\frac{(x_0 - y_0)^2}{z_0} + \frac{(x_0 - \mu - \beta z_0)^2}{z_0} + \delta^2 \frac{1}{z_0} + \gamma^2 z_0 \right) \right] \\ &\propto \frac{1}{z_0} z_0^{\lambda_0-1} \exp \left[-\frac{1}{2} \left(\frac{(x_0 - (y_0 + \mu + \beta z_0)/2)^2}{z_0/2} \right. \right. \\ &\left. \left. + \frac{y_0^2/2 + \mu^2/2 + \beta^2 z_0^2/2 - y_0 \mu}{z_0} + \delta^2 \frac{1}{z_0} + \gamma^2 z_0 \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{z_0}} \exp \left[-\frac{1}{2} \left(\frac{(x_0 - (y_0 + \mu + \beta z_0)/2)^2}{z_0/2} \right) \right] z_0^{\lambda_0 - 1} \\
&\times \exp \left[-\frac{1}{2} \left(\delta_0^2 \frac{1}{z_0} + \gamma_0^2 z_0 \right) \right]
\end{aligned}$$

Assumiamo ora che la distribuzione condizionata di (X_{n-1}, Z_{n-1}) dato $Y_{0:n-1} = y_{0:n-1}$ sia $N(\mu_{n-1} + \beta(Z_{n-1}/2), (Z_{n-1}/2)) \times GIG(\lambda_{n-1}, \delta_{n-1}, \gamma_{n-1})$. Per le prime due equazioni in (3.2), si ha che la distribuzione condizionata di (X_n, Z_n) dato $Y_{0:n-1} = y_{0:n-1}$ é assolutamente continua con densitá:

$$\begin{aligned}
p(x_n, z_n | y_{0:n-1}) &= \int_{\mathbb{R} \times [0, +\infty)} p(x_n, z_n | x_{n-1}, z_{n-1}) \\
&\quad \times p(x_{n-1}, z_{n-1} | y_{0:n-1}) dx_{n-1} dz_{n-1} \\
&= \int_{\mathbb{R} \times [0, +\infty)} p(x_n | z_n, x_{n-1}, z_{n-1}) p(z_n | z_{n-1}) p(x_{n-1} | z_{n-1}, y_{0:n-1}) \\
&\quad \times p(z_{n-1} | y_{0:n-1}) dx_{n-1} dz_{n-1} \\
&= \int_{[0, +\infty)} \left(\int_{\mathbb{R}} p(x_n | z_n, x_{n-1}, z_{n-1}) p(x_{n-1} | z_{n-1}, y_{0:n-1}) dx_{n-1} \right) \\
&\quad \times p(z_n | z_{n-1}) p(z_{n-1} | y_{0:n-1}) dz_{n-1}
\end{aligned}$$

Ora

$$\int_{\mathbb{R}} p(x_n | z_n, x_{n-1}, z_{n-1}) p(x_{n-1} | z_{n-1}, y_{0:n-1}) dx_{n-1}$$

non dipende da z_{n-1} e coincide con la densitá di una variabile casuale gaussiana di media $\mu_{n-1} + \beta z_n$ e varianza z_n . Infatti,

$$p(x_n | z_n, x_{n-1}, z_{n-1})$$

é $N(x_{n-1} + \beta(z_n - z_{n-1}/2); z_n - z_{n-1}/2)$ per (3.2) e

$$p(x_{n-1} | z_{n-1}, y_{0:n-1})$$

é $N(\mu_{n-1} + \beta z_{n-1}/2, z_{n-1}/2)$ per ipotesi induttiva. Inoltre

$$\int_{[0, +\infty)} p(z_n | z_{n-1}) p(z_{n-1} | y_{0:n-1}) dz_{n-1}$$

é la densitá di una variabile casuale $GIG(|\lambda_{n-1}|, \delta_{n-1}, \gamma_{n-1})$. Infatti,

$$p(z_n | z_{n-1})$$

é $GIG(|\lambda_{n-1}|, 0, \gamma_{n-1})$ valutata in $z_n - z_{n-1}$ e

$$p(z_{n-1}|y_{0:n-1})$$

é $GIG(\lambda_{n-1}, \delta_{n-1}, \gamma_{n-1})$ per ipotesi induttiva.

Applicando la proprietà (A.23), si ha la tesi. Quindi $p(x_n, z_n|y_{0:n-1})$ é la densità di una variabile casuale $N(\mu_{n-1} + \beta z_n, z_n) \times GIG(|\lambda_{n-1}|, \delta_{n-1}, \gamma_{n-1})$. Per concludere la dimostrazione, si applica la regola di Bayes, ottenendo:

$$p(x_n, z_n|y_{0:n}) = p(x_n, z_n|y_{0:n-1}, y_n) \propto p(y_n|x_n, z_n, y_{0:n-1})p(x_n, z_n|y_{0:n-1}).$$

Procedendo in modo del tutto analogo al caso $n = 0$, si ottiene che la distribuzione di (X_n, Z_n) dato $Y_{0:n} = y_{0:n}$ é un vettore casuale distribuito come $N(\mu_n + \beta Z_n/2, Z_n/2) \times GIG(\lambda_n, \delta_n, \gamma_n)$. Le ricorsioni (3.3) e (3.4) definiscono il filtro finito dimensionale.

□

Se si considera un processo parzialmente osservabile nelle applicazioni, ci sono ulteriori distribuzioni condizionate di interesse, oltre a quella di filtraggio.

Riguardo al processo $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$ definito dal sistema (3.2), si potrebbe essere interessati a determinare la densità condizionate di $\{X_{n+1}, Z_{n+1}\}$ dato $(Y_{0:n} = y_{0:n})$ e di $Y_{0:n+1}$ dato $(Y_{0:n} = y_{0:n})$, $n \in \mathbb{N}$. Queste distribuzioni sono dette *densità di previsione*.

Proposizione 3.2.2 Sotto (H.1)-(H.3), la densità condizionate di $(Y_{0:n+1})$

dato $Y_{0:n} = y_{0:n}$, $n \in \mathbb{N}$ é $GH(|\lambda_n|, \sqrt{\gamma_n^2/2 + \beta^2/4}, \beta/2, \mu_n, \sqrt{2}\delta_n)$, dove λ_n , $\gamma_n, \mu_n, \delta_n$, $n \in \mathbb{N}$ sono definiti ricorsivamente da (3.3) e (3.4).

Inoltre, $Y_0 \sim GH(\lambda, \sqrt{\gamma^2/2 + \beta^2/4}, \beta/2, \mu, \sqrt{2}\delta)$.

Dimostrazione Da (3.2) e dalle assunzioni (H.1)-(H.3), la distribuzione condizionate di Y_{n+1} dato $Y_{0:n} = y_{0:n}$, per ogni $n \geq 1$, ha densità che soddisfa alla seguente equazione:

$$\begin{aligned} p(y_{n+1}|y_{0:n}) &= \int_{\mathbb{R} \times [0, +\infty)} p(y_{n+1}|x_{n+1}, z_{n+1})p(x_{n+1}, z_{n+1}|y_{0:n})dx_{n+1}dz_{n+1} \\ &= \int_{[0, +\infty)} \left(\int_{\mathbb{R}} p(y_{n+1}|x_{n+1}, z_{n+1})p(x_{n+1}|z_{n+1}, y_{0:n})dx_{n+1} \right) dz_{n+1} \end{aligned}$$

$$\times p(z_{n+1}|y_{0:n})dz_{n+1} \quad (3.5)$$

Essendo

$$p(y_{n+1}|x_{n+1}, z_{n+1}) = \frac{1}{\sqrt{2\pi z_{n+1}}} \exp \left[-\frac{1}{2} \frac{(y_{n+1} - x_{n+1})^2}{z_{n+1}} \right]$$

e

$$\begin{aligned} p(x_{n+1}, z_{n+1}|y_{0:n}) &= p(x_{n+1}|z_{n+1}, y_{0:n})p(z_{n+1}|y_{0:n}) \\ &= \frac{1}{\sqrt{2\pi z_{n+1}}} \exp \left[\frac{(x_{n+1} - \mu_n - \beta z_{n+1})^2}{2z_{n+1}} \right] \frac{(\gamma_n/\delta_n)^{|\lambda_n|}}{2K_{|\lambda_n|}(\delta_n \gamma_n)} z_{n+1}^{|\lambda_n|-1} \\ &\quad \times \exp \left[-\frac{1}{2} \left(\delta_n^2 \frac{1}{z_{n+1}} + \gamma_n^2 z_{n+1} \right) \right] \end{aligned}$$

si ottiene

$$\begin{aligned} &\int_{\mathbb{R}} p(y_{n+1}|x_{n+1}, z_{n+1})p(x_{n+1}|z_{n+1}, y_{0:n})dx_{n+1} \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi z_{n+1}}} \exp \left[-\frac{(x_{n+1} - \mu_n - \beta z_{n+1})^2}{2z_{n+1}} \right] \frac{1}{\sqrt{2\pi z_{n+1}}} \\ &\quad \times \exp \left[-\frac{(y_{n+1} - x_{n+1})^2}{2z_{n+1}} \right] dx_{n+1} \end{aligned}$$

Per le proprietà di convoluzione di variabili casuali gaussiane, si ha che il primo integrale corrisponde alla densità di una $N(\mu_n + (\beta/2)2z_{n+1}, 2z_{n+1})$. Quindi, dalla (3.5), si ha che $p(y_{n+1}|y_{0:n})$ è la densità marginale della prima componente S di un vettore casuale

$$(S, T) \sim N(\mu_n + (\beta/2)T, T) \times GIG(|\lambda_n|, \sqrt{2}\delta_n, (\gamma_n/\sqrt{2})).$$

Come sottolineato nell'Appendice A (A.25), questo significa che $p(y_{n+1}|y_{0:n})$ è la densità di una variabile casuale $GH(|\lambda_n|, \sqrt{\gamma_n^2/2 + \beta^2/4}, \beta/2, \mu_n, \sqrt{2}\delta_n)$. La dimostrazione per Y_0 è del tutto analoga.

□

Osservazione La terza equazione del sistema (3.2) assicura che la distribuzione condizionata di Y_n dato $(X_n, Z_n) = (x_n, z_n)$, $n \in \mathbb{N}$, é $N(x_n, z_n)$. Cosí il processo $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$ fornisce un modello in cui sia la media X_n che la varianza Z_n , $n \in \mathbb{N}$, delle osservazioni gaussiane sono stocastiche. Inoltre, se la prima equazione in (3.2) viene sostituita da $X_{n+1} = \mu + \beta Z_{n+1}$, si ottiene un modello stocastico per la sola varianza.

Considerando le ricorsioni di aggiornamento per i parametri (3.3), (3.4) si possono descrivere altre proprietà interessanti di questo processo parzialmente osservabile. Dato che $\lambda \in [0, \frac{1}{2})$, é immediato mostrare che, per ogni n , $\lambda_n \leq 0$, con $\lambda_{2n} = \lambda - \frac{1}{2}$ e $\lambda_{2n+1} = -\lambda$. Possono essere sottolineati due casi particolari.

1. Se $\lambda = 0$ allora λ_n prende valori $-\frac{1}{2}$ e 0. Allora, (X_n, Z_n) dato $Y_{0:n}$, $n \in \mathbb{N}$, segue, alternativamente, una distribuzione bivariata gaussiana gaussiana inversa e una distribuzione bivariata gaussiana iperbolica, mentre Y_0 segue una distribuzione iperbolicoide e Y_{n+1} dato Y_n , $n \in \mathbb{N}$, alternativamente, una distribuzione gaussiana inversa reciproca e una distribuzione iperbolicoide.
2. Se $\lambda = \frac{1}{4}$, $\lambda_n = -\frac{1}{4}$ per ogni $n \in \mathbb{N}$. In questo contesto, assumiamo $\beta = 0$, $\gamma = 2\delta$, $\mu = 0$, cosí $\{W_{n+1}\}$ é una sequenza di variabili casuali gamma indipendenti di parametri $(\frac{1}{4}, 2\delta^2)$, $Y_0 \sim GH(1/4, \sqrt{2}\delta, 0, 0, \sqrt{2}\delta)$, particolare distribuzione iperbolica generalizzata simmetrica (vedi [5]).

3.2.2 Caso generale

Senza assumere restrizioni sul dominio del parametro λ , si ha il seguente sistema stocastico non lineare:

$$\begin{aligned}
 (X_0, Z_0) & \quad \text{vettore casuale dato} \\
 X_{n+1} & = \frac{Z_{n+1}}{Z_n} X_n + \mu_n + \beta Z_{n+1} - \frac{Z_{n+1}}{Z_n} \left(\mu_n + \beta \frac{Z_n}{2} \right) \\
 & + \sqrt{\frac{Z_{n+1}}{2}} V_{n+1}, n \geq 0 \\
 Z_{n+1} & = Z_n^{-\Phi(\lambda_n)} + W_{n+1}, n \geq 0 \\
 Y_n & = X_n + \sqrt{Z_n} V'_n, n \geq 0
 \end{aligned} \tag{3.6}$$

con $\{V_{n+1}\}_{n \in \mathbb{N}}$, $\{W_{n+1}\}_{n \in \mathbb{N}}$ come nella sezione precedente. Qui, $\lambda \in \mathbb{R}$, μ_n é definita ricorsivamente dalla prima equazione in (3.4) e $\Phi(\cdot)$ é la versione continua a destra della funzione $\text{sgn}(\cdot)$.

Nella seconda equazione di (3.6), che descrive lo sviluppo del processo latente $\{Z_n\}_{n \in \mathbb{N}}$, c'è un termine addizionale rispetto al caso (3.2), cioè l'esponente $-\Phi(\lambda_n)$. Questa aggiunta é necessaria per applicare la proprietá di convoluzione (A.23) delle variabili GIG, se $\lambda_n > 0$. In questo caso l'equazione diventa $Z_{n+1} = Z_n^{-1} + W_{n+1}$ e Z_n^{-1} é una GIG con parametro negativo $-\lambda_n$. Questo non era necessario in (3.2), dato che $\lambda_n \leq 0$ per ogni $n \in \mathbb{N}$. Inoltre, la prima equazione in (3.6) dipende dalle osservazioni y_0, \dots, y_n attraverso μ_n e differisce da quella data in (3.2).

É facile provare che il processo stocastico parzialmente osservabile soluzione di (3.6) ammette un filtro finito dimensionale; inoltre é possibile definire le ricorsioni di aggiornamento anche per le distribuzioni di previsione di y e (x, z) . Le dimostrazioni sono simili a quelle delle Proposizioni 3.2.1 e 3.2.2: l'unica differenza é che, nel caso $\lambda_n > 0$, si considera la formula di convoluzione (A.23) insieme al risultato della distribuzione della reciproca di una GIG.

Proposizione 3.2.3 *Sia $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$ l'unica soluzione del sistema (3.6), assumiamo che valgano le ipotesi (H.1)-(H.3) e che $(X_0, Z_0) \sim N(\mu + \beta Z_0, Z_0) \times GIG(\lambda, \delta, \gamma)$, con $\lambda \in \mathbb{R}$. Allora, per ogni $n \in \mathbb{N}$, la densitá condizionata di (X_n, Z_n) dato $Y_{0:n} = y_{0:n}$ é quella di un vettore casuale $N(\mu_n + \beta(Z_n/2), Z_n/2) \times GIG(\lambda_n, \delta_n, \gamma_n)$, dove $\mu_n, \lambda_n, \delta_n$ e γ_n sono definiti ricorsivamente da (3.3) e (3.4).*

Proposizione 3.2.4 *Sotto le assunzioni della Proposizione (3.2.3), la densitá condizionata di (X_{n+1}, Z_{n+1}) dato $Y_{0:n} = y_{0:n}$, $\forall n \in \mathbb{N}$, é quella di un vettore casuale $N(\mu_n + \beta Z_{n+1}, Z_{n+1}, Z_{n+1}) \times GIG(|\lambda_n|, \delta_n, \gamma_n)$ e la densitá condizionata di Y_{n+1} dato $Y_{0:n} = y_{0:n}$, $\forall n \in \mathbb{N}$, é $GH(|\lambda_n|, \sqrt{\gamma_n^2/2 + \beta^2/4})$,*

$\beta/2, \mu_n, \sqrt{2}\delta_n$), dove $\mu_n, \lambda_n, \delta_n$ e γ_n sono definiti ricorsivamente da (3.3) e (3.4). Inoltre $Y_0 \sim GH(\lambda, \sqrt{\gamma^2/2 + \beta^2/4}, \mu, \sqrt{2}\delta)$.

Osservazione Le ricorsioni di aggiornamento per $\mu_n, \lambda_n, \delta_n$ e $\gamma_n, n \in \mathbb{N}$ date da (3.3) e (3.4) definiscono il filtro finito dimensionale sia per il modello piú semplice che per quello piú generale. Si puó mostrare che, dopo un numero opportuno di passi, l'indice λ_n varia come nel modello con $\lambda_n \in [0, \frac{1}{2})$. Questo significa che esiste un intero $\nu \in \mathbb{N}$ tale che $\lambda_\nu \in [0, \frac{1}{2})$ e $\lambda_{\nu+n} \leq 0$ con $\lambda_{\nu+2n} = \lambda_\nu - \frac{1}{2}$ e $\lambda_{\nu+2n+1} = \lambda_\nu, n \in \mathbb{N}$.

Capitolo 4

Filtraggio nel modello per media e varianza

In questo capitolo vengono riportati dei risultati di tipo empirico ottenuti implementando l'algoritmo Monte Carlo sequenziale (particle filter) descritto precedentemente. Il problema di filtraggio considerato riguarda il modello state space presentato nel Capitolo 3. Successivamente i risultati empirici ottenuti riguardo alla convergenza verranno analizzati teoricamente per dimostrarne la validità. L'implementazione è stata realizzata in R.

4.1 Modello

Il modello considerato è descritto dal seguente sistema di equazioni stocastiche alle differenze:

$$\begin{aligned} (X_0, Z_0) & \quad \text{vettore casuale dato} \\ X_{n+1} & = X_n + \beta \left(Z_{n+1} - \frac{Z_n}{2} \right) + \sqrt{Z_{n+1} - \frac{Z_n}{2}} V_{n+1} \\ Z_{n+1} & = Z_n + W_{n+1} \\ Y_n & = X_n + \sqrt{Z_n} V'_n, \quad n \in \mathbb{N} \end{aligned}$$

con $\beta \in \mathbb{R}$ e $\{V_{n+1}\}_{n \in \mathbb{N}}$, $\{V'_n\}_{n \in \mathbb{N}}$ e $\{W_{n+1}\}_{n \in \mathbb{N}}$ tre sequenze indipendenti di variabili casuali indipendenti. Valgono inoltre le seguenti assunzioni:

H.1 $\{V_{n+1}\}_{n \in \mathbb{N}}$ e $\{V'_n\}_{n \in \mathbb{N}}$ sono sequenze di variabili casuali gaussiane standard indipendenti

H.2 $\{W_{n+1}\}_{n \in \mathbb{N}}$ é una sequenza di variabili gamma indipendenti con parametro di forma $|\lambda_n|$ e parametro di scala $\gamma_n^2/2$, dove

$$\begin{aligned}\lambda_0 &= \lambda - \frac{1}{2}, & \lambda_n &= |\lambda_{n-1}| - \frac{1}{2}, & n &\geq 1 \\ \gamma_0 &= \sqrt{\gamma^2 + \frac{\beta^2}{2}}, & \gamma_n &= \sqrt{\gamma_{n-1}^2 + \frac{\beta^2}{2}}, & n &\geq 1\end{aligned}$$

H.3 $(X_0, Z_0) \sim N(\mu + \beta Z_0, Z_0) \times GIG(\lambda, \delta, \gamma)$ con $\lambda \in [0, \frac{1}{2})$; (X_0, Z_0) indipendente dai processi $\{V_{n+1}\}_{n \in \mathbb{N}}, \{V'_n\}_{n \in \mathbb{N}}$ e $\{W_{n+1}\}_{n \in \mathbb{N}}$.

Consideriamo il processo stocastico parzialmente osservabile $\{X_n, Z_n, Y_n\}_{n \in \mathbb{N}}$ soluzione del sistema sopra riportato. In questo caso abbiamo a disposizione il filtro esatto ottenuto nel Capitolo 3. Saremo quindi in grado di confrontare l'approssimazione empirica ottenuta mediante l'implementazione dell'algoritmo con la soluzione analitica del problema di filtraggio.

4.2 Simulazione

Consideriamo il caso in cui il parametro λ é compreso tra 0 e $\frac{1}{2}$.

