Indice

Indice delle figure

NOTAZIONI

Simboli

$\langle\ ,\ \rangle$	prodotto scalare
$\|\ \|$	norma (generica)
$\|f\|_p$	norma della funzione f in $L_p(\mathbb{R})$
$\oplus$	somma diretta
$\otimes$	prodotto tensoriale
$\overline{x}$	complesso coniugato del numero complesso x
$Supp(f)$	supporto di f
$Span\{f_n\}$	spazio generato dalla sequenza f_n
Z_f	trasformata di Zak di f
$L_p(\mathbb{R})$	spazio delle funzioni di p-potenza integrabile
$L_1(\mathbb{R})$	spazio delle funzioni integrabili
$L_2(\mathbb{R})$	spazio delle funzioni di quadrato integrabile
C^k	spazio delle funzioni con k derivate continue
$l_p(\mathbb{Z})$	spazio delle sequenze di p potenza sommabile
$l_2(\mathbb{Z})$	spazio delle sequenze di quadrato sommabile
S_r	spazio di Schwartz di ordine r
S'	spazio delle distribuzioni temperate
H^α, H_p^α	spazi di Sobolev
B_{pq}^s	spazio di Besov
δ_{ij}	simbolo di Kronecker
$\delta(x=y)$	funzione delta di Dirac
$\mathcal{F}_f$	trasformata di Fourier di f
$\mathcal{F}_f^{-1}$	trasformata inversa di Fourier di f
$W_f(a,b)$	trasformata continua wavelet di f
$W_f^{-1}(a,b)$	trasformata continua wavelet inversa di f
ϕ	funzione di scala
ψ	wavelet madre
j	livello di MRA (vedi abbreviazioni)
k	parametro di traslazione
V_j	sottospazio di MRA
P_j	operatore di proiezione in V_j
W_j	sottospazio complemento ortogonale di V_j in V_{j+1}

h_k	coefficienti di filtro
$c_{j,k}$	coefficienti di scala (livello j, posizione k)
$d_{j,k}$	coefficienti wavelet (livello j, posizione k)
$q_j(x, y)$	nucleo autoriproducente di V_j

Abbreviazioni

MRA	MultiResolution Analysis (analisi a più livelli di risoluzione)
DWT	Discrete Wavelet Transform (trasformata wavelet discreta)
IWT	Inverse Wavelet Transform (trasformata wavelet inversa)
STFT	Short Time Fourier Transform (trasformata di Fourier con finestra)
EQM	Errore Quadratico Medio
EAM	Errore Assoluto Medio
EQI	Errore Quadratico Integrale
EQMI	Errore Quadratico Medio Integrale

Capitolo 1

Introduzione

1.1 Struttura del lavoro

In questa tesi si propone l'esplorazione delle "wavelets" come strumento di analisi statistica. L'argomento, sebbene recente, suggerisce molti temi di ricerca; quelli qui trattati sono la stima non parametrica di funzioni, con particolare riferimento a funzioni di densità e di regressione, e l'analisi di serie storiche. Tra le applicazioni recentemente emerse, la cui trattazione è, per il momento, rinviata, si ricordano quelle relative all'analisi di dati per componenti principali, all'analisi discriminante e alla risoluzione di problemi inversi.

Date le finalità applicative, una ampia descrizione della teoria inerente le funzioni wavelet non rientra negli scopi di questa tesi. Pertanto, solo una definizione "intuitiva" delle wavelets ed una descrizione di quelle caratteristiche che ne fanno uno valido strumento applicativo è riportata nella sezione seguente, mentre ognuno dei capitoli successivi contiene una breve introduzione alle nozioni teoriche strettamente necessarie alla comprensione del contenuto.

Le funzioni wavelet saranno utilizzate nel secondo capitolo per la costruzione di stimatori non parametrici di funzioni di densità e di regressione e nel terzo capitolo per la rimozione di disturbi Gaussiani da serie storiche, per la stima di spettri, per la eliminazione di outliers e per la definizione di uno spettro wavelet.