Fissiamo un orizzonte temporale $N = 200$ e simuliamo i processi $\{X_n, Z_n, Y_n\}_{n=1, \dots, 200}$ con valori dei parametri fissati:

Parametri	
λ	0.25
δ	0.4
γ	0.4
μ	3
β	0.13

Procediamo nel seguente modo:

- definiamo i parametri della distribuzione del processo $\{W_n\}_{n=1, \dots, 200}$ utilizzando le ricorsioni (3.3);
- generiamo il processo $\{W_n\}_{n=1, \dots, 200}$ da una distribuzione Gamma di parametri $|\lambda_n|$ e $\gamma_n^2/2$;

- generiamo i processi degli errori $\{V_n\}_{n=1,\dots,200}$ e $\{V'_n\}_{n=1,\dots,200}$, realizzazioni di variabili casuali normali standard;
- simuliamo $N = 200$ valori dalla distribuzione iniziale $(X_0, Z_0) \sim N(\mu + \beta Z_0, Z_0) \times GIG(\lambda, \delta, \gamma)$ specificata nel modello;
- creiamo il processo $\{X_n, Z_n\}_{n=1,\dots,200}$;
- creiamo il processo delle osservazioni.

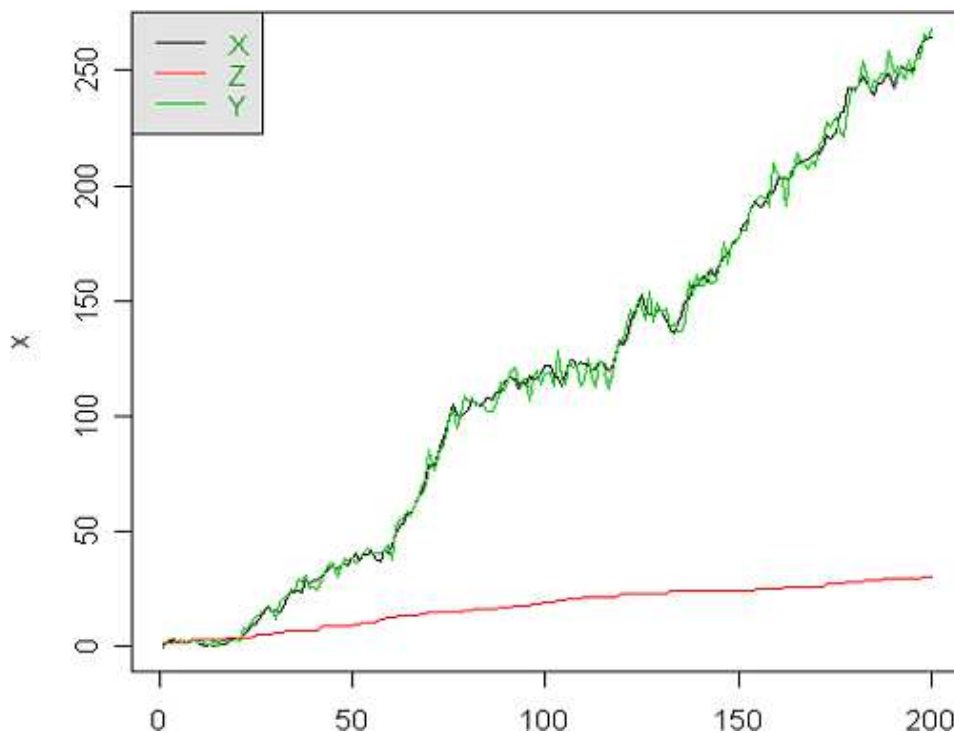


Figura 4.1: Processi Simulati

La Figura 4.1 rappresenta i tre processi simulati: in verde il processo osservato $\{Y_n\}_{n=1,\dots,200}$ di media $\{X_n\}_{n=1,\dots,200}$ (in nero) e varianza $\{Z_n\}_{n=1,\dots,200}$ (in rosso).

Procediamo ora con il metodo particle filter per stimare la distribuzione di filtraggio. Dato che in questo caso abbiamo a disposizione il filtro finito dimensionale esatto, possiamo confrontare il risultato approssimato dato

dall'algoritmo di filtraggio con la distribuzione condizionata esatta trovata nella Proposizione 3.2.1.

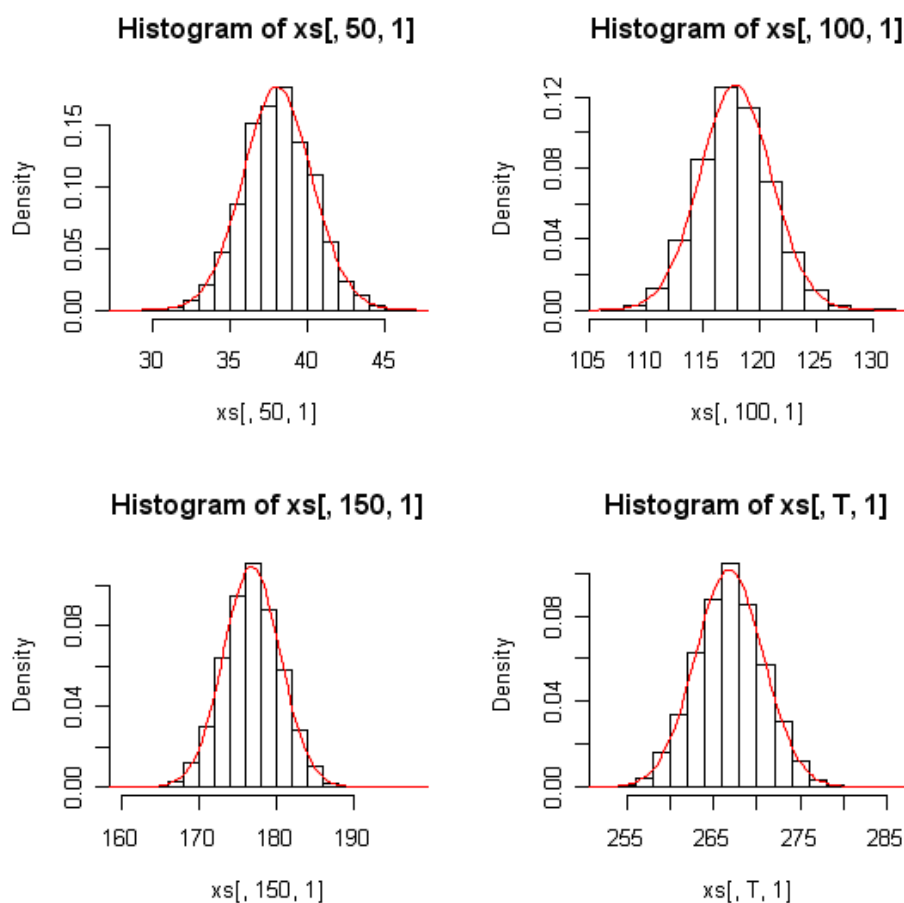


Figura 4.2: Istogramma e distribuzione esatta di $X_n|Y_{0:n}$

Utilizziamo un numero di particle $M = 10000$ per avere una buona stima della distribuzione di filtraggio.

Implementiamo il metodo Monte Carlo sequenziale con selezione dei particles multinomiale.

1. Inizializzazione, $n = 0$

- Per $i = 1, \dots, M$ genero i particles $z_0^{(i)} \sim GIG(\lambda, \delta, \gamma)$;
- Per $i = 1, \dots, M$ genero i particles $x_0^{(i)} \sim N(\mu + \beta z_0^{(i)}, z_0^{(i)})$;
- pongo $n = 1$.

2. Passo di campionamento per importanza

- Per $i = 1, \dots, M$ genero i particles

$$\tilde{z}_n^{(i)} \sim \text{Gamma}(|\lambda_n|, \frac{\gamma^2}{2}) + z_0^{(i)} \quad \text{e pongo}$$

$$\tilde{x}_{0:n}^{(i)} = (x_{0:n-1}^{(i)}, \tilde{x}_n^{(i)});$$

- Per $i = 1, \dots, M$ genero i particles

$$\tilde{x}_n^{(i)} \sim N\left(x_0^{(i)} + \beta\left(\tilde{z}_n^{(i)} - \frac{z_0^{(i)}}{2}\right), \left(\tilde{z}_n^{(i)} - \frac{z_0^{(i)}}{2}\right)^2\right)$$

e pongo

$$\tilde{x}_{0:n}^{(i)} = (x_{0:n-1}^{(i)}, \tilde{x}_n^{(i)});$$

- per $i = 1, \dots, M$ calcolo i pesi di importanza

$$\tilde{\omega}_n^{(i)} = \text{dnorm}(y_n; \tilde{x}_n^{(i)}, \tilde{z}_n^{(i)});$$

- normalizzo i pesi di importanza.

3. Passo di selezione

- campiono con ripetizione M particles $(x_{0:n}^{(1)}, \dots, x_{0:n}^{(M)})$ dall'insieme $\{\tilde{x}_{0:n}^{(1)}, \dots, \tilde{x}_{0:n}^{(M)}\}$ secondo i pesi di importanza;
- pongo $n := n+1$ e torno al passo 2.

Nel seguito tenteremo di analizzare graficamente il problema della convergenza dell'approssimazione empirica della distribuzione di filtraggio ottenuta tramite l'algoritmo alla distribuzione esatta e il problema della sta-

bilità dell'equazione ricorsiva di filtraggio.

4.2.1 Convergenza

Una importante informazione che si può ottenere riguarda la forma della distribuzione di filtraggio.

Per ogni istante temporale è possibile rappresentare l'approssimazione empirica della densità di filtraggio ottenuta tramite l'algoritmo standard.

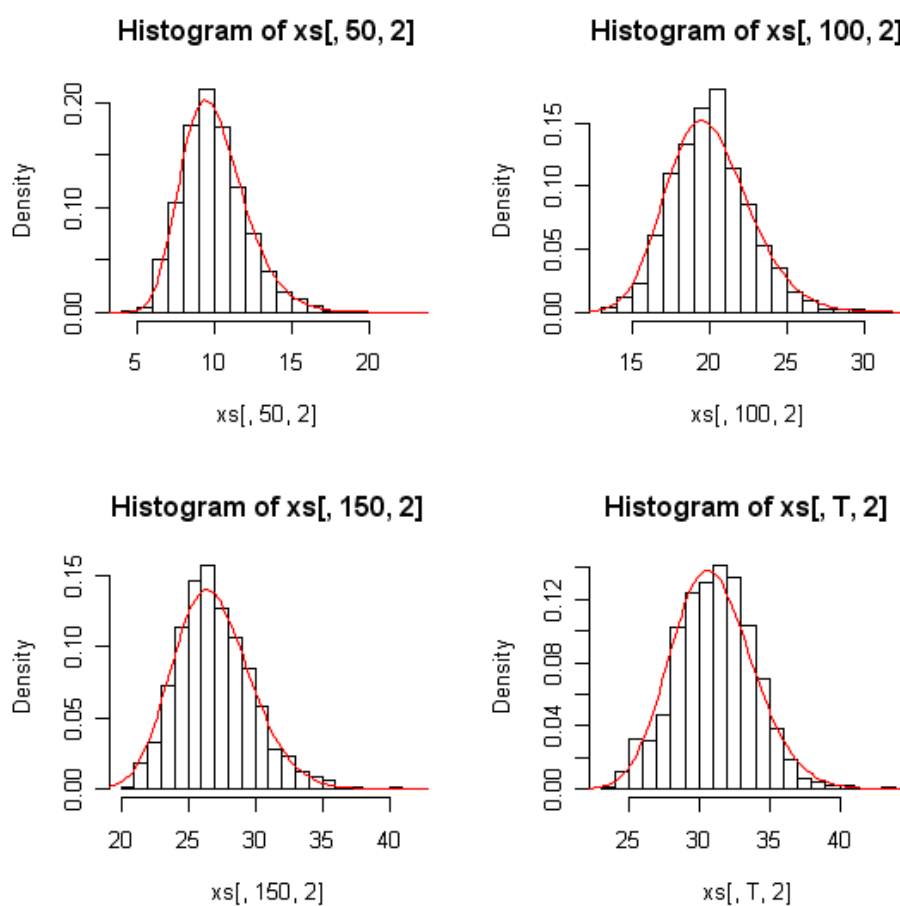


Figura 4.3: Istogramma e distribuzione esatta di $Z_n|Y_{0:n}$

Confronteremo quindi la distribuzione marginale di X_n condizionata a $Y_{0:n}$ ottenuta con l'algoritmo (distribuzione empirica associata ai particles) con la distribuzione esatta che, nel caso in questione, coincide con una distribuzione iperbolica generalizzata, $GH(\sqrt{\beta^2 + 2\gamma_n^2}, \beta, \delta_n/\sqrt{2}, \mu_n, \lambda_n)$.

La Figura 4.2 mostra l'istogramma relativo al campione di $M = 10000$ particles riferito alla variabile X_n e rappresentato per i tempi $n = 50, 100, 150, 200$. In rosso é riportata la distribuzione iperbolica generalizzata esatta. Dai grafici di può notare che l'approssimazione é buona.

Una analisi del tutto analoga é stata fatta per la variabile Z_n , dove l'istogramma relativo alla distribuzione empirica associata al campione rappresentato dai particles é stato confrontato con la distribuzione marginale esatta di Z_n condizionata a $Y_{0:n}$, cioè con una distribuzione GIG, $GIG(\lambda_n, \delta_n, \gamma_n)$ (Figura 4.3).

Anche in questo caso possiamo affermare che l'approssimazione é buona.

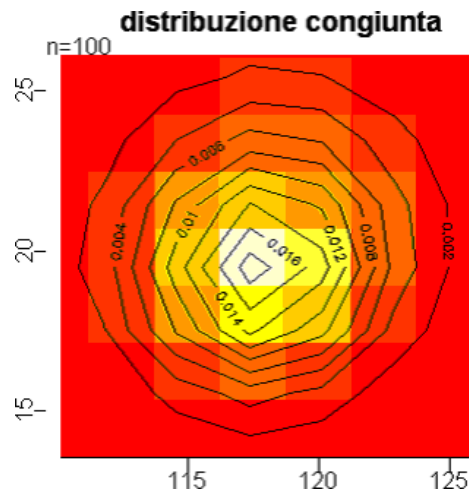


Figura 4.4: Distribuzione empirica congiunta e curve di livello $n = 100$

Guardiamo la distribuzione congiunta di $\{X_n, Z_n\}$ condizionata a $Y_{0:n}$. Rappresentiamo le curve di livello della distribuzione di filtraggio esatta data dalla Proposizione 3.2.1, cioè $N(\mu_n + \beta(Z_n/2), Z_n/2) \times GIG(\lambda_n, \delta_n, \gamma_n)$ e aggiungiamo nel grafico la distribuzione empirica del campione congiunto

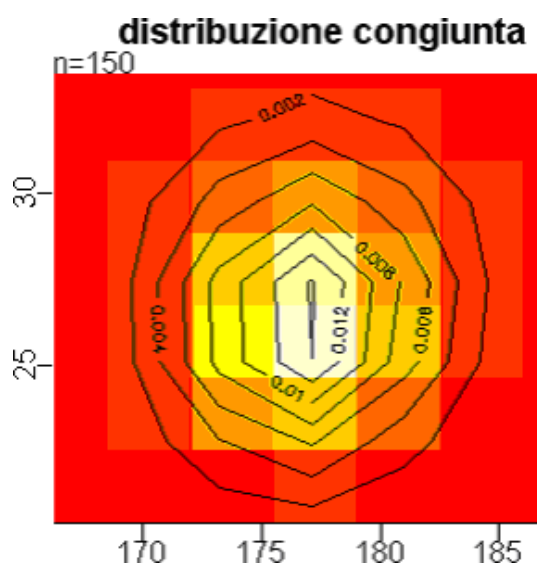


Figura 4.5: Distribuzione empirica congiunta e curve di livello $n = 150$

$\{X_n, Z_n\}$ ottenuto tramite l'algoritmo di filtraggio.

Riportiamo i grafici relativi agli istanti $n = 100$ e $n = 150$ (Figure 4.4-4.5). Le distribuzioni empiriche (sia marginali che congiunte) dei particles ottenuti tramite l'algoritmo sembrano approssimare in modo soddisfacente le distribuzioni esatte. Questi risultati di convergenza verranno in seguito analizzati da un punto di vista teorico.

4.2.2 Stabilità

Interessiamoci ora al problema della stabilità della formula ricorsiva di filtraggio. Il sistema sarà *asintoticamente stabile* se il suo comportamento a lungo termine non dipende dalla condizione iniziale. Ripetiamo quindi l'analisi partendo da una distribuzione iniziale concentrata nei punti $x = 6.5$ e $z = 2.5$. In questo modo possiamo giudicare empiricamente la proprietà di stabilità dell'equazione ricorsiva di filtraggio.

Riportiamo anche in questo caso l'istogramma relativo al campione di particles riferito alla variabile X_n e rappresentato per i tempi $n = 50, 100, 150, 200$, insieme alla distribuzione iperbolica generalizzata esatta (Figura 4.6) e

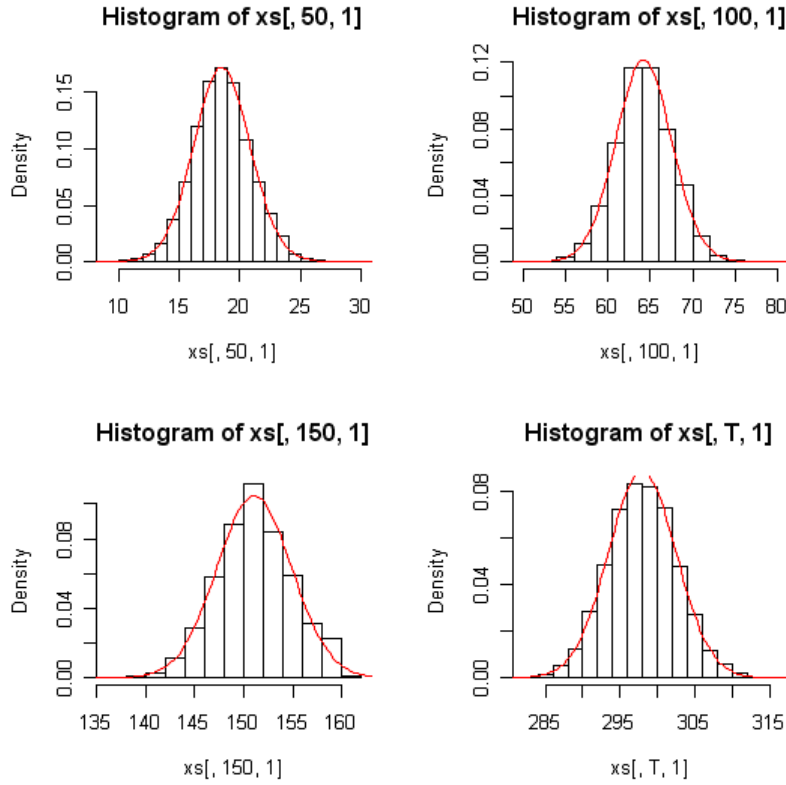


Figura 4.6: Istogramma e distribuzione esatta di $X_n|Y_{0:n}$

l'istogramma relativo alla variabile Z_n , insieme alla distribuzione marginale esatta di Z_n condizionata a $Y_{0:n}$, cioè una distribuzione GIG, $GIG(\lambda_n, \delta_n, \gamma_n)$ (Figura 4.7).

Rappresentiamo le curve di livello della distribuzione di filtraggio esatta $N(\mu_n + \beta(Z_n/2), Z_n/2) \times GIG(\lambda_n, \delta_n, \gamma_n)$, aggiungendo nel grafico la distribuzione empirica del campione congiunto $\{X_n, Z_n\}$ ottenuto tramite l'algoritmo di filtraggio. Riportiamo i grafici relativi agli istanti $n = 100$ e $n = 150$ (Figure 4.8-4.9).

Per convincerci che le cose, almeno da un punto di vista empirico, effettivamente funzionano, ripetiamo l'analisi partendo da una distribuzione iniziale concentrata nei punti $x = 2.9$ e $z = 3.8$. I grafici riportati hanno lo stes-

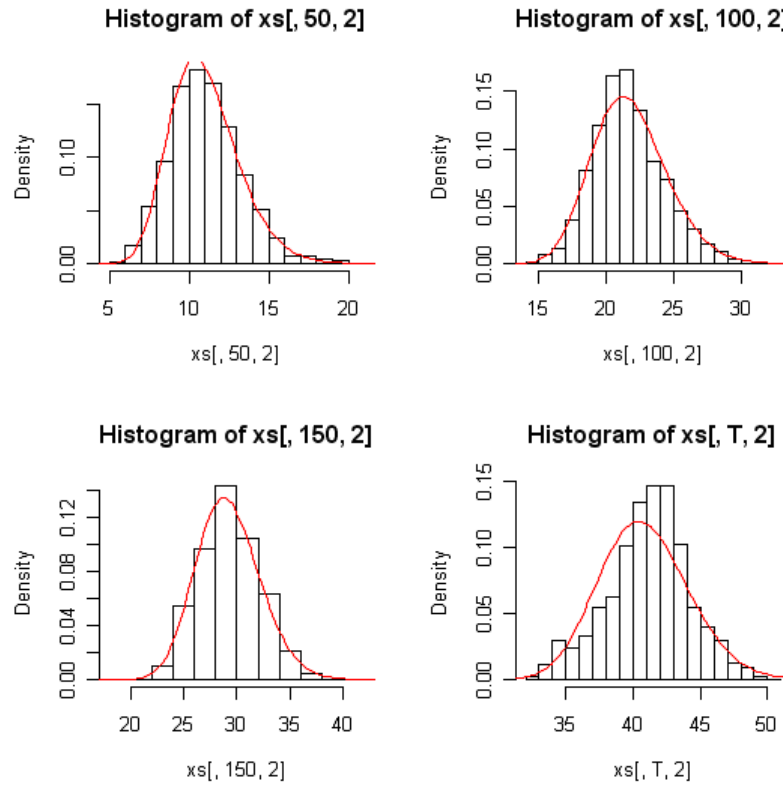
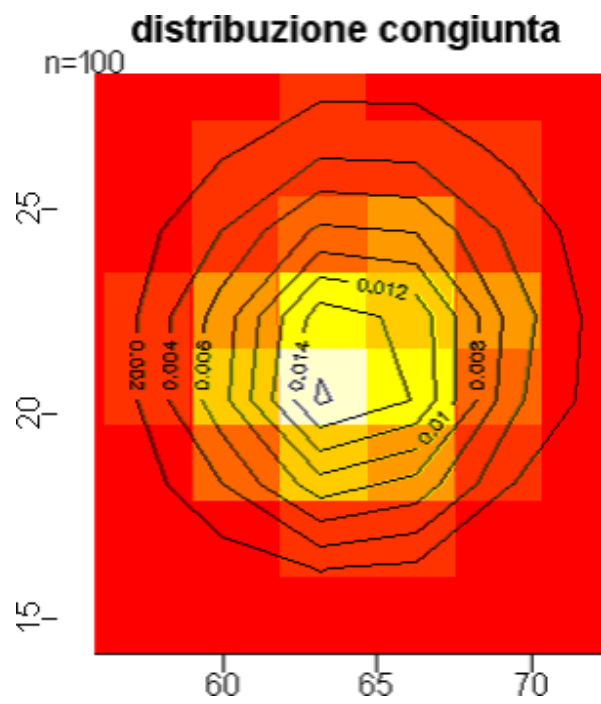
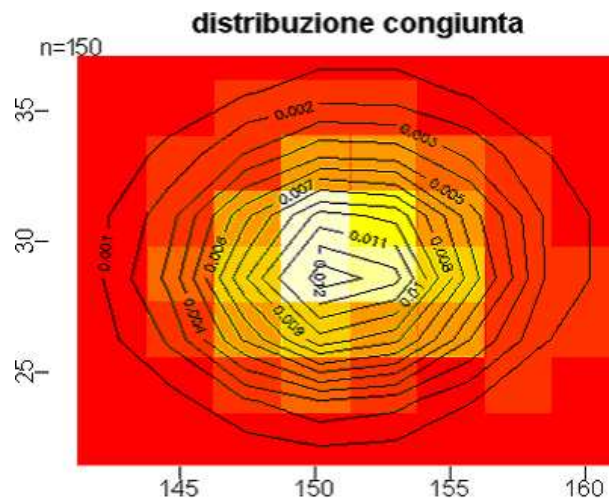


Figura 4.7: Istogramma e distribuzione esatta di $Z_n|Y_{0:n}$

so significato dei precedenti (Figure 4.10-4.11-4.12-4.13). Proviamo ancora partendo da una distribuzione marginale di $(X_0) \sim N(0, 1)$ e da una distribuzione marginale di $(Z_0) \sim Gamma(\lambda, \gamma)$. L'analisi effettuata é analoga a quella presentata per i precedenti casi (Figure 4.14-4.15-4.16-4.17). Possiamo notare come in tutti i casi presentati, la distribuzione empirica ottenuta con l'algoritmo particle filter approssimi 'bene' il filtro esatto.

Figura 4.8: Distribuzione empirica congiunta e curve di livello $n = 100$ Figura 4.9: Distribuzione empirica congiunta e curve di livello $n = 150$

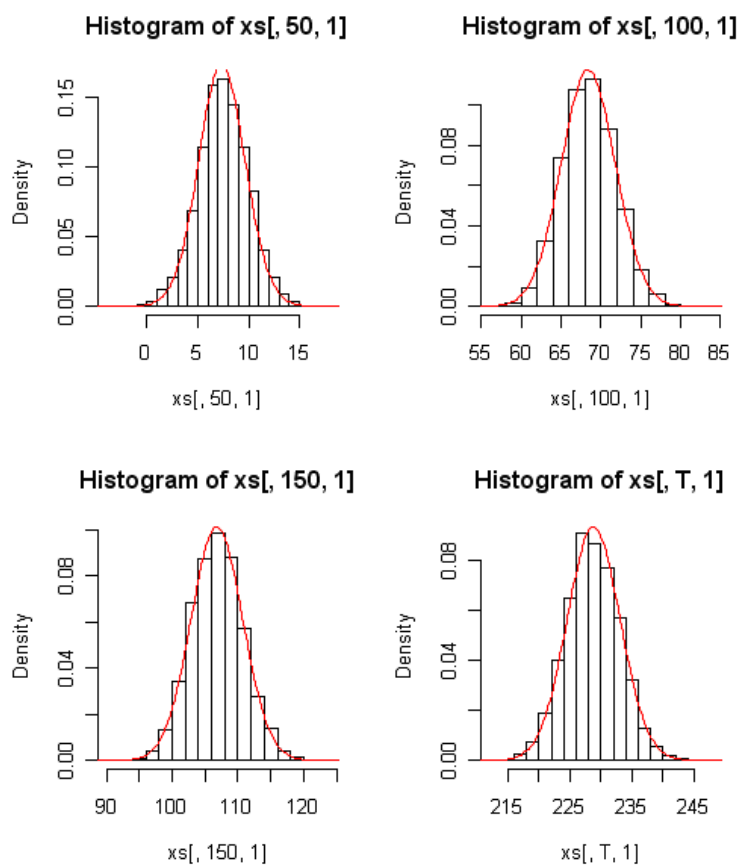
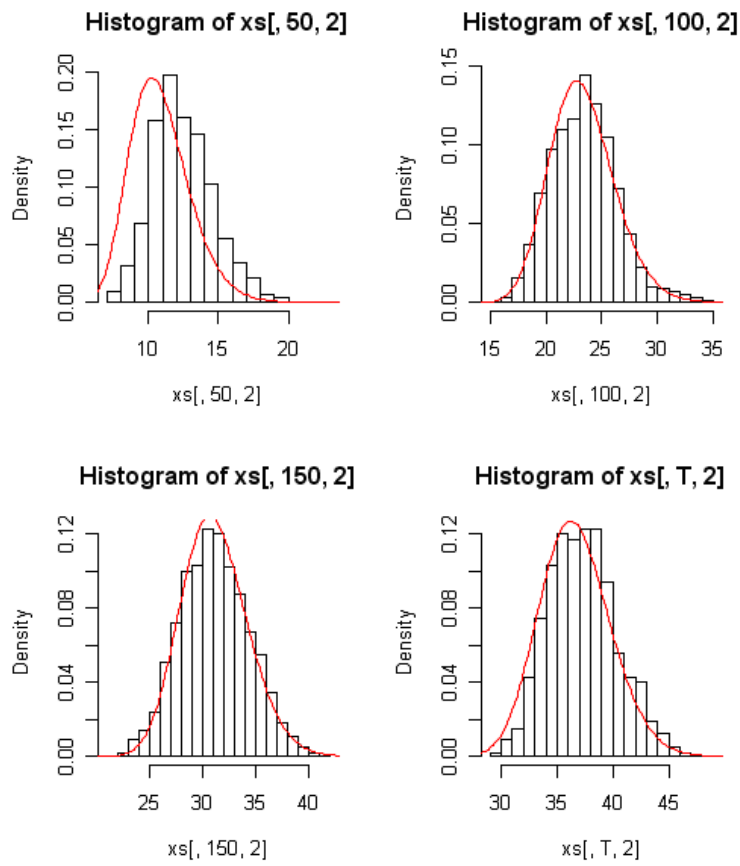


Figura 4.10: Istogramma e distribuzione esatta di $X_n|Y_{0:n}$

Figura 4.11: Istogramma e distribuzione esatta di $Z_n|Y_{0:n}$

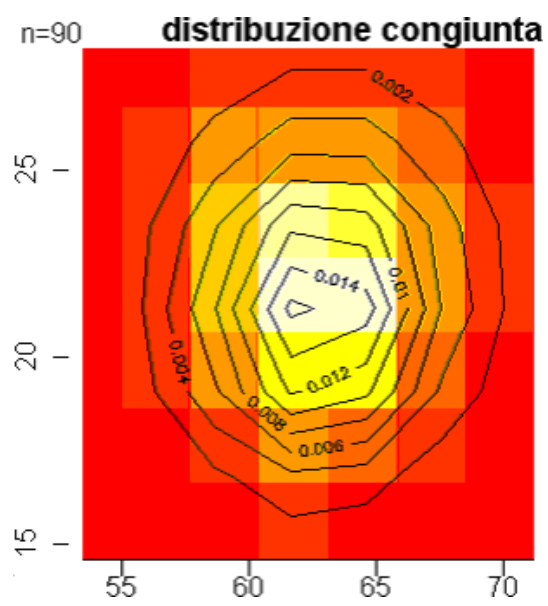


Figura 4.12: Distribuzione empirica congiunta e curve di livello $n = 90$

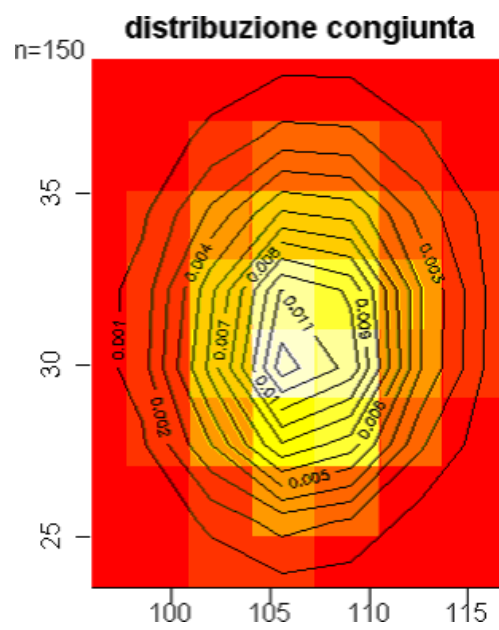


Figura 4.13: Distribuzione empirica congiunta e curve di livello $n = 150$

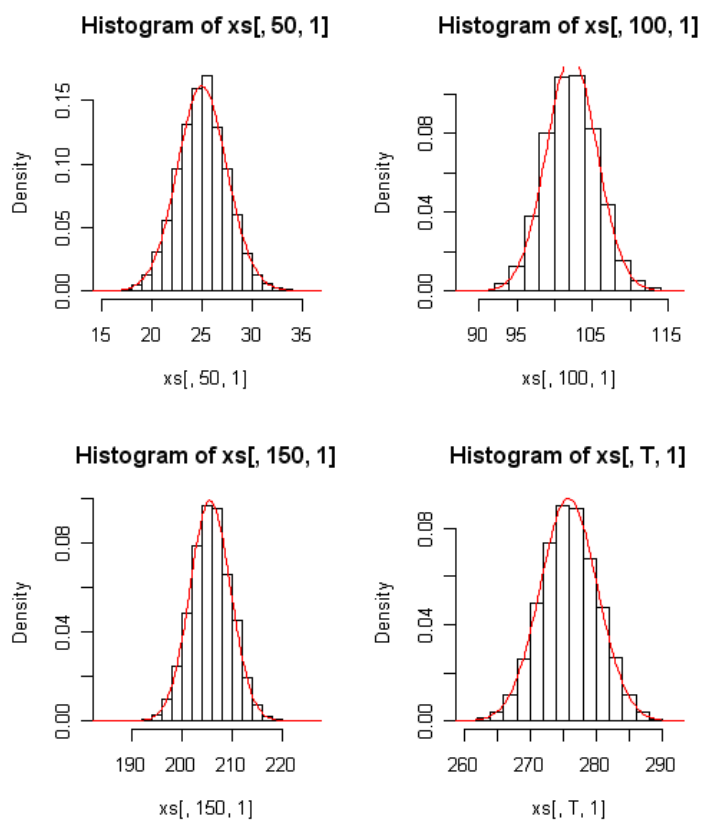
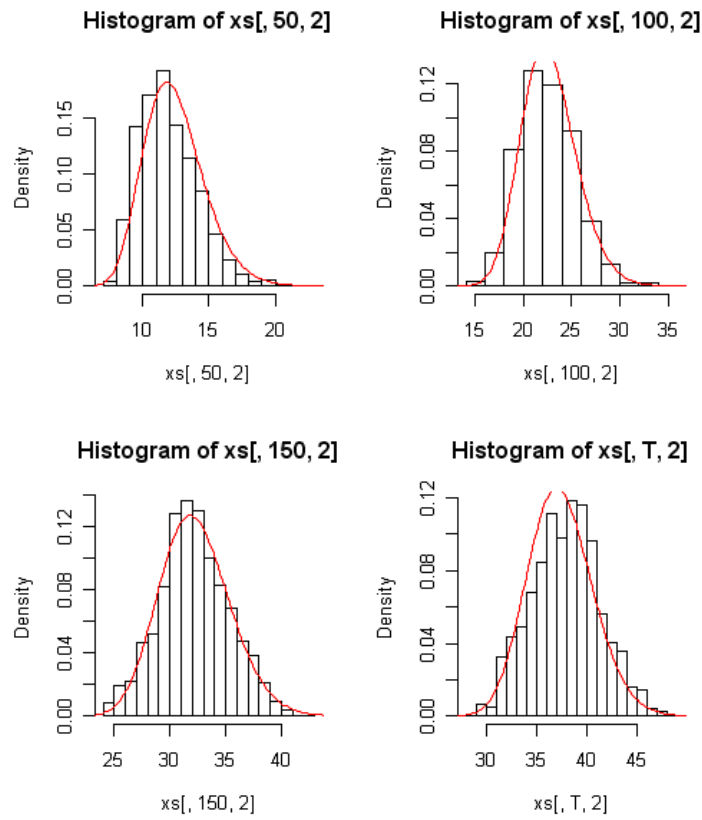


Figura 4.14: Istogramma e distribuzione esatta di $X_n|Y_{0:n}$

Figura 4.15: Istogramma e distribuzione esatta di $Z_n | Y_{0:n}$

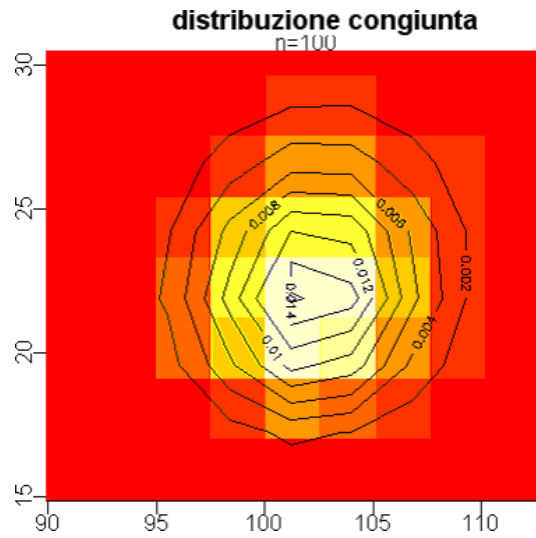


Figura 4.16: Distribuzione empirica congiunta e curve di livello $n = 100$

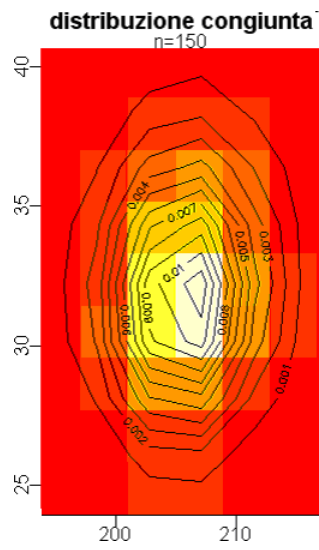


Figura 4.17: Distribuzione empirica congiunta e curve di livello $n = 150$

Capitolo 5

Convergenza di metodi

Monte Carlo sequenziali

In questo capitolo verrà analizzata la convergenza di metodi Monte Carlo sequenziali. Seguendo il lavoro di Crisan e Doucet (2002), verranno fornite condizioni sufficienti per la convergenza del particle filter alle distribuzioni di filtraggio e di previsione in ogni istante di tempo fissato.

5.1 Notazione e terminologia

Siano \mathbb{R}^d lo spazio euclideo d -dimensionale e $\mathcal{B}(\mathbb{R}^d)$ l'associata σ -algebra di Borel. Verranno utilizzati i seguenti insiemi di funzioni:

- $B(\mathbb{R}^d)$: insieme delle funzioni $\mathcal{B}(\mathbb{R}^d)$ -misurabili limitate definite su \mathbb{R}^d ,
- $\mathcal{C}_b(\mathbb{R}^d)$: l'insieme delle funzioni continue limitate definite su \mathbb{R}^d ,
- $\mathcal{C}_k(\mathbb{R}^d)$: l'insieme delle funzioni continue a supporto compatto definite su \mathbb{R}^d .

Se $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$, definiamo la norma:

$$\|\varphi\| = \sup_{x \in \mathbb{R}^d} |\varphi(x)|.$$

Verranno considerati inoltre i seguenti insiemi di misure:

- $\mathcal{M}_F(\mathbb{R}^d)$: l'insieme delle misure finite su $\mathcal{B}(\mathbb{R}^d)$,
- $\mathcal{P}(\mathbb{R}^d)$: l'insieme delle misure di probabilità su $\mathcal{B}(\mathbb{R}^d)$.

Se $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ (o $\mu \in \mathcal{P}(\mathbb{R}^d)$) e $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$, l'integrale di φ rispetto a μ verrà indicato con (μ, φ) :

$$(\mu, \varphi) = \int_{\mathbb{R}^d} \varphi(x) \mu(dx).$$

Si assume che $\mathcal{M}_F(\mathbb{R}^d)$ e, rispettivamente, $\mathcal{P}(\mathbb{R}^d)$ siano dotati della topologia debole. Sia $(\mu^M)_{M \in \mathbb{N}^+}$ una successione di misure finite; si dirà che μ^M converge debolmente a $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ e si scriverà $\lim_{M \rightarrow +\infty} \mu^M = \mu$ se

$$\lim_{M \rightarrow +\infty} (\mu^M, \varphi) = (\mu, \varphi) \quad \forall \varphi \in \mathcal{C}_b(\mathbb{R}^d),$$

e si ha la stessa definizione se $(\mu^M)_{M \in \mathbb{N}^+}$ è una successione di misure di probabilità. Inoltre si indicherà la misura concentrata in $a \in \mathbb{R}^d$ con δ_a e la funzione costante 1 con $\bar{\mathbf{1}}$.