Oltre ad una sistematizzazione dei contributi relativi alla utilizzazione delle wavelets in tali ambiti di ricerca, questo lavoro di tesi contiene la presentazione di alcuni contributi metodologici originali e la realizzazione di applicazioni a insiemi di dati di varia natura che hanno lo scopo di verificare la validità delle wavelets come strumento statistico per le applicazioni. In particolare, nell'ambito della stima di funzioni di densità è proposto un metodo nuovo che utilizza l'informazione di Fisher, o più precisamente una sua stima wavelet, come funzione di penalità allo scopo di determinare il valore del parametro di scala al quale troncare lo sviluppo in serie wavelet della funzione di densità. Originali sono anche alcuni aspetti della estensione multivariata. Le applicazioni degli stimatori wavelet presentati riguardano un insieme di dati simulati e il dataset bivariato “Old Faithfull Geyser”. Nell'ambito della stima di funzioni di regressione, gli stimatori wavelet sono applicati ad un insieme di tassi di cambio e i risultati ottenuti sono confrontati con quelli ottenuti tramite l'utilizzo di stimatori classici.

Per quanto riguarda invece la rimozione di disturbi Gaussiani da serie storiche, particolare enfasi è data ad un approccio che utilizza un modello Bayesiano per i coefficienti wavelet. L'applicazione della analisi Bayesiana in questo contesto rappresenta un nuovo tema di ricerca che sta riscuotendo particolare attenzione; i risultati ottenuti dalla applicazione del modello presentato ad una serie storica di prezzi azionari incoraggiano ulteriori approfondimenti. La tesi si conclude con la sistemazione di alcune applicazioni delle wavelets già presenti in letteratura; tra queste, si segnala la descrizione dello spettro wavelet che, data l'analogia con lo spettro di Fourier, promette interessanti risultati nel campo dell'analisi spettrale di serie storiche.

1.2 Wavelets: brevi cenni

Nate nei primi anni '80 come trasformazione integrale che coinvolge traslazioni e dilatazioni (o compressioni) di una funzione base, le wavelets furono inizialmente utilizzate come strumento applicativo nella rappresentazione di segnali sismici, [83]. Tale rappresentazione risultava migliore di quella ottenibile con la trasformata di Fourier. Negli anni immediatamente successivi, la teoria matematica è stata rigorosamente formalizzata, [55], fino ad arrivare, alla metà degli anni '80, alla formulazione di una rappresentazione discreta ortogonale, [75] [78], e alla costruzione di basi di funzioni wavelet ortonormali e con supporto compatto, [27]. Il crescente interesse dimostrato dalla comunità scientifica verso questo nuovo strumento di analisi è dovuto al fatto che le wavelets rappresentano una sintesi di teorie e concetti che si sono sviluppati negli ultimi anni in campi diversi, quali l'ingegneria, la fisica, la matematica pura e l'analisi numerica, [80]. Ma soprattutto, le wavelets sono uno strumento matematico semplice che presenta una notevole varietà di possibili applicazioni quali l'analisi di segnali, suoni e immagini, l'analisi numerica e la geometria frattale.

Lo scopo di questa sezione è quello di fornire al lettore una descrizione intuitiva delle funzioni wavelet evidenziando quelle caratteristiche che ne hanno fatto uno strumento applicativo “migliore” di quelli tradizionali. Una chiara esposizione della trasformata wavelet richiede, infatti, l'introduzione della trasformata di Fourier; le proprietà e soprattutto le limitazioni di quest'ultima motiveranno l'introduzione della trasformata wavelet.