Dato che \mathbb{R}^d è uno spazio metrico localmente compatto separabile, esiste un insieme numerabile \mathcal{M} denso in $\mathcal{C}_k(\mathbb{R}^d)$. Allora $\mathcal{M} \cup \{\bar{\mathbf{1}}\}$ è un insieme che determina la convergenza, cioè, se $\nu_M, M \in \mathbb{N}^+$, e ν sono misure finite e $\lim_{M \rightarrow \infty} (\nu_M, \varphi) = (\nu, \varphi)$ per ogni $\varphi \in \mathcal{M} \cup \{\bar{\mathbf{1}}\}$, allora $\lim_{M \rightarrow \infty} \nu_M = \nu$. Usando questo insieme, si può definire la seguente distanza su $\mathcal{P}(\mathbb{R}^d)$

$$d(\mu, \nu) = \sum_{\varphi_k \in \mathcal{M}} \frac{|(\mu, \varphi_k) - (\nu, \varphi_k)|}{2^k \|\varphi_k\|} + |(\mu, \bar{\mathbf{1}}) - (\nu, \bar{\mathbf{1}})|$$

che induce la topologia debole

$$\lim_{M \rightarrow +\infty} \nu_M = \nu \iff \lim_{M \rightarrow +\infty} d(\nu_M, \nu) = 0.$$

5.1.1 Processi di Markov e nuclei di transizione

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e sia $X = \{X_n, n \in \mathbb{N}\}$ un processo stocastico definito su (Ω, \mathcal{F}, P) a valori in \mathbb{R}^{n_x} . Sia \mathcal{F}_n^X la σ -algebra generata dal processo fino all'istante n , cioè $\mathcal{F}_n^X = \sigma(X_s, s \leq n)$. Allora X è un *processo di Markov* se, per ogni $n \in \mathbb{N}$ e per ogni $A \in \mathcal{B}(\mathbb{R}^{n_x})$, dove $\mathcal{B}(\mathbb{R}^{n_x})$ rappresenta la σ -algebra di Borel associata allo spazio euclideo n_x -dimensionale \mathbb{R}^{n_x} ,

$$P(X_{n+1} \in A | \mathcal{F}_n^X) = P(X_{n+1} \in A | X_n).$$

Il *nucleo di transizione* di un processo di Markov X é la funzione $K_n(\cdot, \cdot)$ definita su $\mathbb{R}^{n_x} \times \mathcal{B}(\mathbb{R}^{n_x})$ tale che, per ogni $n \in \mathbb{N}^+$ e per ogni $x \in \mathbb{R}^{n_x}$,

$$K_n(x, A) = P(X_n \in A | X_{n-1} = x). \quad (5.1)$$

Il nucleo di transizione K_n ha le seguenti proprietà:

- $K_n(x, \cdot)$ é una misura di probabilità su \mathbb{R}^{n_x} , per ogni $n \in \mathbb{N}^+$ e per ogni $x \in \mathbb{R}^{n_x}$;
- $K_n(\cdot, A) \in \mathcal{B}(\mathbb{R}^{n_x})$, per ogni $n \in \mathbb{N}^+$ e per ogni $A \in \mathcal{B}(\mathbb{R}^{n_x})$.

La distribuzione di X é univocamente determinata dalla sua distribuzione iniziale e dal suo nucleo di transizione. Sia q_n la distribuzione di probabilità della variabile casuale X_n :

$$q_n(A) = P(X_n \in A).$$

Allora dalla (5.1) si deduce che q_n soddisfa alla formula ricorsiva

$$q_{n+1}(A) = \int_{\mathbb{R}^{n_x}} K_{n+1}(x, A) q_n(dx).$$

Quindi $q_n = q_0 K_1 K_2 \dots K_n$.

5.1.2 Modello considerato

In questo capitolo si considererá il seguente modello state space. Sia $X = \{X_n, n \in \mathbb{N}\}$ un processo di Markov a valori in \mathbb{R}^{n_x} con nucleo di transizione K_n . Sia inoltre $Y = \{Y_n, n \in \mathbb{N}\}$ un processo stocastico a valori in \mathbb{R}^{n_y} definito da

$$Y_n = h_n(X_n, W_n), \quad n > 0,$$

con $Y_0 = 0$. Inoltre $h_n(\cdot, W_n)$ é continua su \mathbb{R}^{n_x} per ogni $n \in \mathbb{N}^+$ e i vettori casuali $W_n : \Omega \rightarrow \mathbb{R}^{n_y}$, $n \in \mathbb{N}^+$ sono indipendenti e le loro distribuzioni sono assolutamente continue rispetto alla misura di Lebesgue λ su \mathbb{R}^{n_y} . Il processo X ha distribuzione iniziale $X_0 \sim a_0(x) d\mu(x)$ e nucleo di transizione $K_n(x_{n-1}, dx_n)$ tale che

$$P(X_n \in A | X_{n-1} = x_{n-1}) = \int_A K_n(x_{n-1}, dx_n), \quad A \in \mathcal{B}(\mathbb{R}^{n_x}).$$

Le osservazioni hanno distribuzione marginale

$$P(Y_n \in B | X_n = x_n) = \int_B b_n(x_n, dy_n), \quad B \in \mathcal{B}(\mathbb{R}^{n_y}).$$

Per semplicitá assumiamo che $K_n(x_{n-1}, dx_n)$ e $b_n(x_n, dy_n)$ ammettano densitá rispetto alla misura di Lebesgue. Questo significa che $P(X_n \in dx_n | X_{n-1} = x_{n-1}) = K_n(x_{n-1}, dx_n) = K_n(x_{n-1}, x_n)dx_n$, e $P(Y_n \in dy_n | X_n = x_n) = b_n(x_n, dy_n) = b_n(x_n, y_n)dy_n$.

Esempio: Consideriamo il modello dinamico scalare

$$\begin{aligned} X_n &= g_n(X_{n-1}) + V_n, \\ Y_n &= h_n(X_n) + W_n \end{aligned}$$

dove $\{V_n\}_{n \in \mathbb{N}}$ e $\{W_n\}_{n \in \mathbb{N}^+}$ sono entrambe sequenze indipendenti e identicamente distribuite con

$$\begin{aligned} P(V_n \in C) &= \int_C P_V(dv) = \int_C p_V(v)dv, & C \in \mathcal{B}(\mathbb{R}) \\ P(W_n \in D) &= \int_D P_W(dw) = \int_D p_W(w)dw, & D \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

Allora si ha

$$\begin{aligned} K_n(x_{n-1}, x_n) &= p_V(x_n - g_n(x_{n-1})) \\ b_n(x_n, y_n) &= p_W(y_n - h_n(x_n)). \end{aligned}$$

5.2 Filtro Ottimo

Il problema qui considerato consiste nel calcolare le distribuzioni di filtraggio, cioè nel determinare la misura di probabilitá $\pi_{k:l|m}$, dove

$$\pi_{k:l|m}(dx_{k:l}) = P(X_{k:l} \in dx_{k:l} | Y_{1:m} = y_{1:m}).$$

Ricordando le ricorsioni (1.6) e (1.7) determinate nel Capitolo 1, si ha

$$\begin{aligned} \text{previsione } \pi_{n|n-1}(dx_n) &= \int_{\mathbb{R}^{n_x}} \pi_{n-1|n-1}(dx_{n-1}) K_n(x_{n-1}, dx_n) \\ \text{aggiornamento } \pi_{n|n}(dx_n) &= \frac{b_n(x_n, y_n) \pi_{n|n-1}(dx_n)}{\int_{\mathbb{R}^{n_x}} b_n(x, y_n) \pi_{n|n-1}(dx_n)}. \end{aligned} \tag{5.2}$$

Usando la notazione introdotta precedentemente, é facile vedere che per ogni funzione $\varphi : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$, la formula ricorsiva (5.2) implica che

$$\begin{aligned} \text{previsione } (\pi_{n|n-1}, \varphi) &= (\pi_{n-1|n-1}, K_n \varphi) \\ \text{aggiornamento } (\pi_{n|n}, \varphi) &= \frac{(\pi_{n|n-1}, \varphi b_n)}{(\pi_{n|n-1}, b_n)}. \end{aligned} \quad (5.3)$$

Esempio: Per un modello dinamico in cui $\pi_{n|n-1}(dx_n)$ e $\pi_{n|n}(dx_n)$ ammettono densità $p(x_n|y_{1:n-1})$ e $p(x_n|y_{1:n})$, la (5.2) diventa

$$\begin{aligned} p(x_n|y_{1:n-1}) &= \int_{\mathbb{R}^{n_x}} p(x_{n-1}|y_{1:n-1}) K_n(x_{n-1}, x_n) dx_{n-1} \\ p(x_n|y_{1:n}) &= \frac{\int_{\mathbb{R}^{n_x}} b_n(x, y_n) p(x_n|y_{1:n-1}) \varphi(x_n) dx_n}{\int_{\mathbb{R}^{n_x}} b_n(x, y_n) p(x_n|y_{1:n-1}) dx_n} \end{aligned}$$

e la (5.3)

$$\begin{aligned} \int_{\mathbb{R}^{n_x}} p(x_n|y_{1:n-1}) \varphi(x_n) dx_n &= \int_{\mathbb{R}^{n_x}} p(x_{n-1}|y_{1:n-1}) \cdot \\ &\quad \cdot \left[\int_{\mathbb{R}^{n_x}} K_n(x_{n-1}, x_n) \varphi(x_n) dx_n \right] dx_{n-1} \\ \int_{\mathbb{R}^{n_x}} p(x_n|y_{1:n}) \varphi(x_n) dx_n &= \frac{\int_{\mathbb{R}^{n_x}} b_n(x, y_n) p(x_n|y_{1:n-1}) \varphi(x_n) dx_n}{\int_{\mathbb{R}^{n_x}} b_n(x, y_n) p(x_n|y_{1:n-1}) dx_n}. \end{aligned}$$

5.3 Particle filter

In questa sezione verranno presentati degli algoritmi che comportano l' utilizzo di un sistema di M particles che evolvono secondo la distribuzione di un dato processo di Markov e, in istanti di tempo fissati, danno origine a un certo numero di discendenti. Verranno descritti diversi possibili meccanismi di ramificazione e si dimostrerà che la misura empirica associata ai sistemi di particles converge (al tendere di M a $+\infty$) alla distribuzione condizionata del segnale date le osservazioni.

5.3.1 Algoritmo

Ricordiamo la descrizione dell'algoritmo generale data nel Capitolo 2. Ad ogni istante n l'algoritmo genera M particles $\{x_n^{(i)}\}_{i=1}^M$ con una misura em-

pirica associata $\pi_{n|n}^M$

$$\pi_{n|n}^M := \frac{1}{M} \sum_{i=1}^M \delta_{X_n^{(i)}}$$

che é ‘vicina’ a $\pi_{n|n}$. L’algoritmo é ricorsivo nel senso che $\{x_n^{(i)}\}_{i=1}^M$ é determinato utilizzando le osservazioni ottenute al tempo n e l’insieme dei particles $\{x_{n-1}^{(i)}\}_{i=1}^M$ prodotto al tempo $n-1$ (la cui misura empirica $\pi_{n-1|n-1}^M$ é ‘vicina’ a $\pi_{n-1|n-1}$).

Il particle filter può essere descritto nel seguente modo.

Inizializzazione. Il filtro parte con una misura di probabilità empirica, indicata con $\pi_{0|0}^M$, associata a un campione di dimensione M da $\pi_{0|0}$. In altre parole, il filtro parte con M particles casuali di massa $1/M$, con posizioni $X_0^{(i)}$, $i = 1, 2, \dots, M$, tali che, per ogni i , $X_0^{(i)}$ ha distribuzione di probabilità $\pi_{0|0}$:

$$\pi_{0|0}^M := \frac{1}{M} \sum_{i=1}^M \delta_{X_0^{(i)}}$$

Iterazione. Descriviamo come ottenere $\pi_{n|n}^M$ da $\pi_{n-1|n-1}^M$.

L’approssimazione $\pi_{n-1|n-1}^M$ ha la forma $\frac{1}{M} \sum_{i=1}^M \delta_{X_{n-1}^{(i)}}$, cioè é la misura empirica associata a un sistema di M particles casuali. Il primo passo consiste nello spostare ogni particle usando il nucleo di transizione del segnale. Se $x_{n-1}^{(i)}$ é la posizione del particle i -esimo, allora la nuova posizione $\tilde{X}_n^{(i)}$ ha distribuzione di probabilità $K_n(x_{n-1}^{(i)}, \cdot)$ e i particles si spostano in modo mutuamente indipendente; $\tilde{\pi}_{n|n-1}^M$ sarà la distribuzione di probabilità empirica associata al nuovo sistema di particles:

$$\tilde{\pi}_{n|n-1}^M := \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{X}_n^{(i)}}$$

Ricordando le ricorsioni (5.2), si ha

$$\begin{aligned} \tilde{\pi}_{n|n}^M(dx_n) &= \frac{b_n(x_n, y_n) \tilde{\pi}_{n|n-1}^M(dx_n)}{\int_{\mathbb{R}^{n_x}} b_n(x_n, y_n) \tilde{\pi}_{n|n-1}^M(dx_n)} \\ &= \frac{\sum_{i=1}^M b_n(\tilde{x}_n^{(i)}, y_n) \delta_{\tilde{x}_n^{(i)}}(dx_n)}{\sum_{i=1}^M b_n(\tilde{x}_n^{(i)}, y_n)} \end{aligned}$$

cioé

$$\tilde{\pi}_{n|n}^M(dx_n) = \sum_{i=1}^M \tilde{\omega}_n^{(i)} \delta_{\tilde{x}_n^{(i)}}(dx_n), \quad \sum_{i=1}^M \omega_n^{(i)} = 1 \quad (5.4)$$

dove $\tilde{\omega}_n^{(i)} \propto b_n(\tilde{x}_n^{(i)}, y_n)$ sono i pesi di importanza. Lo scopo del passo di selezione é quello di ottenere una distribuzione empirica non pesata

$$\pi_{n|n}^M(dx_n) = \frac{1}{M} \sum_{i=1}^M \delta_{x_n^{(i)}}(dx_n)$$

moltiplicando i particles $\tilde{x}_n^{(i)}$ che hanno pesi di importanza elevati e scartando quelli con pesi bassi, in modo da focalizzarsi nelle zone con piú alta probabilità a posteriori. Ogni particle viene quindi sostituito da un numero di discendenti, indicato con $\xi_n(i)$, tale che il numero medio di discendenti sia $M\tilde{w}_n^{(i)}$, dove $\tilde{w}_n^{(i)}$ é il valore osservato della variabile casuale $\tilde{W}_n^{(i)}$, e di varianza finita. Si impone la condizione che il numero totale di particles non cambi, cioè che $\sum_{i=1}^M \xi_n(i) = M$. Le posizioni dei particle appena ottenute vengono indicate con $X_n^{(i)}, i = 1, \dots, M$ e si definisce $\pi_{n|n}^M$ nel seguente modo:

$$\pi_{n|n}^M(dx_n) = \frac{1}{M} \sum_{i=1}^M \delta_{x_n^{(i)}}(dx_n).$$

Sia A_n^M la matrice di covarianza del vettore casuale $\xi_n = (\xi_n^{(1)}, \dots, \xi_n^{(M)})'$, $A_n^M = E[(\xi_n - M\tilde{\omega}_n)(\xi_n - M\tilde{\omega}_n)']$ e sia $\tilde{\omega}_n = (\tilde{\omega}_n^{(1)}, \dots, \tilde{\omega}_n^{(M)})'$ il vettore dei pesi (e quindi $M\tilde{\omega}_n$ il vettore delle medie di ξ_n). Assumiamo che esista una costante c_n tale che

$$q' A_n^M q \leq M\bar{c}_n, \quad \forall q \in \mathbb{R}^{n_x}, q = (q^{(1)}, \dots, q^{(M)})', |q^{(i)}| \leq 1, i = 1, \dots, M. \quad (5.5)$$

Proposizione 5.3.1 *La condizione (5.5) é equivalente alla*

$$q' A_n^M q \leq M\bar{c}_n, \quad \forall q \in \mathbb{R}^{n_x}, q = (q^{(1)}, \dots, q^{(M)})', 0 \leq q^{(i)} \leq 1, i = 1, \dots, M \quad (5.6)$$

per una costante fissata \bar{c}_n .

Dimostrazione. La (5.5) implica immediatamente la (5.6), quindi dimostriamo l'implicazione inversa. Sia $q \in \mathbb{R}^{n_x}$ un vettore arbitrario tale

che $q = (q^{(1)}, \dots, q^{(M)})'$, $|q^{(i)}| \leq 1$, $i = 1, \dots, M$. Siano inoltre $q_+ = (q_+^{(1)}, \dots, q_+^{(M)})'$, $q_- = (q_-^{(1)}, \dots, q_-^{(M)})'$, $0 \leq q_+^{(i)}$, $q_-(i) \leq 1$, per $i = 1, \dots, M$, dove $q_+^{(i)} = \max(q^{(i)}, 0)$, $q_-^{(i)} = \max(-q^{(i)}, 0)$. Data una matrice A simmetrica semidefinita positiva, si indicherá con $\|\cdot\|_A$ la seminorma associata alla matrice A (se A é definita positiva, allora $\|\cdot\|_A$ é una vera norma), $\|q\|_A = \sqrt{q' A q}$. Allora $q = q_+ - q_-$ e, usando la disuguaglianza triangolare e la (5.6),

$$\|q\|_{A_n^M} \leq \|q_+\|_{A_n^M} + \|q_-\|_{A_n^M} \leq 2\sqrt{M\bar{c}_n}$$

e ciò implica che la (5.5) é soddisfatta con $c_n = 4\bar{c}_n$.

□

L'unica parte dell'algoritmo che resta da specificare é il meccanismo di ramificazione a cui sono associate le variabili casuali ξ_n . Nel prossimo paragrafo verranno presentati due meccanismi di ramificazione che mantengono costante il numero dei particles e soddisfano alla (5.5).

5.3.2 Meccanismi di ramificazione

In questo paragrafo vengono presentati due meccanismi di ramificazione: il *meccanismo di ramificazione multinomiale* e il *meccanismo di ramificazione basato su alberi*.

Ramificazione multinomiale

In questo caso si sceglie $\xi_n \sim \text{Multinomiale}(M, \tilde{\omega}_n^{(1)}, \dots, \tilde{\omega}_n^{(M)})$. Si ha che

$$\begin{aligned} E[\xi_n^{(i)}] &= M\tilde{\omega}_n^{(i)} \\ V[\xi_n^{(i)}] &= E[(\xi_n^{(i)} - M\tilde{\omega}_n^{(i)})^2] = M\tilde{\omega}_n^{(i)}(1 - \tilde{\omega}_n^{(i)}) \\ COV[\xi_n^{(i)}, \xi_n^{(j)}] &= E[(\xi_n^{(i)} - M\tilde{\omega}_n^{(i)})(\xi_n^{(j)} - M\tilde{\omega}_n^{(j)})] = -M\tilde{\omega}_n^{(i)}\tilde{\omega}_n^{(j)}. \end{aligned}$$

Per ogni $q \in \mathbb{R}^{n_x}$, $q = (q^{(1)}, \dots, q^{(M)})'$, $|q^{(i)}| \leq 1$, $i = 1, \dots, M$,

$$\begin{aligned} q' A_n^M q &= \sum_{i=1}^M M\tilde{\omega}_n^{(i)}(1 - \tilde{\omega}_n^{(i)})(q^{(i)})^2 - 2 \sum_{1 \leq i < j \leq M} M\tilde{\omega}_n^{(i)}\tilde{\omega}_n^{(j)}q^{(i)}q^{(j)} \\ &= M \sum_{i=1}^M \tilde{\omega}_n^{(i)}(q^{(i)})^2 - M \left(\sum_{i=1}^M \tilde{\omega}_n^{(i)}q^{(i)} \right)^2 \\ &\leq M \sum_{i=1}^M \tilde{\omega}_n^{(i)}. \end{aligned}$$

Dato che $\sum_{i=1}^M \tilde{\omega}_n^{(i)} = 1$, la (5.5) é soddisfatta con $c_n = 1$.

Ramificazione basata su alberi

Consideriamo un albero binario con M foglie e indichiamo con r la radice dell'albero (vedi Figura 5.1). Al particle i -esimo é associata l' i -esima foglia dell'albero, alla quale viene assegnato il corrispondente peso $\tilde{\omega}_i$. Sia m un nodo arbitrario dell'albero che non sia una foglia, siano m_1 e m_2 i suoi figli (discendenti diretti) e I_m (rispettivamente I_{m_1} e I_{m_2}) l'insieme delle foglie/particles che sono discendenti di m (rispettivamente di m_1 e m_2). Il peso $\tilde{\omega}_m$ é definito come la somma dei pesi di tutte le foglie sue discendenti:

$$\tilde{\omega}_m = \sum_{i \in I_m} \tilde{\omega}_i.$$

Si ha che $\tilde{\omega}_r = 1$, poiché I_r comprende l'intero insieme di foglie/particles e che $\tilde{\omega}_m = \tilde{\omega}_{m_1} + \tilde{\omega}_{m_2}$ poiché I_{m_1} e I_{m_2} rappresentano una partizione di I_m . Per ogni nodo dell'albero m , comprese le foglie, si costruirá un insieme di variabili casuali ξ_m tale che

$$\xi_m = \begin{cases} \lfloor M\tilde{\omega}_m \rfloor & \text{con probabilità } 1 - \{M\tilde{\omega}_m\}, \\ \lfloor M\tilde{\omega}_m \rfloor + 1 & \text{con probabilità } \{M\tilde{\omega}_m\} \end{cases} \quad (5.7)$$

Dove $\lfloor x \rfloor$ é la parte intera di $x \in \mathbb{R}$ e $\{x\} = x - \lfloor x \rfloor$. Le variabili casuali associate alle foglie costituiranno il meccanismo di ramificazione dei particles corrispondenti. La costruzione viene fatta ricorsivamente partendo dalla radice e terminando con le foglie in modo tale che la somma delle variabili casuali associate ai nodi figli sia uguale alla variabile casuale associata al nodo genitore.