Data f appartenente ad $L_2(\mathbb{R})$, spazio delle funzioni di quadrato integrabile (vedi Appendice A), la *trasformata di Fourier* di f è definita come, [14],

$$\mathcal{F}_f(\omega) = \int_{\mathbb{R}} e^{-ix\omega} f(x) dx \tag{1.1}$$

mentre la *trasformata inversa di Fourier* è data da

$$\mathcal{F}_f^{-1}(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{ix\omega} \mathcal{F}_f(\omega) d\omega. \tag{1.2}$$

La trasformata di Fourier riveste un ruolo molto importante nella teoria della elaborazione di segnali: una funzione f appartenente a $L_2(\mathbb{R})$ descrive infatti un fenomeno che varia con il tempo o, secondo la terminologia ingegneristica, un segnale analogico con energia (la norma L_2) finita. Tale fenomeno può essere analizzato in due diversi domini, quello temporale e quello frequenziale. La trasformata di Fourier di f e la sua inversa consentono il passaggio dal dominio temporale a quello frequenziale e viceversa [1].

Tra le proprietà della trasformata di Fourier, che saranno utili in seguito, si ricordano l'*Identità di Parseval*

$$\langle f, g \rangle = \frac{1}{2\pi} \langle \mathcal{F}_f, \mathcal{F}_g \rangle \qquad f, g \in L_2(\mathbb{R})$$

e la *Formula di Plancherel*

$$\|f\|_2^2 = \frac{1}{2\pi} \|\mathcal{F}_f\|^2.$$

Il tipo di analisi ottenibile utilizzando la trasformata di Fourier e la sua inversa presenta una grossa limitazione costituita dal fatto che tali strumenti rendono impossibile una analisi simultanea del fenomeno nei due diversi domini. Tale limitazione ha motivato la ricerca di uno strumento che consentisse una analisi tempo/frequenza di un segnale. Una analogia spesso utilizzata in letteratura per questo tipo di rappresentazione è quella con la notazione musicale, [29]; in musica, infatti, uno spartito permette di leggere sia l'ordine nel quale le note saranno suonate (tempo) sia quale sarà la durata di ciascuna nota (frequenza).

Tra le varie tecniche proposte, particolare successo è stato ottenuto da quella suggerita da Gabor, [50]: utilizzare la funzione Gaussiana come "finestra" temporale su f. Tale concetto è stato successivamente generalizzato, [24], ed ha portato alla introduzione di una *trasformata di Fourier con finestra* (Short-Time Fourier Transform, STFT) che permette di estrarre il contenuto frequenziale del

[1] Se il tempo è misurato in secondi, la frequenza è data da cicli per secondo ed è misurata in Hertz mentre la frequenza angolare, $\omega = 2\pi * frequenza$, che compare nella trasformata di Fourier, è in radianti per secondo.

segnale limitatamente ad un determinato intervallo di tempo. Tale trasformata è definita come

$$STFT_f(x_0,\omega) = \int_{\mathbb{R}} f(x)g(x-x_0)e^{-i\omega x}dx, \qquad (1.3)$$

dove g è una *funzione finestra* [2] di *centro* $x^* = \int_{\mathbb{R}} x|g(x)|^2dx$. Il prodotto di f per g consente, al variare del parametro di traslazione x_0, di individuare la parte di f relativa a intervalli di tempo centrati in $x^* + x_0$.

Rimane ancora un problema: l'ampiezza delle finestre ottenute al variare di x_0 è costante e non permette una buona analisi di segnali non stazionari, il cui contenuto frequenziale varia con il tempo. La necessità di una trasformata che utilizzasse funzioni finestra di ampiezza variabile (finestre strette per l'analisi delle alte frequenze e più larghe per le basse) ha determinato l'introduzione di una *trasformata continua wavelet* di f, [55], definita come

$$W_f(a,b) = \langle f, \psi_{a,b}\rangle = \int_{\mathbb{R}} f(x)\overline{\psi_{a,b}(x)}dx \qquad (1.4)$$

dove il parametro a di *scala* e quello b di *traslazione* sono numeri reali e dove $\{\psi_{a,b}\ ;\ a,b \in \mathbb{R},\ a \neq 0\}$ costituisce una famiglia di funzioni definite come traslazioni e dilatazioni di una funzione di base, ossia