Dato che $\tilde{\omega}_r = 1$, ξ_r é identicamente uguale a M . La regola di ricorsione é la seguente. Si suponga di aver costruito ξ_m con la distribuzione data dalla (5.7). Per prima cosa si assegnano i valori $\lfloor M\tilde{\omega}_{m_1} \rfloor$ al nodo m_1 e $\lfloor M\tilde{\omega}_{m_2} \rfloor$ al nodo m_2 ; si puó notare che $\xi_m - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor$ é uguale a 0, 1 o 2. Queste sono le uniche possibilitá poiché ξ_m puó assumere soltanto due valori, $\lfloor M\tilde{\omega}_m \rfloor$ o $\lfloor M\tilde{\omega}_m \rfloor + 1$ e $\lfloor M\tilde{\omega}_m \rfloor - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor$ puó essere uguale o a 0 o a 1. Se $\xi_m - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor = 2$ si incrementa di una unitá il valore assegnato a ciascuno dei due nodi, se $\xi_m - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor = 0$ non si modificano i valori giá assegnati e se $\xi_m - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor = 1$ si incrementa di una unitá il valore assegnato a uno dei due nodi, scelti casualmente, in modo tale che le variabili casuali risultanti abbiano la distribuzione desiderata. Piú precisamente

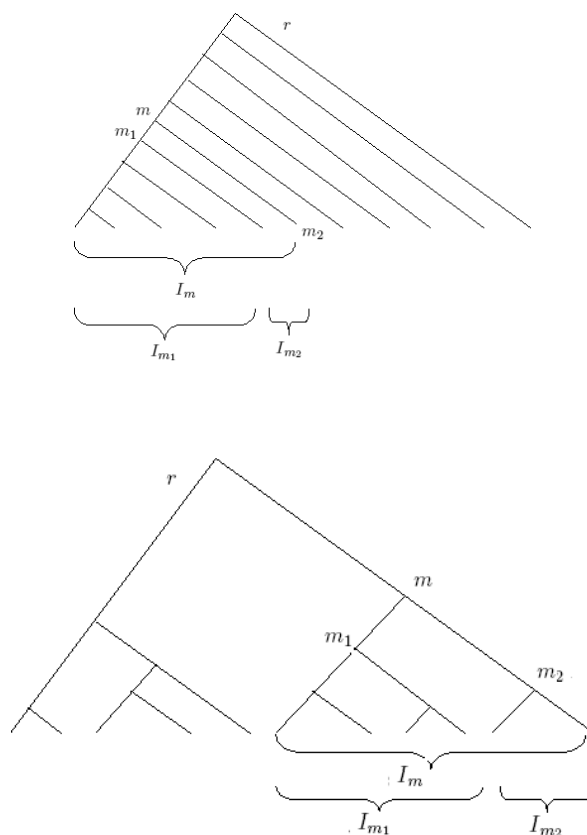


Figura 5.1: Alberi binari di ramificazione

- se $\lfloor M\tilde{\omega}_m \rfloor = \lfloor M\tilde{\omega}_{m_1} \rfloor + \lfloor M\tilde{\omega}_{m_2} \rfloor$, allora $\xi_m - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor$ é uguale a 0 (se $\xi_m = \lfloor M\tilde{\omega}_m \rfloor$) oppure 1 (se $\xi_m = \lfloor M\tilde{\omega}_m \rfloor + 1$). Si definisce

$$\begin{aligned}\xi_{m_1} &= \lfloor M\tilde{\omega}_{m_1} \rfloor + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor)\eta_m \\ \xi_{m_2} &= \lfloor M\tilde{\omega}_{m_2} \rfloor + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor)(1 - \eta_m),\end{aligned}$$

dove

$$\eta_m = \begin{cases} 0 & \text{con probabilità } \frac{\{M\tilde{\omega}_{m_2}\}}{\{M\tilde{\omega}_m\}}, \\ 1 & \text{con probabilità } \frac{\{M\tilde{\omega}_{m_1}\}}{\{M\tilde{\omega}_m\}} \end{cases}$$

La distribuzione della variabile casuale ausiliaria η_m é ben definita

poiché

$$\left. \begin{aligned} \tilde{\omega}_m &= \tilde{\omega}_{m_1} + \tilde{\omega}_{m_2} \\ \lfloor M\tilde{\omega}_m \rfloor &= \lfloor M\tilde{\omega}_{m_1} \rfloor + \lfloor M\tilde{\omega}_{m_2} \rfloor \end{aligned} \right\} \Rightarrow \frac{\{M\tilde{\omega}_{m_1}\}}{\{M\tilde{\omega}_m\}} + \frac{\{M\tilde{\omega}_{m_2}\}}{\{M\tilde{\omega}_m\}} = 1.$$

- Se $\lfloor M\tilde{\omega}_m \rfloor = \lfloor M\tilde{\omega}_{m_1} \rfloor + \lfloor M\tilde{\omega}_{m_2} \rfloor + 1$, allora $\xi_m - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor$ é uguale a 1 (se $\xi_m = \lfloor M\tilde{\omega}_m \rfloor$) oppure 2 (se $\xi_m = \lfloor M\tilde{\omega}_m \rfloor + 1$). Si definisce

$$\begin{aligned} \xi_{m_1} &= \lfloor M\tilde{\omega}_{m_1} \rfloor + 1 + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor - 1)\eta_m \\ \xi_{m_2} &= \lfloor M\tilde{\omega}_{m_2} \rfloor + 1 + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor - 1)(1 - \eta_m), \end{aligned}$$

dove

$$\eta_m = \begin{cases} 0 & \text{con probabilità } \frac{1 - \{M\tilde{\omega}_{m_2}\}}{1 - \{M\tilde{\omega}_m\}}, \\ 1 & \text{con probabilità } \frac{1 - \{M\tilde{\omega}_{m_1}\}}{1 - \{M\tilde{\omega}_m\}} \end{cases}$$

La distribuzione della variabile casuale ausiliaria η_m é ben definita poiché

$$\left. \begin{aligned} \tilde{\omega}_m &= \tilde{\omega}_{m_1} + \tilde{\omega}_{m_2} \\ \lfloor M\tilde{\omega}_m \rfloor &= \lfloor M\tilde{\omega}_{m_1} \rfloor + \lfloor M\tilde{\omega}_{m_2} \rfloor + 1 \end{aligned} \right\} \Rightarrow \frac{1 - \{M\tilde{\omega}_{m_1}\}}{1 - \{M\tilde{\omega}_m\}} + \frac{1 - \{M\tilde{\omega}_{m_2}\}}{1 - \{M\tilde{\omega}_m\}} = 1.$$

Si prendono le variabili casuali η_m mutuamente indipendenti. In entrambi i casi

$$\begin{aligned} \xi_{m_1} &= \lfloor M\tilde{\omega}_{m_1} \rfloor + u_m + (\xi_m - v_m)\eta_m \\ \xi_{m_2} &= \lfloor M\tilde{\omega}_{m_2} \rfloor + u_m + (\xi_m - v_m)(1 - \eta_m), \end{aligned}$$

dove $u_m = \lfloor M\tilde{\omega}_m \rfloor - \lfloor M\tilde{\omega}_{m_1} \rfloor - \lfloor M\tilde{\omega}_{m_2} \rfloor = \{M\tilde{\omega}_{m_1}\} + \{M\tilde{\omega}_{m_2}\} - \{M\tilde{\omega}_m\}$ e $v_m = \lfloor M\tilde{\omega}_m \rfloor + u_m$.

Siano $\xi_i, i = 1, \dots, M$, le variabili casuali risultanti corrispondenti alle foglie dell'albero, e quindi ai particles del sistema. Dalla (5.7) si nota che

$$\begin{aligned} E[\xi_i] &= \lfloor M\tilde{\omega}_i \rfloor (1 - \{M\tilde{\omega}_i\}) + (\lfloor M\tilde{\omega}_i \rfloor + 1)\{M\tilde{\omega}_i\} = M\tilde{\omega}_i \\ V[\xi_i] &= E[\xi_i^2] - E^2[\xi_i] = \lfloor M\tilde{\omega}_i \rfloor^2 (1 - \{M\tilde{\omega}_i\}) + (\lfloor M\tilde{\omega}_i \rfloor + 1)^2 \{M\tilde{\omega}_i\} - M^2 \tilde{\omega}_i^2 \\ &= \lfloor M\tilde{\omega}_i \rfloor^2 + \{M\tilde{\omega}_i\} + 2 \lfloor M\tilde{\omega}_i \rfloor \{M\tilde{\omega}_i\} - M^2 \tilde{\omega}_i^2 \\ &= \lfloor M\tilde{\omega}_i \rfloor^2 + \{M\tilde{\omega}_i\} + 2M\tilde{\omega}_i \lfloor M\tilde{\omega}_i \rfloor - 2 \lfloor M\tilde{\omega}_i \rfloor^2 - M^2 \tilde{\omega}_i^2 \\ &= -\lfloor M\tilde{\omega}_i \rfloor^2 + \{M\tilde{\omega}_i\} + 2M\tilde{\omega}_i \lfloor M\tilde{\omega}_i \rfloor - M^2 \tilde{\omega}_i^2 \\ &= -(M\tilde{\omega}_i - \lfloor M\tilde{\omega}_i \rfloor)^2 + \{M\tilde{\omega}_i\} \\ &= -\{M\tilde{\omega}_i\}^2 + \{M\tilde{\omega}_i\} = \{M\tilde{\omega}_i\}(1 - \{M\tilde{\omega}_i\}) \leq \frac{1}{4}. \end{aligned}$$

Inoltre

$$\begin{aligned}
E[\xi_{m_1} - E[\xi_{m_1}] | \xi_m] &= E[\lfloor M\tilde{\omega}_{m_1} \rfloor + u_m + (\xi_m - v_m)\eta_m - M\tilde{\omega}_{m_1} | \xi_m] \\
&= E[u_m - \{M\tilde{\omega}_{m_1}\} + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor - u_m)\eta_m | \xi_m] \\
&= \frac{u_m - \{M\tilde{\omega}_{m_1}\}}{u_m - \{M\tilde{\omega}_m\}} (\xi_m - M\tilde{\omega}_m)
\end{aligned}$$

e

$$\begin{aligned}
COV[\xi_{m_1}, \xi_{m_2} | \xi_m] &= E[(\xi_{m_1} - E[\xi_{m_1}])(\xi_{m_2} - E[\xi_{m_2}]) | \xi_m] \\
&= E[(u_m - \{M\tilde{\omega}_{m_1}\} + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor - u_m)\eta_m)(u_m \\
&\quad - \{M\tilde{\omega}_{m_2}\} + (\xi_m - \lfloor M\tilde{\omega}_m \rfloor - u_m)(1 - \eta_m)) | \xi_m] \\
&= (u_m - \{M\tilde{\omega}_{m_1}\})(u_m - \{M\tilde{\omega}_{m_2}\}) \left(\frac{2(\xi_m - M\tilde{\omega}_m)}{u_m - \{M\tilde{\omega}_m\}} - 1 \right).
\end{aligned}$$

Quindi, se i e j sono due particles o foglie distinti, allora

$$\begin{aligned}
COV[\xi_i, \xi_j] &= - \prod_{s=1}^{h_i-1} \frac{u_{m_s} - \{M\tilde{\omega}_{m_{s-1}}\}}{u_{m_s} - \{M\tilde{\omega}_{m_s}\}} \prod_{t=1}^{h_j-1} \frac{u_{n_t} - \{M\tilde{\omega}_{n_{t-1}}\}}{u_{n_t} - \{M\tilde{\omega}_{n_t}\}} (u_{m_{n_i}} \\
&\quad - \{M\tilde{\omega}_{m_{h_i-1}}\})(u_{n_{h_j}} - \{M\tilde{\omega}_{m_{h_j-1}}\}),
\end{aligned}$$

dove $m_{h_i} = n_{h_j}$ é il piú recente antenato comune di i e j , h_i e h_j sono i numeri di generazioni dal piú recente antenato comune a i e, rispettivamente, j , m_1, m_2, \dots, m_{h_i} é la linea genealogica di i e n_1, n_1, \dots, n_{h_j} é la linea genealogica di j . Poiché valgono le disuguaglianze $V[\xi_i] \leq 1/4$ e $COV[\xi_i, \xi_j] \leq 0$, si ha che $q'A_n^M q \leq 1/4M$, per ogni $q \in \mathbb{R}^{n_x}$, $q = (q^{(1)}, \dots, q^{(M)})'$, $0 \leq q^{(i)} \leq 1$, $i = 1, \dots, M$. Quindi la condizione (5.5) é soddisfatta con $c_n = 1$.

L'algoritmo appena descritto é molto intuitivo e facile da usare. Nei prossimi paragrafi si mostrerá che tale algoritmo produce una approssimazione che converge al 'vero' filtro ottimo sotto assunzioni minime.

5.4 Teoremi sulla convergenza

Assumiamo che il processo delle osservazioni sia fissato. Prima di analizzare la convergenza dell'algoritmo descritto precedentemente, presentiamo alcune osservazioni preliminari che ci aiuteranno a capire perché converge e quali sono le condizioni che si devono imporre.

5.4.1 Considerazioni preliminari

Cominciamo la trattazione della convergenza con una formulazione astratta e successivamente identificheremo gli elementi che riguardano il problema di filtraggio. Sia (E, d) uno spazio metrico, su questo spazio siano $(\alpha_n)_{n=1}^\infty$ e $(\beta_n)_{n=1}^\infty$ due sequenze di funzioni continue, $\alpha_n, \beta_n: E \rightarrow E$. Inoltre, siano k_n e $k_{1:n}$ altre due sequenze di funzioni definite da

$$k_n := \alpha_n \circ \beta_n, \quad k_{1:n} := k_n \circ k_{n-1} \circ \dots \circ k_1$$

dove l'operazione 'o' denota la composizione di funzioni. Ovviamente, sia k_n che $k_{1:n}$ sono continue.

Nel contesto del filtraggio stocastico, lo spazio E sarà $\mathcal{P}(\mathbb{R}^{n_x})$, lo spazio di tutte le misure di probabilità sullo spazio euclideo \mathbb{R}^{n_x} , β_n sarà la funzione che mappa $\pi_{n-1|n-1}$ in $\pi_{n|n-1}$ e α_n sarà la funzione che mappa $\pi_{n|n-1}$ in $\pi_{n|n}$. Per definizione, $k_{1:n}$ sarà la trasformazione $\pi_{0|0} \rightarrow \pi_{n|n}$.

Perturbiamo k_n e $k_{1:n}$ usando una funzione c^M , non necessariamente continua, $c^M: E \rightarrow E$ nel seguente modo

$$k_n^M = c^M \circ \alpha_n \circ c^M \circ \beta_n, \quad k_{1:n}^M := k_n^M \circ k_{n-1}^M \circ \dots \circ k_1^M.$$

Nel contesto del filtraggio stocastico, c^M sarà la funzione che mappa una misura in un campione casuale di dimensione M della misura. Successivamente assumeremo che se M cresce, la perturbazione diventa sempre più piccola. In altre parole, assumiamo che c^M converge a i , cioè alla funzione identità su E ($i(e) = e$, per ogni $e \in E$). A questo punto risulta naturale chiedersi se k_n^M converge a k_n e se $k_{1:n}^M$ converge a $k_{1:n}$.

Esempio: Siano $E = [0, 1]$ e d l'usuale metrica su $[0, 1]$, $d(a, b) = |a - b|$. Siano α_n e β_n uguali alla funzione identità su E . Dunque anche k_n è la funzione identità. Modifichiamo k_n usando la seguente funzione continua a tratti

$$c^M(e) := \begin{cases} e + \frac{e}{M}, & \text{se } e \in [0, \frac{1}{2}] \\ 1 - (M-1) \left| \frac{1}{2} + \frac{1}{2M} - e \right|, & \text{se } e \in \left(\frac{1}{2}, \frac{1}{2} + \frac{1}{M} \right) \\ e + \frac{e-1}{M-2}, & \text{se } e \in \left(\frac{1}{2} + \frac{1}{M}, 1 \right]. \end{cases}$$

È facile vedere che sebbene $\lim_{M \rightarrow \infty} c^M(e) = e$ per ogni $e \in [0, 1]$, si ha

$$\lim_{M \rightarrow \infty} k_n^M(1/2) = \lim_{M \rightarrow \infty} c^M(c^M(1/2))$$

$$\begin{aligned}
&= \lim_{M \rightarrow \infty} c^M \left(\frac{1}{2} + \frac{1}{2M} \right) \\
&= 1 \neq \frac{1}{2} = k_n(1/2).
\end{aligned}$$

Dunque, non é vero che successive piccole perturbazioni di una funzione portano ad una piccola perturbazione. Per avere che $\lim_{M \rightarrow \infty} k_{1:n}^M = k_{1:n}$ serve un tipo di convergenza piú forte per c^M . Chiediamo che, per ogni e^M , $e \in E$ tale che $\lim_{M \rightarrow \infty} e^M = e$

$$\lim_{M \rightarrow \infty} c^M(e^M) = e \quad (5.8)$$

La condizione (5.8) é equivalente alla

$$\lim_{M \rightarrow \infty} d(c^M(e^M), e^M) = 0. \quad (5.9)$$

Vale il seguente lemma.

Lemma 5.4.1 *Siano $\alpha_n, \beta_n, k_n, k_{1:n}$ e c^M definite come sopra. Allora, se c^M soddisfa alla (5.8), si ha*

$$\lim_{M \rightarrow \infty} k_n^M = k_n \quad e \quad \lim_{M \rightarrow \infty} k_{1:n}^M = k_{1:n}. \quad (5.10)$$

Inoltre, k_n^M e $k_{1:n}^M$ soddisfano a

$$\begin{aligned}
\lim_{M \rightarrow \infty} e^M = e \quad \Rightarrow \quad \lim_{M \rightarrow \infty} k_n^M(e^M) = k_n(e) \\
\lim_{M \rightarrow \infty} k_{1:n}^M(e^M) = k_{1:n}(e).
\end{aligned} \quad (5.11)$$

Dimostrazione: La condizione (5.11) implica la (5.10) (basta prendere $e^M = e$ per ogni M nella (5.11)). Resta quindi da dimostrare soltanto la (5.11). Dato che β_n é continua, si ha

$$\lim_{M \rightarrow \infty} e^M = e \Rightarrow \lim_{M \rightarrow \infty} b_n(e^M) = b_n(e). \quad (5.12)$$

Usando la (5.8),

$$\lim_{M \rightarrow \infty} b_n(e^M) = b_n(e) \Rightarrow \lim_{M \rightarrow \infty} c^M(b_n(e^M)) = b_n(e) \quad (5.13)$$

e, dato che α_n é continua

$$\begin{aligned} \lim_{M \rightarrow \infty} c^M(b_n(e^M)) &= b_n(e) \\ \Rightarrow \lim_{M \rightarrow \infty} a_n(c^M(b_n(e^M))) &= a_n(b_n(e)) \end{aligned} \quad (5.14)$$

e, usando ancora la (5.8),

$$\begin{aligned} \lim_{M \rightarrow \infty} a_n(c^M(b_n(e^M))) &= a_n(b_n(e)) \\ \Rightarrow \lim_{M \rightarrow \infty} c^M(a_n(c^M(b_n(e^M)))) &= a_n(b_n(e)). \end{aligned} \quad (5.15)$$

Infine, mettendo assieme le condizioni (5.12)-(5.15), si ha che

$$\lim_{M \rightarrow \infty} k_n^M(e^M) = k_n(e),$$

che a sua volta implica, per induzione su n , che

$$\lim_{M \rightarrow \infty} k_{1:n}^M(e^M) = k_{1:n}(e).$$

□

5.4.2 Applicazione al Filtro Ottimo

In questo paragrafo, applicheremo i risultati precedenti al problema di filtraggio stocastico. La convergenza dell'algorithmo particle filter deriverá direttamente dal Lemma 5.4.1.

Sia $E = \mathcal{P}(\mathbb{R}^{n_x})$ l'insieme delle misure di probabilitá sullo spazio euclideo \mathbb{R}^{n_x} . Ricordiamo che, data una sequenza di misure di probabilitá $(\mu_M)_{M=1}^{\infty}$, μ_M converge (debolmente) a $\mu \in \mathcal{P}(\mathbb{R}^{n_x})$ e scriviamo $\lim_{M \rightarrow \infty} \mu_M = \mu$ se, per ogni $\varphi \in \mathcal{C}_b(\mathbb{R}^{n_x})$

$$\lim_{M \rightarrow \infty} (\mu_M, \varphi) = (\mu, \varphi).$$

Definiamo $\beta_n : \mathcal{P}(\mathbb{R}^{n_x}) \rightarrow \mathcal{P}(\mathbb{R}^{n_x})$ come

$$b_n(\nu)(dx_n) := \int_{(\mathbb{R}^{n_x})} \int_{(\mathbb{R}^{n_x})} \varphi(x_n) K_n(x_{n-1}, dx_n) \nu(dx_{n-1}) = (\nu, K_n \varphi).$$

Si ha dunque

$$\pi_{n|n-1} = \beta_n(\pi_{n-1|n-1}). \quad (5.16)$$

Vogliamo assicurarci che β_n sia continua. Nel contesto del filtraggio, significa semplicemente che il segnale si muove in modo continuo e che due realizzazioni del segnale che partono da posizioni 'vicine' rimarranno 'vicine' al

tempo successivo. Per assicurare che questo accada, richiediamo che il nucleo di transizione del segnale sia di Feller¹, cioè che goda della proprietà

$$\forall \varphi \in \mathcal{C}_b(\mathbb{R}^{n_x}) \Rightarrow K_n \varphi \in \mathcal{C}_b(\mathbb{R}^{n_x}). \quad (5.17)$$

Se $\lim_{M \rightarrow \infty} \nu_M = \nu$, allora, per definizione, $\lim_{M \rightarrow \infty} (\mu_M, \varphi) = (\mu, \varphi)$ per ogni $\varphi \in \mathcal{C}_b(\mathbb{R}^{n_x})$. Dunque

$$\lim_{M \rightarrow \infty} (\mu_M, K_n \varphi) = (\mu, K_n \varphi), \quad \forall \varphi \in \mathcal{C}_b(\mathbb{R}^{n_x})$$

e

$$\begin{aligned} \lim_{M \rightarrow \infty} (\beta_n(\nu_M), \varphi) &= \lim_{M \rightarrow \infty} (\nu_M, K_n \varphi) \\ &= (\nu, K_n \varphi) = (\beta_n(\nu), \varphi), \quad \forall \varphi \in \mathcal{C}_b(\mathbb{R}^{n_x}). \end{aligned}$$

Definiamo ora l'applicazione α_n . Sia $\alpha_n : \mathcal{P}(\mathbb{R}^{n_x}) \rightarrow \mathcal{P}(\mathbb{R}^{n_x})$ la funzione che, per ogni $\nu \in \mathcal{P}(\mathbb{R}^{n_x})$, $\alpha_n(\nu)$ è definita da

$$(\alpha_n(\nu), \varphi) = (\nu, b_n)^{-1}(\nu, \varphi b_n), \quad \forall \varphi \in \mathcal{C}_b(\mathbb{R}^{n_x}). \quad (5.18)$$

Allora

$$\pi_{n|n} = \alpha_n(\pi_{n|n-1}) = \alpha_n \circ \beta_n(\pi_{n-1|n-1}). \quad (5.19)$$

Ancora, nel contesto del filtraggio, è naturale assumere che α_n sia continua. Questo significa che una piccola variazione nella distribuzione condizionata (iniziale) del segnale X_n non indurrà una grande variazione nella distribuzione condizionata del segnale quando viene presa in considerazione la nuova osservazione y_n .

Un modo per assicurare che ciò accada è richiedere che $b_n(\cdot, y_n)$ sia una funzione continua, limitata e strettamente positiva

$$b_n(\cdot, y_n) \in \mathcal{C}_b(\mathbb{R}^{n_x}), \quad b_n(x_n, y_n) > 0, \quad \forall x_n \in \mathbb{R}^{n_x}. \quad (5.20)$$

L'assunzione di positività è necessaria per assicurare che (ν, b_n) non si annulli mai e che quindi sia permessa la divisione nella (5.18). Inoltre, se $b_n(\cdot, y_n)$ soddisfa alla (5.20), allora per la (5.18), si ha che $\lim_{M \rightarrow \infty} \nu_M = \nu$ implica

$$\lim_{M \rightarrow \infty} (\alpha_n(\nu_M), \varphi) = \frac{\lim_{M \rightarrow \infty} (\nu_M, \varphi b_n)}{\lim_{M \rightarrow \infty} (\nu_M, b_n)} = \frac{(\nu, \varphi b_n)}{(\nu, b_n)} = (\alpha_n(\nu), \varphi)$$

¹In realtà questa proprietà dovrebbe essere definita come proprietà di Feller debole, ma non facendo uso in questa tesi della proprietà di Feller forte, la chiameremo così per semplicità

per ogni funzione $\varphi \in \mathcal{C}_b(\mathbb{R}^{n_x})$. Dunque

$$\lim_{M \rightarrow \infty} \alpha_n(\nu_M) = \alpha_n(\nu),$$

e dunque α_n é continua. Ovviamente, se α_n e β_n sono continue, lo sono anche k_n e $k_{1:n}$ e

$$\pi_n|_n = k_n(\pi_{n-1}|_{n-1}) = k_{1:n}(\mu). \quad (5.21)$$

Nel contesto del particle filter, la perturbazione c^M sará casuale. Tuttavia, con probabilitá 1, avrá ancora le proprietá richieste precedentemente. Sia $c^{M,\omega}$, $M > 0$, $\omega \in \Omega$ la seguente perturbazione casuale. Per ogni $\nu \in \mathcal{P}(\mathbb{R}^{n_x})$

$$c^{M,\omega}(\nu) := \frac{1}{M} \sum_{j=1}^M \delta_{\{V_j(\omega)\}} \quad (5.22)$$

dove $V_j : \Omega \rightarrow \mathbb{R}^d$ sono variabili casuali i.i.d. con distribuzione comune ν .

Lemma 5.4.2 *Se $c^{M,\omega}$ é definito come in (5.22), allora, per quasi ogni $\omega \in \Omega$, $c^{M,\omega}$ soddisfa alla (5.8).*

Dimostrazione: Sia $\nu_M, \nu \in \mathcal{P}(\mathbb{R}^{n_x})$ tali che $\lim_{M \rightarrow \infty} \nu_M = \nu$. $\forall \varphi \in \mathcal{M}$ (\mathcal{M} insieme che determina la convergenza), usando l'indipendenza della sequenza (V_j) si ha che

$$\begin{aligned} E \left[\left((c^{M,\cdot}(\nu_M), \varphi_i) - (\nu_M, \varphi_i) \right)^4 \right] &= E \left[\left(\frac{1}{M} \sum_{j=1}^M \varphi_i(V_j) - (\nu_M, \varphi_i) \right)^4 \right] \\ &= \frac{1}{M^4} \sum_{j=1}^M E[(\varphi_i(V_j) - (\nu_M, \varphi_i))^4] \\ &\quad + \frac{2}{M^4} \sum_{1 \leq j_1 < j_2 \leq M} E[(\varphi_i(V_{j_1}) - (\nu_M, \varphi_i))^2 (\varphi_i(V_{j_2}) - (\nu_M, \varphi_i))^2] \\ &\leq \frac{2^4 \|\varphi_i\|^4 M + 2^4 \|\varphi_i\|^4 M(M-1)}{M^4} = \frac{2^4 \|\varphi_i\|^4 (M^2)}{M^4} = \frac{2^4 \|\varphi_i\|^4}{M^2} \end{aligned}$$

Segue che

$$E \left[\sum_{M=1}^{\infty} \left((c^{M,\cdot}(\nu_M), \varphi_i) - (\nu_M, \varphi_i) \right)^4 \right] \leq 16 \|\varphi\|^4 \sum_{M=1}^{\infty} \frac{1}{M^2} < \infty$$

e dunque

$$\sum_{M=1}^{\infty} ((c^{M,\cdot}(\nu_M), \varphi_i) - (\nu_M, \varphi_i))^4 < \infty$$

per quasi ogni $\omega \in \Omega$. Ciò implica che

$$\lim_{M \rightarrow \infty} |(c^{M,\omega}(\nu_M), \varphi_i) - (\nu_M, \varphi_i)| = 0$$

per quasi ogni $\omega \in \Omega$.

Così, esiste un sottoinsieme $\bar{\Omega} \subset \Omega$, $P(\bar{\Omega}) = 1$ tale che

$$\forall \omega \in \bar{\Omega}, \forall i \in I, \quad \lim_{M \rightarrow \infty} |(c^{M,\omega}(\nu_M), \varphi_i) - (\nu_M, \varphi_i)| = 0$$

che implica che, per ogni $\omega \in \bar{\Omega}$

$$\lim_{M \rightarrow \infty} d(c^{M,\omega}(\nu_M), \nu_M) = 0$$

e dunque, $c^{M,\omega}$ soddisfa alla (5.8) per ogni $\omega \in \bar{\Omega}$.

□

Consideriamo ora $\pi_{n|n}^M$, che é la misura empirica associata all'insieme dei particles ottenuti alla fine del passo di ricampionamento nel filtro bootstrap. Si ha che

$$\begin{aligned} \pi_{n|n}^M &= c^M \circ \alpha_n \circ c^M \circ \beta_n(\pi_{n-1|n-1}^M) = k_n^M(\pi_{n-1|n-1}^M) \\ \pi_{n|n}^M &= k_{1:n}^M \circ c^M(\mu) = k_{1:n}^M(\mu^M) \end{aligned}$$

dove $\mu^M = c^M(\mu)$. Inoltre, osserviamo che $\tilde{\pi}_{n|n-1}^M = c^M \circ \beta_n(\pi_{n-1|n-1}^M)$.

Teorema 5.4.3 *Assumiamo che il nucleo di transizione K_n sia di Feller e che la funzione di verosimiglianza g_n sia continua, limitata e strettamente positiva. Allora, quasi certamente*

$$\lim_{M \rightarrow \infty} \pi_{n|n}^M = \pi_{n|n}.$$

Dimostrazione: Questo risultato segue dal Lemma 5.4.1 e dalla (5.21) dato che $\lim_{M \rightarrow \infty} \mu^M = \mu$; allora

$$\lim_{M \rightarrow \infty} \pi_{n|n}^M = \lim_{M \rightarrow \infty} k_{1:n}^M(\mu^M) = k_{1:n}(\mu) = \pi_{n|n}.$$

□

Capitolo 6

Convergenza nel modello per media e varianza

In questo capitolo si applicheranno i risultati sulla convergenza ottenuti nel Capitolo 5 al modello per media e varianza stocastica presentato nel Capitolo 3.

$$\begin{aligned}(X_0, Z_0) & \quad \text{vettore casuale dato} \\ X_{n+1} & = X_n + \beta \left(Z_{n+1} - \frac{Z_n}{2} \right) + \sqrt{Z_{n+1} - \frac{Z_n}{2}} V_{n+1}, \\ Z_{n+1} & = Z_n + W_{n+1}, \\ Y_n & = X_n + \sqrt{Z_n} V'_n, \quad n \in \mathbb{N}\end{aligned}$$

con $\beta \in \mathbb{R}$ e $\{V_{n+1}\}_{n \in \mathbb{N}}$, $\{V'_n\}_{n \in \mathbb{N}}$ e $\{W_{n+1}\}_{n \in \mathbb{N}}$ tre sequenze indipendenti di variabili casuali indipendenti. Assumiamo che:

H.1 $\{V_{n+1}\}_{n \in \mathbb{N}}$ e $\{V'_n\}_{n \in \mathbb{N}}$ sono sequenze di variabili casuali gaussiane standard indipendenti

H.2 $\{W_{n+1}\}_{n \in \mathbb{N}}$ é una sequenza di variabili gamma indipendenti con parametro di forma $|\lambda_n|$ e parametro di scala $\gamma_n^2/2$, dove

$$\lambda_0 = \lambda - \frac{1}{2}, \quad \lambda_n = |\lambda_{n-1}| - \frac{1}{2}, \quad n \geq 1,$$

$$\gamma_0 = \sqrt{\gamma^2 + \frac{\beta^2}{2}}, \quad \gamma_n = \sqrt{\gamma_{n-1}^2 + \frac{\beta^2}{2}}, \quad n \geq 1$$

H.3 $(X_0, Z_0) \sim N(\mu + \beta Z_0, Z_0) \times GIG(\lambda, \delta, \gamma)$ con $\lambda \in [0, \frac{1}{2})$; (X_0, Z_0) indipendente dai processi $\{V_{n+1}\}_{n \in \mathbb{N}}, \{V'_n\}_{n \in \mathbb{N}}$ e $\{W_{n+1}\}_{n \in \mathbb{N}}$.

Perché l'algoritmo particle filter converga (debolmente) al filtro esatto, devono valere le ipotesi del Teorema (5.4.3). Se tali ipotesi sono soddisfatte nel modello sopra riportato, si ha che, quasi certamente,

$$\lim_{M \rightarrow \infty} \pi_{n|n}^M = \pi_{n|n},$$

dove $\pi_{n|n}^M(dx_n) = \frac{1}{M} \sum_{i=1}^M \delta_{x_n^{(i)}}(dx_n)$ é la misura empirica associata all'insieme di particles generati dell'algoritmo e $\pi_{n|n}$ é la distribuzione di filtraggio esatta.

6.1 Proprietá di Feller

La prima ipotesi del Teorema 5.4.3 riguarda la proprietá di Feller.

Tale proprietá é spesso definita richiedendo che il nucleo di transizione K_n mappi $\mathcal{C}_b(X)$ in $\mathcal{C}_b(X)$. Per ogni $\varphi \in \mathcal{C}_b(X)$ si ha che

$$K_n \varphi(x) = \int K_n(x, dy) \varphi(y) \in \mathcal{C}_b(X), \quad x \in X$$

6.1.1 Modello state space non lineare

Uno dei piú semplici esempi di catena di Feller riguarda il modello state space non lineare. Sia $\mathbf{X} = \{X_n\}$, dove

1. per ogni $n \geq 0$ X_n e W_n sono, rispettivamente, variabili casuali su \mathbb{R}^n , \mathbb{R}^p soddisfano per $n \geq 1$ alla formula ricorsiva

$$X_n = F(X_{n-1}, W_n) \tag{6.1}$$

dove $F : X \times O_w \rightarrow X$ é C^∞ e X é un aperto di \mathbb{R}^n e O_w é un sottoinsieme aperto di \mathbb{R}^p

2. le variabili W_n sono una sequenza di rumore con densitá γ

In queste ipotesi, vale il seguente lemma.

Lemma 6.1.1 *Il modello state space non lineare sopra descritto gode sempre della proprietá di Feller.*

Dimostrazione: Per definizione la funzione $x \rightarrow F(x, w)$ é continua per ogni $w \in \mathbb{R}$ fissato. Se $\varphi : X \rightarrow \mathbb{R}$ é continua e limitata, $\varphi \circ F(x, w)$ é essa stessa continua e limitata per ogni $w \in \mathbb{R}$ fissato. Per il teorema della Convergenza limitata

$$K_n \varphi(x) = E[h(F(x, W))] = \int \Gamma(dw) h \circ F(x, w)$$

é una funzione continua di $x \in X$.

□

Questo lemma ci permette di affermare che il modello state space non lineare con le proprietá riportate gode della proprietá di Feller, cioè il segnale si muove in modo continuo e due realizzazioni del segnale che partono da posizioni ‘vicine’ rimarranno ‘vicine’ al tempo successivo.

Dato il modello

$$\begin{aligned} X_n &= X_{n-1} + \beta(Z_n - Z_{n-1}/2) + \sqrt{Z_n - Z_{n-1}/2} V_n \\ Z_n &= Z_{n-1} + W_n \\ Y_n &= X_n + \sqrt{Z_n} V'_n, \end{aligned}$$

si ha

$$\begin{aligned} X_n &= X_{n-1} + \beta(Z_{n-1}/2 + W_n) + \sqrt{Z_{n-1}/2 + W_n} V_n \\ Z_n &= Z_{n-1} + W_n \\ Y_n &= X_n + \sqrt{Z_n} V'_n. \end{aligned}$$

Quindi la sequenza $\{X_n, Z_n\}$ evolve nel seguente modo

$$\begin{pmatrix} X_n \\ Z_n \end{pmatrix} = \begin{pmatrix} X_{n-1} + \beta(Z_{n-1}/2 + W_n) + \sqrt{Z_{n-1}/2 + W_n} V_n \\ Z_{n-1} + W_n \end{pmatrix} \quad (6.2)$$

ed é rappresentabile nella forma (6.1) con

$$\begin{aligned}
F(x, z, w, v) &= \\
&= \left(x + \beta \left(z/2 + w \right) + \sqrt{z/2 + wv}, z + w \right).
\end{aligned}$$

L'applicazione $F : \mathbb{R} \times (0, +\infty) \times (0, +\infty) \times \mathbb{R} \rightarrow \mathbb{R} \times (0, +\infty)$, così definita, soddisfa chiaramente alle ipotesi del Lemma 6.1.1.

Inoltre le variabili (W_n, V_n) sono indipendenti: il vettore ha distribuzione $\text{Gamma}(|\lambda_n|, \frac{\gamma_n^2}{2}) \times N(0, 1)$ Valgono quindi le ipotesi del Lemma 6.1.1 e quindi il nucleo di transizione del modello analizzato gode della proprietà di Feller.

La prima ipotesi del Teorema 5.4.3 é dunque soddisfatta nel modello preso in esame.

6.2 Condizioni sulla funzione di verosimiglianza

La seconda ipotesi per la validità del Teorema 5.4.3 pone delle condizioni sulla funzione di verosimiglianza $b_n(x_n, y_n)$, in particolare richiede che sia continua, limitata e strettamente positiva

$$b_n(\cdot, y_n) \in \mathcal{C}_b(\mathbb{R}^{n_x}), \quad b_n(x_n, y_n) > 0, \quad \forall x_n \in \mathbb{R}^{n_x}.$$

Nel nostro caso, deve essere

$$b_n(\cdot, y_n) \in \mathcal{C}_b(\mathbb{R} \times (0, +\infty)), \quad b_n(x_n, z_n, y_n) > 0, \quad \forall x_n, z_n \in \mathbb{R} \times (0, +\infty),$$

dove

$$b_n(x_n, z_n, y_n) = \frac{1}{\sqrt{2\pi z_n}} \exp \left[-\frac{1}{2z_n} (y_n - x_n)^2 \right].$$

La funzione risulta strettamente positiva $\forall x_n, z_n \in \mathbb{R} \times (0, +\infty)$, dato che si tratta di un rapporto tra una funzione esponenziale e una radice.

Tuttavia la funzione non risulta limitata. Infatti, sezionando la funzione $b_n(\cdot, y_n)$ nel piano $x_n = y_n$, si ha che $\lim_{z_n \rightarrow 0^+} b_n(y_n, z_n, y_n) = +\infty$. La seconda ipotesi del Teorema 5.4.3 non é soddisfatta nel modello analizzato.

Non siamo quindi in grado, nelle ipotesi attuali, di assicurare la convergenza (debole) dell'algoritmo particle filter alla distribuzione esatta di filtraggio.

Il problema può essere risolto modificando la distribuzione iniziale del vettore casuale (X_0, Z_0) . Invece di considerare come distribuzione iniziale $\sim N(\mu + \beta Z_0, Z_0) \times GIG(\lambda, \delta, \gamma)$ con $\lambda \in [0, \frac{1}{2})$, possiamo prendere come distribuzione marginale della variabile casuale Z_0 una distribuzione concentrata in un punto $z_0 > 0$. In questo modo

$$\begin{aligned} Z_1 &= z_0 + W_1, \\ Z_2 &= Z_1 + W_2 = z_0 + W_1 + W_2, \\ &\dots \\ Z_n &= z_0 + W_1 + \dots + W_n, \end{aligned}$$

e, dato che la sequenza $\{W_n\}$ ha distribuzione con supporto $(0, +\infty)$, $Z_n \in [z_0, +\infty)$.