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi(\frac{x-b}{a}).$$

La funzione di base $\psi \in L_2(\mathbb{R})$ è detta *wavelet madre* e deve soddisfare la seguente *condizione di ammissibilità*

$$C_\psi = \int_{\mathbb{R}} \frac{|\mathcal{F}_\psi(\omega)|^2}{\omega}d\omega < +\infty. \qquad (1.5)$$

La costante $1/\sqrt{|a|}$ è tale che

$$\int_{\mathbb{R}} |\psi_{a,b}(x)|^2dx = \int_{\mathbb{R}} |\psi(x)|^2dx, \quad \forall a.$$

Naturalmente, nota la trasformata continua wavelet di f, è possibile definire la *trasformata continua inversa* come

$$f(x) = C_\psi^{-1}\int_{\mathbb{R}}\int_{\mathbb{R}} W_f(a,b)\psi_{a,b}(x)\frac{dadb}{a^2} \qquad (1.6)$$

[2] Si definisce funzione finestra una funzione $g \in L_2(\mathbb{R})$ tale che $xg(x) \in L_2(\mathbb{R})$. Intuitivamente, una funzione finestra è una funzione a supporto finito e sufficientemente regolare.

che è nota anche come *formula di ricostruzione di f* poichè consente di ricostruire la funzione a partire dall'insieme (continuo) di coefficienti wavelet $\{W_f(a,b) = \langle f, \psi_{a,b}\rangle;\ a,b \in \mathbb{R}\ \ a \neq 0\}$ ottenuti dalla trasformata wavelet continua. La convergenza ad f è assicurata dalla condizione di ammissibilità (1.5).

E' possibile dimostrare che, se ψ appartiene allo spazio $L_1(\mathbb{R})$ delle funzioni integrabili (vedi Appendice A), la condizione (1.5) implica $\mathcal{F}_\psi(0) = 0$, ovvero

$$\int_{\mathbb{R}} \psi(x)dx = 0.$$

Una wavelet madre è quindi una funzione con media zero che ha, in generale, un comportamento oscillatorio. La condizione di ammissibilità implica anche che ψ decada a zero per $|x| \to +\infty$. Tali proprietà motivano il nome di *wavelet madre* assegnato a ψ; tale funzione è infatti una "piccola onda" dalla quale, mediante traslazioni e dilatazioni, si ottiene la famiglia di "piccole onde"

$$\{\psi_{a,b},\ a,b \in \mathbb{R},\ a \neq 0\}.$$

Si osservi che la trasformata continua wavelet può essere vista come una trasformata con finestra dove però l'introduzione di un secondo parametro, quello di scala, permette di avere funzioni finestra di ampiezza variabile. In particolare, mentre la trasformata di Fourier con finestra è costruita utilizzando funzioni seno e coseno moltiplicate per una funzione finestra, in quella wavelet la funzione finestra è già oscillante (la wavelet madre) e la trasformata è costruita semplicemente utilizzando traslazioni e dilatazioni di questa. La figura 1.1 mostra un particolare tipo di wavelet madre, la funzione *Sombrero* [3] (Mexican hat) $\psi(x) = (2/\sqrt{3})\pi^{-1/4}(1-x^2)\exp(-x^2/2)$, ed esemplifica come piccoli (grandi) valori del parametro di scala a consentano di ottenere versioni compresse (dilatate) di questa, ovvero funzioni finestra strette (larghe). Valori diversi del parametro b consentono, come nel caso della trasformata con finestra di Fourier, di traslare la wavelet madre lungo l'asse temporale.

Si ricorda che una funzione è detta ben localizzata nel tempo se decade rapidamente a zero per $|x| \to +\infty$ e ben localizzata in frequenza se la sua trasformata di Fourier decade rapidamente a zero per $|\omega| \to +\infty$. La figura 1.1 mostra come le funzioni wavelet siano ben localizzate nel tempo (per piccoli valori di a) e nella frequenza (per grandi valori di a). Le funzioni trigonometriche, utilizzate nella trasformata di Fourier, sono invece localizzate solo in frequenza.