La funzione $b_n(x_n, z_n, y_n)$ resta definita allo stesso modo, l'unica cosa che cambia è che ora l'insieme di definizione resta lontano dallo zero. La funzione $b_n(\cdot, y_n)$ risulta quindi limitata e il Teorema 5.4.3 può essere applicato. In questo caso, con un vettore casuale (X_0, Z_0) tale che $X_0 \sim N(\mu + \beta z_0, z_0)$ e $Z_0 \equiv z_0$, siamo in grado di affermare che l'algoritmo particle filter converge (debolmente) al filtro esatto, che tuttavia, in queste nuove ipotesi, non siamo in grado di esplicitare analiticamente.

Appendice A

Funzioni di Bessel e distribuzione GIG

La funzione di Bessel (che prende il nome dal matematico e astronomo tedesco Friedrich Wilhelm Bessel (1784–1846)) rappresenta la soluzione canonica $y(x)$ dell'equazione differenziale di Bessel

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (\lambda^2 x^2 - \nu^2) y = 0. \quad (\text{A.1})$$

Se poniamo $t = \lambda x$,

$$\frac{dy}{dx} = \lambda \frac{dy}{dt},$$

e

$$\frac{d^2 y}{dx^2} = \frac{d}{dx} \left(\frac{dy}{dx} \right) = \lambda \frac{d}{dt} \left(\frac{dy}{dx} \right) = \lambda^2 \frac{d^2 y}{dt^2}.$$

Sostituendo nella (A.1), si ha

$$t^2 \frac{d^2 y}{dt^2} + t \frac{dy}{dt} + (t^2 - \nu^2) y = 0.$$

Riscrivendo questa forma per $t = x$, si ha

$$x^2 y'' + xy' + (x^2 - \nu^2)y = 0. \quad (\text{A.2})$$

La (A.2) é la rappresentazione piú comune dell' equazione differenziale di Bessel, anche se la (A.1) é una forma piú generale. La soluzione della (A.2) é la *funzione di Bessel del primo tipo* di ordine ν , definita da

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{\nu+2k}}{k! \Gamma(\nu + k + 1)}.$$

Osservando il grafico delle funzioni di Bessel, si puó notare un comportamento oscillatorio del tipo seno o coseno. Le funzioni di Bessel sono generalizzazioni (non banali) delle funzioni trigonometriche. Infatti

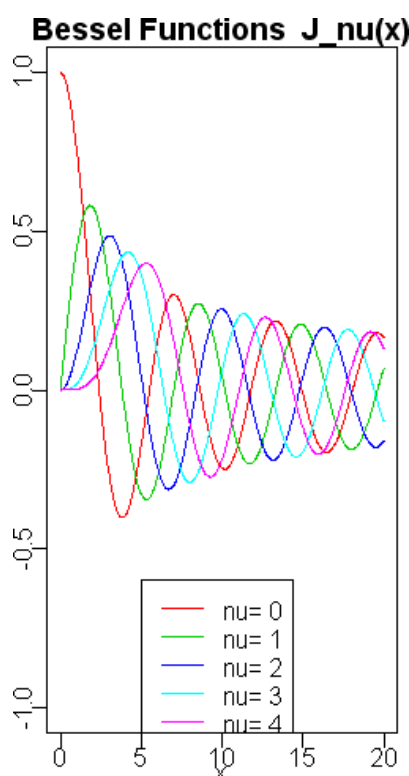


Figura A.1: Grafico della funzione di Bessel del primo tipo $J_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$

$$J_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin(x), \quad J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cos(x)$$

Osserviamo che se J_ν é soluzione della (A.2), anche $J_{-\nu}$ lo é, e se $\nu \notin \mathbb{N}$, le due soluzioni sono linearmente indipendenti. La soluzione generale dell'equazione differenziale di Bessel diventa

$$y(x) = c_0 J_\nu(x) + c_1 J_{-\nu}(x). \quad (\text{A.3})$$

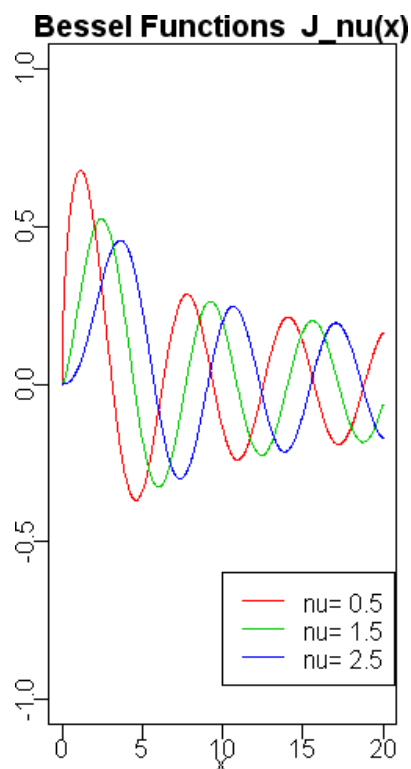


Figura A.2: Grafico della funzione di Bessel del primo tipo $J_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$

Dunque, se $\nu \notin \mathbb{N}$ la (A.3) rappresenta la soluzione generale dell'equazione differenziale (A.2). Ora definiamo una nuova funzione $Y_\nu(x)$ in termini delle due soluzioni linearmente indipendenti $J_\nu(x)$ e $J_{-\nu}(x)$

$$Y_\nu(x) := \frac{\cos(\pi\nu)J_\nu - J_{-\nu}}{\sin(\pi\nu)}. \quad (\text{A.4})$$

Dato che J_ν e Y_ν sono linearmente indipendenti per $\nu \notin \mathbb{N}$, possiamo scrivere la soluzione generale dell'equazione di Bessel come

$$y(x) = A_0 J_\nu(x) + A_1 Y_\nu(x), \quad (\text{A.5})$$

dove é facile vedere la corrispondenza con la soluzione (A.3) se i valori di c_0 e c_1 sono dati da

$$c_0 = A_0 + A_1 \frac{\cos(\pi\nu)}{\sin(\pi\nu)}, \quad c_1 = -\frac{A_1}{\sin(\pi\nu)}.$$

$Y_\nu(x)$ é nota come *funzione di Bessel del secondo tipo*.

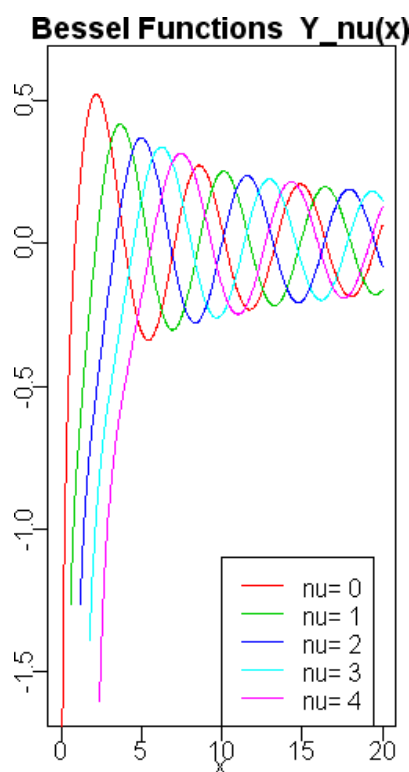


Figura A.3: Grafico della funzione di Bessel del secondo tipo $Y_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$

Se invece della (A.2) si analizza l'equazione modificata

$$x^2 y'' + xy' - (x^2 + \nu^2)y = 0, \quad (\text{A.6})$$

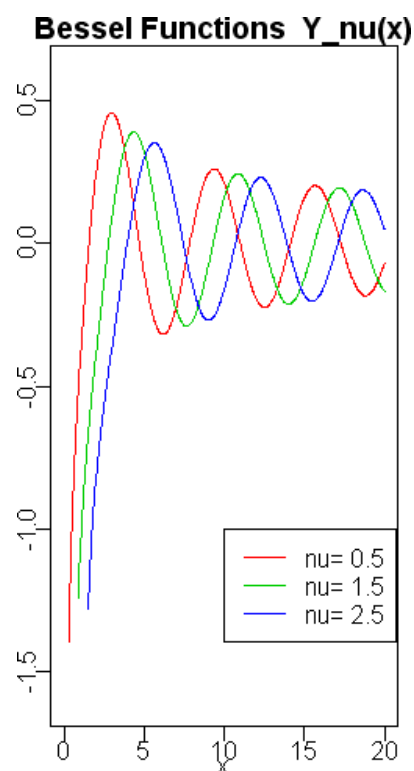


Figura A.4: Grafico della funzione di Bessel del secondo tipo $Y_{\nu}(x)$ di ordine $\nu = 0.5, 1.5, 2.5$

le soluzioni sono date dalle *funzioni di Bessel modificate del primo tipo*

$$I_{\nu}(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{\nu+2k}}{k! \Gamma(\nu + k + 1)}$$

e dalle *funzioni di Bessel modificate del secondo tipo*¹

$$K_{\nu}(x) = \frac{\pi}{2 \sin(\pi\nu)} [I_{-\nu}(x) - I_{\nu}(x)].$$

¹Le funzioni di Bessel modificate del secondo tipo sono note anche con i nomi meno comuni di *funzioni di Basset*, *funzioni di Bessel modificate del terzo tipo* o *funzioni di MacDonald*.

Diversamente dalle funzioni di Bessel ordinarie, che presentano un comportamento oscillatorio, le funzioni di Bessel modificate I_ν e K_ν sono funzioni rispettivamente crescenti e decrescenti con velocità esponenziale.

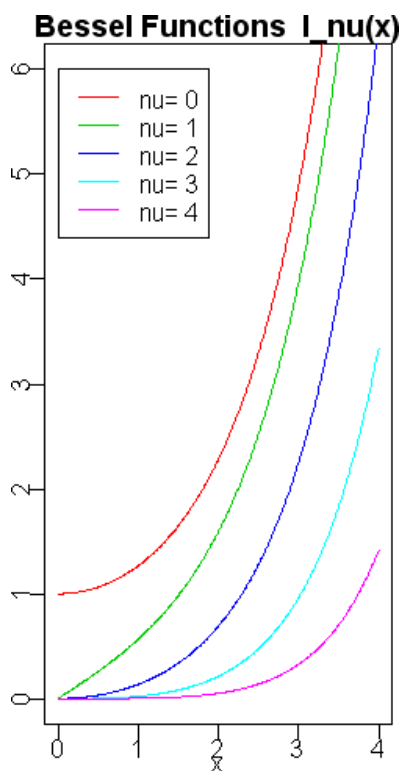


Figura A.5: Grafico della funzione di Bessel modificata del primo tipo $I_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$

Si vede che

$$I_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sinh(x), \quad I_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cosh(x)$$

e dunque

$$K_{\frac{1}{2}}(x) = \frac{\pi}{2} \sqrt{\frac{2}{\pi x}} [\cosh(x) - \sinh(x)] = \sqrt{\frac{\pi}{2x}} e^{-x}. \quad (\text{A.7})$$

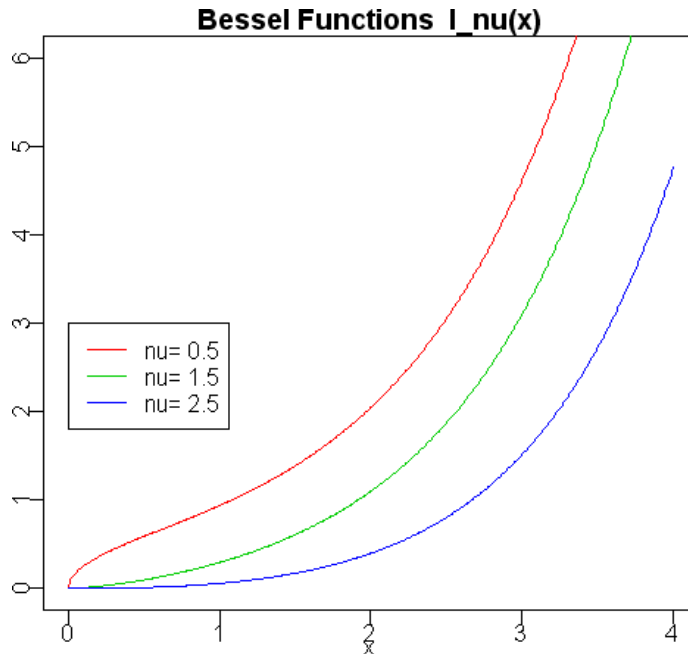


Figura A.6: Grafico della funzione di Bessel modificata del primo tipo $I_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$

Pertanto, come le J_ν generalizzano le funzioni seno e coseno, le I_ν generalizzano l'esponenziale. Le funzioni I_ν e K_ν sono linearmente indipendenti, quindi la soluzione generale dell'equazione differenziale (A.6) é la seguente

$$y(x) = C_0 I_\nu(x) + C_1 K_\nu(x).$$

A.1 Funzione di Bessel modificata del secondo tipo

In questo paragrafo riportiamo alcuni risultati riguardanti la funzione di Bessel modificata del secondo tipo che compare come fattore di normalizzazione nella funzione di densità della distribuzione gaussiana inversa generalizzata.

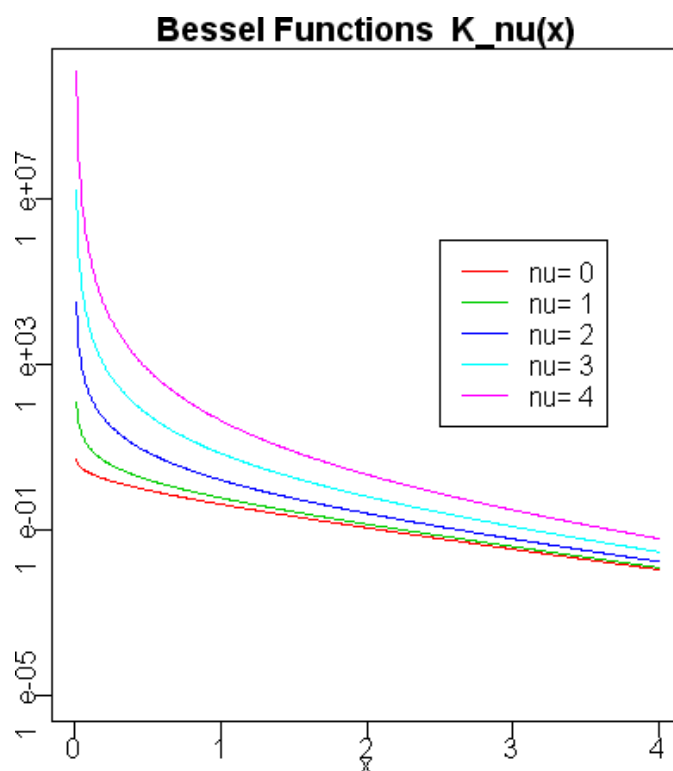


Figura A.7: Grafico della funzione di Bessel modificata del secondo tipo $K_\nu(x)$ di ordine $\nu = 0, 1, 2, 3, 4$

Tra le molte rappresentazioni integrali di $K_\nu(w)$, $w > 0$, vale la seguente

$$K_\nu(w) = \frac{1}{2} \int_0^{+\infty} x^{\nu-1} \exp \left[-\frac{1}{2} w (x + x^{-1}) \right] dx.$$

Riportiamo alcuni risultati riguardanti $K_\nu(\cdot)$. Per una trattazione piú approfondita, vedere Abramowitz e Stegun (1965). La funzione di Bessel $K_\nu(\cdot)$, $\nu \in \mathbb{R}$ soddisfa alle seguenti relazioni

$$K_\nu(w) = K_{-\nu}(w) \quad (\text{A.8})$$

$$K_{\nu+1}(w) = \frac{2\nu}{w} K_\nu(w) + K_{\nu-1}(w) \quad (\text{A.9})$$

$$K_{\nu-1}(w) + K_{\nu+1}(w) = -2K'_\nu(w). \quad (\text{A.10})$$

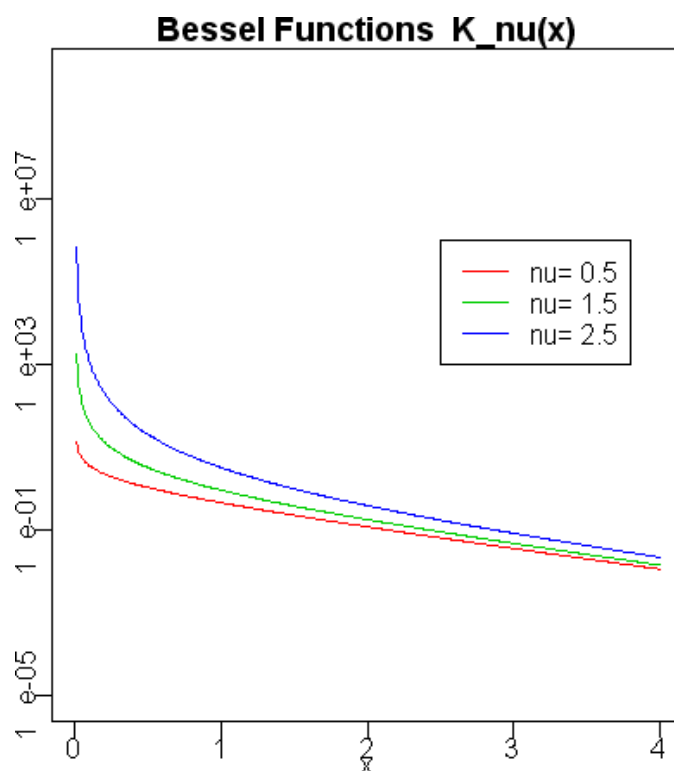


Figura A.8: Grafico della funzione di Bessel modificata del secondo tipo $K_\nu(x)$ di ordine $\nu = 0.5, 1.5, 2.5$

$$K_\nu(w) \approx \Gamma(\nu)2^{\nu-1}w^{-\nu}, \quad (w \downarrow 0), \quad (\nu > 0). \quad (\text{A.11})$$

Per $\nu = 0$ si ha

$$K_0(w) \approx -\ln w, \quad (w \downarrow 0). \quad (\text{A.12})$$

Da Ismail (1977) si ha la seguente relazione asintotica

$$K_\nu(w) \approx \sqrt{\frac{\pi}{2}}2^\nu \nu^{\nu-1/2} e^{-\nu} w^{-\nu}, \quad (\nu \rightarrow \infty) \quad (\text{A.13})$$

che può essere vista come l'analogo della formula di Stirling per la funzione Gamma.

A.2 Le funzioni D_ν e R_ν

Definiamo le funzioni R_ν e D_ν

$$R_\nu(w) := \frac{K_{\nu+1}(w)}{K_\nu(w)} \quad (\text{A.14})$$

$$D_\nu(w) := \frac{K_{\nu+1}(w)K_{\nu-1}(w)}{K_\nu^2(w)}. \quad (\text{A.15})$$

Valgono le seguenti relazioni (utilizzando le proprietà (A.8) e (A.9)):

$$\begin{aligned} D_{-\nu}(w) &= \frac{K_{-\nu+1}(w)K_{-\nu-1}(w)}{K_{-\nu}^2(w)} \\ &= \frac{K_{-(\nu-1)}(w)K_{-(\nu+1)}(w)}{K_{-\nu}^2(w)} \\ &= \frac{K_{\nu-1}(w)K_{\nu+1}(w)}{K_\nu^2(w)} = D_\nu(w) \end{aligned} \quad (\text{A.16})$$

$$D_\nu(w) = R_\nu(w)R_{-\nu}(w) \quad (\text{A.17})$$

$$R_{-\nu}(w) = \frac{K_{\nu-1}(w)}{K_\nu(w)} = R_{\nu-1}^{-1} \quad (\text{A.18})$$

$$R_\nu(w) = \frac{K_{\nu+1}(w)}{K_\nu(w)} = \frac{2\nu}{w} + \frac{K_{\nu-1}(w)}{K_\nu(w)} = \frac{2\nu}{w} + R_{-\nu}(w) \quad (\text{A.19})$$

$$R_\nu(w) = \frac{\nu}{w} + \sqrt{\left(\frac{\nu}{w}\right)^2 + D_\nu(w)} \quad (\text{A.20})$$

A.3 Proprietá statistiche della distribuzione GIG

Lo scopo di questa sezione é quello di studiare la distribuzione gaussiana inversa generalizzata di parametri $(\lambda, \delta, \gamma)$, $\text{GIG}(\lambda, \delta, \gamma)$, la cui funzione di densitá é data da

$$\frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\left[-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right], \quad x > 0, \quad (\text{A.21})$$

dove K_λ é la funzione di Bessel modificata del secondo tipo con indice λ .

Il dominio dei parametri λ, δ e γ é dato dall'insieme

$$D_{\text{GIG}} = \{(\lambda, \delta, \gamma) \in \mathbb{R} \times [0, +\infty)^2 : \lambda^+ + \delta > 0, \lambda^- + \gamma > 0\},$$

dove, per $a \in \mathbb{R}$, $a^+ = \max(a, 0)$ e $a^- = -\min(a, 0)$. Dalla A.21, segue che $\delta\gamma$ é il parametro di forma e δ/γ é il parametro di scala. Dato che λ

é l'indice della funzione di Bessel nella costante di normalizzazione, λ verrà chiamato parametro indice.

Introduciamo i parametri ω e η

$$\omega := \delta\gamma, \quad \eta := \frac{\gamma}{\delta}.$$

Ora $\omega = 0$ comprende sia il caso in cui $\delta = 0$, $\lambda > 0$, sia il caso in cui $\gamma = 0$, $\lambda < 0$. Cosí si ha $\omega > 0$ se entrambi i parametri γ e δ sono positivi. Allora la (A.21) assume la forma

$$\frac{\eta^{-\lambda}}{2K_\lambda(\omega)} x^{\lambda-1} \exp \left[-\frac{1}{2}\omega(\eta x^{-1} + \eta^{-1}x) \right]. \quad (\text{A.22})$$

Se $\delta\gamma \neq 0$ e $X \sim GIG(\lambda, \delta, \gamma)$ allora $X^{-1} \sim GIG(-\lambda, \gamma, \delta)$. Infatti, posto $y = 1/x$, per la proprietá (A.8):

$$\begin{aligned} \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} y^{\lambda-1} \exp \left[-\frac{1}{2}(\delta^2 y + \gamma^2 y^{-1}) \right] \frac{1}{y^2} = \\ \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} y^{-\lambda-1} \exp \left[-\frac{1}{2}(\delta^2 y + \gamma^2 y^{-1}) \right] = \\ \frac{(\delta/\gamma)^{-\lambda}}{2K_\lambda(\delta\gamma)} y^{-\lambda-1} \exp \left[-\frac{1}{2}(\delta^2 y + \gamma^2 y^{-1}) \right]. \end{aligned}$$

□

Casi speciali della (A.21) si ottengono per particolari valori dei parametri. Ad esempio, se $\lambda = -1/2$ ($\lambda = 1/2$) é la funzione di densitá della gaussiana inversa (reciproca). Per mostrare questo risultato basta ricordare la (A.7).

Per la (A.11), quando $\delta = 0$ (che implica che $\lambda > 0$), si ha la distribuzione gamma di parametri λ e $\frac{\gamma^2}{2}$, mentre se $\gamma = 0$ (e allora $\lambda < 0$) la (A.21) definisce la distribuzione associata gamma reciproca.

Un altro caso di interesse si ha quando $\lambda = 0$ che determina la distribuzione iperbolica.

La distribuzione gaussiana inversa generalizzata gode di alcune utili proprietá rispetto al prodotto di convoluzione

$$GIG(-\lambda, \delta, \gamma) * GIG(\lambda, 0, \gamma) = GIG(\lambda, \delta, \gamma), \quad \lambda > 0 \quad (\text{A.23})$$

Infatti

$$\int_0^{+\infty} p_1(x)p_2(y-x)dx =$$

dove $p_1 \sim GIG(-\lambda, \delta, \gamma)$ e $p_2 \sim GIG(\lambda, 0, \gamma)$ che coincide con $Gamma(\lambda, \gamma^2/2)$

$$\begin{aligned} &= \int_0^{+\infty} \frac{(\gamma/\delta)^{-\lambda}}{2K_{-\lambda}(\delta\gamma)} x^{-\lambda-1} \exp\left[-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right] \cdot \\ &\quad \cdot \frac{(\gamma^2/2)^\lambda}{\Gamma(\lambda)} (y-x)^{\lambda-1} \exp\left[-\frac{\gamma^2(y-x)}{2}\right] dx \\ &= \int_0^{+\infty} \frac{(\gamma/\delta)^{-\lambda}}{2K_{-\lambda}(\delta\gamma)} x^{-\lambda-1} \exp\left[-\frac{1}{2}(\delta^2 x^{-1})\right] \frac{(\gamma^2/2)^\lambda}{\Gamma(\lambda)} (y-x)^{\lambda-1} \exp\left[-\frac{\gamma^2 y}{2}\right] dx \\ &= \frac{(\gamma/\delta)^{-\lambda}}{2K_\lambda(\delta\gamma)} \frac{(\gamma^2/2)^\lambda}{\Gamma(\lambda)} \exp\left[-\frac{\gamma^2 y}{2}\right] \int_0^{+\infty} x^{-\lambda-1} \exp\left[-\frac{1}{2}(\delta^2 x^{-1})\right] (y-x)^{\lambda-1} dx \\ &= \frac{(\gamma/\delta)^{-\lambda}}{2K_\lambda(\delta\gamma)} \frac{(\gamma^2/2)^\lambda}{\Gamma(\lambda)} \exp\left[-\frac{\gamma^2 y}{2}\right] 2^\lambda \Gamma(\lambda) \delta^{-2\lambda} y^{\lambda-1} \exp[-\delta^2/(2y)] \\ &= \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} \exp\left[-\frac{\gamma^2 y}{2} - \frac{\delta^2}{2y}\right] y^{\lambda-1}. \end{aligned}$$

I primi due momenti di una variabile casuale $X \sim GIG(\lambda, \delta, \gamma)$ sono dati da:

$$E[X] = R_\lambda(\delta\gamma) \frac{\delta}{\gamma}, \quad V[X] = \left[R_\lambda(\delta\gamma) \frac{\delta}{\gamma} \right]^2 \cdot [D_{\lambda+1}(\delta\gamma) - 1], \quad (\text{A.24})$$

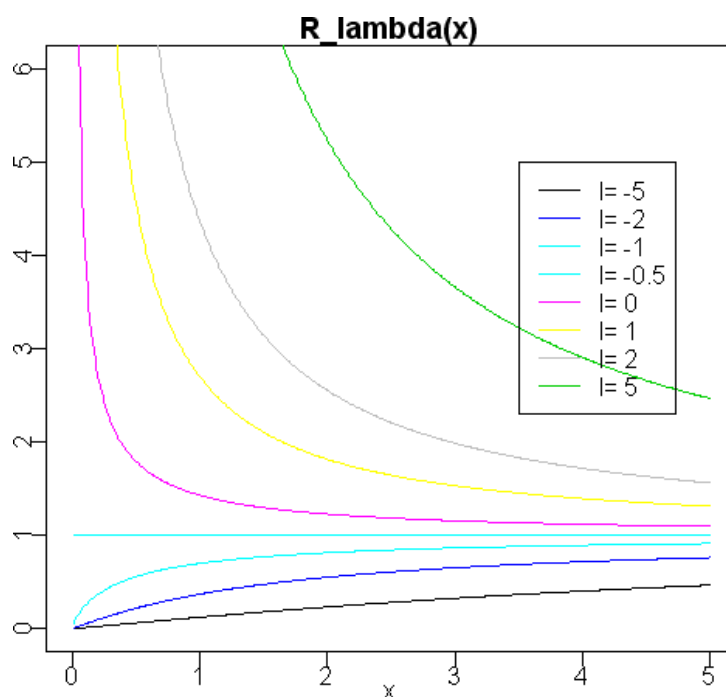
dove

$$R_\lambda(x) = \frac{K_{\lambda+1}(x)}{K_\lambda(x)}, \quad D_\lambda(x) = \frac{K_{\lambda+1}(x)K_{\lambda-1}(x)}{(K_\lambda(x))^2}, \quad x > 0.$$

I grafici riportati mostrano che $R_\lambda(x)$ é una funzione crescente (decrecente) di x , se $\lambda < -\frac{1}{2}$ ($\lambda > -\frac{1}{2}$), mentre $R_{-\frac{1}{2}} \equiv 1$. Per ogni $x > 0$ fissato, $R_\lambda(x)$ cresce come funzione di λ . Inoltre, $D_\lambda(x)$ é decrescente per ogni λ dato e crescente (decrecente) come funzione di λ , per $\lambda < 0$ ($\lambda > 0$). Se x tende all'infinito, sia $R_\lambda(x)$ che $D_\lambda(x)$ convergono a 1.

Per una dimostrazione formale, si veda, ad esempio, Jørgensen, 1982.

La distribuzione gaussiana inversa generalizzata é stata considerata da Barndorff-Nielsen (1977) per caratterizzare la distribuzione iperbolica generalizzata come una mistura normale media-varianza, essendo la distribuzione di mixing la (A.21). Più precisamente, assumiamo che $(X, Z) \sim$

Figura A.9: Grafico di $R_\lambda(\cdot)$ per diversi valori di λ

$N(\mu + \beta Z, Z) \times GIG(\lambda, \delta, \gamma)$, cioè la densità di un vettore casuale (X, Z) é data dal prodotto della densità condizionata di X dato $Z = z$, che é una densità normale di media $\mu + \beta z$ e varianza z , e la densità di una variabile casuale $z \sim GIG(\lambda, \delta, \gamma)$. Diciamo che (X, Z) ha distribuzione bivariata gaussiana inversa generalizzata.

Con queste assunzioni, la distribuzione marginale di X é una distribuzione iperbolica generalizzata di parametri $(\lambda, \alpha, \beta, \mu, \delta)$ ($GH(\lambda, \alpha, \beta, \mu, \delta)$), con $\alpha = \sqrt{\beta^2 + \gamma^2}$. La funzione di densità associata é

$$c(\lambda, \alpha, \beta, \mu, \delta) \left[q\left(\frac{x - \mu}{\delta}\right) \right]^{\lambda - 1/2} K_{\lambda - 1/2}\left(\delta \alpha q\left(\frac{x - \mu}{\delta}\right)\right) \exp(\beta x), \quad (x > 0), \quad (\text{A.25})$$

dove $q(x) = \sqrt{x^2 + 1}$ e

$$c(\lambda, \alpha, \beta, \mu, \delta) = (2\pi)^{-1/2} \delta^{-1/2} \alpha^{-\lambda + 1/2} (\alpha^2 - \beta^2)^{\lambda/2} \cdot [K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})]^{-1} \exp(-\beta \mu).$$

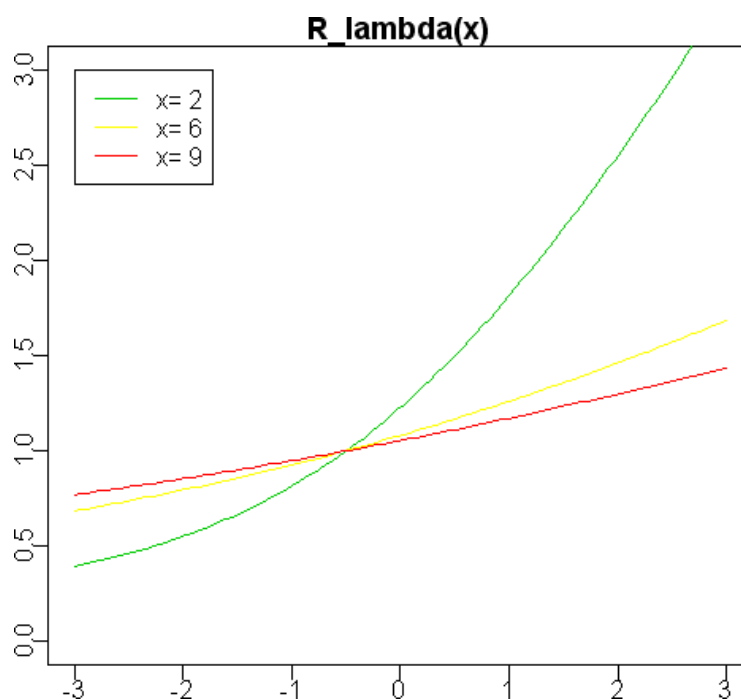


Figura A.10: Grafico di $R(x)$ per diversi valori di x

Il dominio dei cinque parametri $(\lambda, \alpha, \beta, \mu, \delta)$ é dato dall'insieme

$$D_{GH} = \{(\lambda, \alpha, \beta, \mu, \delta) \in \mathbb{R}^3 \times [0, +\infty)^2 : \lambda^+ + \delta > 0, \lambda^- + \alpha > 0, \\ |\beta| \leq \alpha, |\beta| \neq \alpha \text{ se } \lambda \geq 0\}$$

Dalla (A.25) segue che μ é il parametro di location, δ il parametro di scala e β il parametro di asimmetria (se $\beta = 0$ si ha una distribuzione simmetrica). Le principali proprietà di queste distribuzioni sono studiate in [4]. Appartengono a questa classe alcune distribuzioni di interesse. In particolare, la distribuzione gaussiana normale inversa e la distribuzione iperbolica corrispondono ai valori $\lambda = -\frac{1}{2}$ e $\lambda = 1$ rispettivamente. Altri casi interessanti sono $\lambda = \frac{1}{2}$ e $\lambda = 0$, che indicano rispettivamente la distribuzione gaussiana normale inversa reciproca e la distribuzione iperbolicoide. I primi due momenti di una variabile casuale $X \sim GH(\lambda, \alpha, \beta, \mu, \delta)$ sono dati da

$$E[X] = \mu + a\beta, \quad V[X] = a + b\beta^2, \quad (\text{A.26})$$

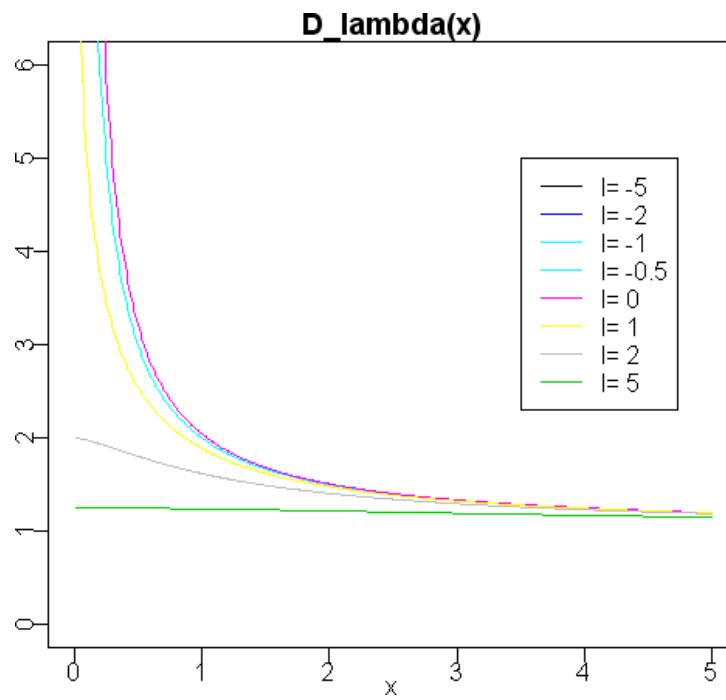


Figura A.11: Grafico di $D_\lambda(\cdot)$ per diversi valori di λ

dove

$$a = \frac{\delta}{\sqrt{\alpha^2 - \beta^2}} R_\lambda(\delta\sqrt{\alpha^2 - \beta^2})$$

$$b = \frac{\delta^2}{\alpha^2 - \beta^2} \left[1 + 2 \frac{\lambda + 1}{\delta\sqrt{\alpha^2 - \beta^2}} R_\lambda(\delta\sqrt{\alpha^2 - \beta^2}) - [R_\lambda(\delta\sqrt{\alpha^2 - \beta^2})]^2 \right].$$

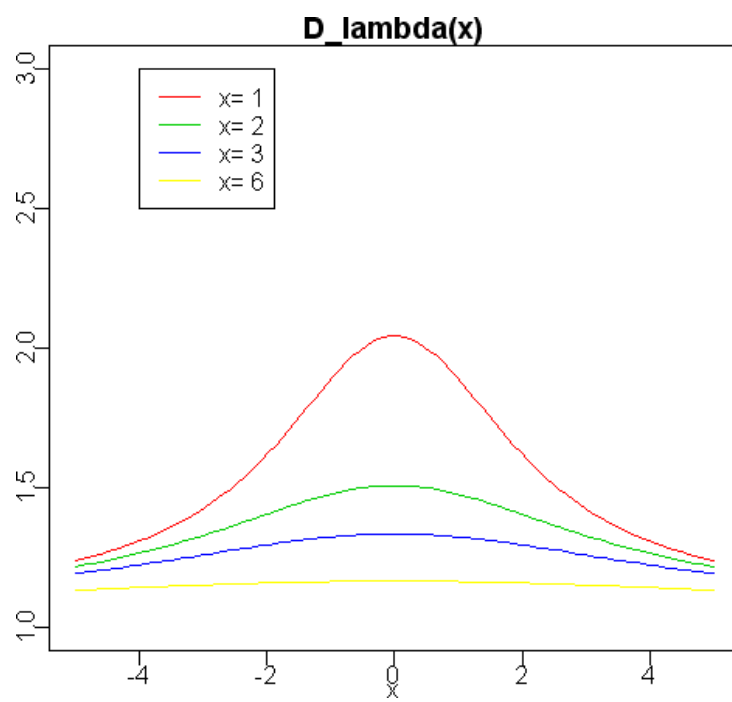


Figura A.12: Grafico di $D_\lambda(x)$ per diversi valori di λ

Bibliografia

- [1] Abramovitz, M. e Stegun, I.A. *Handbook of Mathematical Functions*, Dover Publications New York (1965)
- [2] Akaike, H. *Markovian representation of stochastic processes and its application to the analysis of autoregressive and moving average processes* Ann. Inst. Statist. Math., **26**, 363-387 (1974)
- [3] Barndorff-Nielsen, O.E., *Exponential decreasing distributions for the logarithm of particle size.*, Proc. Royal Soc. London A, **353**, 401-419 (1977)
- [4] Barndorff-Nielsen, O.E., *Hyperbolic distributions and distributions on hyperbolae*, Scan. J. Statist.,**5**, 151-157 (1978)
- [5] Barndorff-Nielsen, O.E. e Blæsild, P., *Hyperbolic distributions and ramifications: contributions to theory and applications*, In: Tallie, C., Patil, G.P., Baldessarri, B.A. (Eds), *Statistical Distributions in Scientific Work 4.D*. Reidel Publishing Company, Dordrecht (1981)
- [6] Baum, L.E. e Petrie, T. *Statistical inference for probabilistic functions of finite state Markov chains*, Ann. Math. Statist., **37**, 1554-1563 (1966)
- [7] Crisan, D. e Doucet, A. *A survey of convergence results on particle filtering for practitioners*, IEEE Trans. Signal Processing, **50** 736-746 (2002)
- [8] Ferrante, M e Vidoni, P. *Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises* Stochastic Process. Appl., **77**, 69-81 (1998)

-
- [9] Ferrante, M e Vidoni, P. *A Gaussian-generalized inverse Gaussian finite-dimensional filter* In Stochastic processes and their applications (1999)
- [10] Harrison, P.J. e Stevens, C.F. *Bayesian forecasting (with discussion)* J. Royal Statist. Soc. B, **38**, 205-247 (1976)
- [11] Ismail, M.E.H. *Integral Representations and complete monotonicity of various quotients of Bessel functions* Can.J.Math. **24**, 1198-1207 (1977)
- [12] Jørgensen, B. *Statistical Properties of the generalized inverse gaussian distribution* Springer-Verlag Berlin (1982)
- [13] Kalman, R.E. *A new approach to linear filtering and prediction problems.* Trans. Amer. Soc. Mech. Eng., J. Basic Eng., **82**, 35-45 (1960)
- [14] Kalman, R.E. e Bucy, R.S. *New results in linear filtering and prediction theory* Trans. Amer. Soc. Mech. Eng., J. Basic Eng., **83**, 95-108 (1961)
- [15] Kitagawa, G. *Non-Gaussian state-space modeling of nonstationary time series (with discussions)* J. Amer. Statist. Assoc., **82**, 1032-1063 (1987)
- [16] Lauritzen, S.L. *Time series analysis in 1880: A discussion of contributions made by T.N.Thiele* International Statist. Review, **49**, 319-331 (1981)
- [17] Plackett, R.L. *Some theorems in least squares* Biometrika, **37**, 149-157 (1950)
- [18] Vidoni, P. *Exponential family state space models based on a conjugate latent process* J. Royal Statist. Soc. Ser. B, **61**, 213-221 (1999)