[3] La funzione sombrero è ottenuta come derivata seconda della funzione Gaussiana $e^{x^2/2}$. La funzione è ri-normalizzata in modo tale che la norma L_2 sia uno. Il grafico ottenuto ruotando la funzione intorno al proprio asse di simmetria è simile all'immagine di un cappello messicano.

In conclusione, la trasformata continua wavelet o, più in generale, la famiglia di funzioni wavelet $\{\psi_{a,b}\}$, presentano rispetto all'analisi di Fourier una maggiore flessibilità nell'analisi locale di segnali, particolarmente utile nel caso non stazionario.

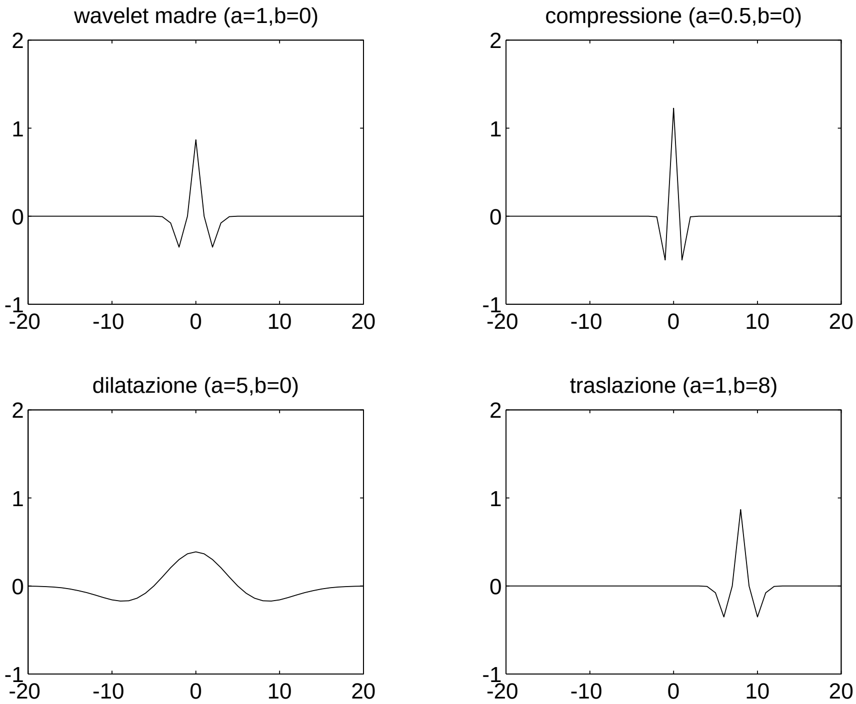

Figura 1.1: Wavelet madre e versioni scalate e traslate

Interessante, soprattutto da un punto di vista applicativo, è la versione discreta della trasformata wavelet ora introdotta. Questa è ottenuta imponendo che i parametri a e b assumano valori discreti. In particolare

$$a = a_0^j \qquad b = kb_0a_0^j \qquad \text{con } j, k \in \mathbb{Z} \text{ e } a_0, b_0 \text{ fissati.}$$

La versione discreta della trasformata wavelet è quindi

$$W_f(j,k) = \langle f, \psi_{j,k} \rangle = \int_{\mathbb{R}} f(x)\overline{\psi_{j,k}(x)}dx \tag{1.7}$$

dove

$$\psi_{j,k}(x) = \frac{1}{a_0^{j/2}}\psi(a_0^{-j}x - kb_0)$$

e che consente, adesso, di rappresentare la funzione f attraverso l'insieme (discreto) di coefficienti wavelet $\{W_f(j,k) = \langle f, \psi_{j,k} \rangle\}_{j,k \in \mathbb{Z}}$. A differenza del caso continuo, la trasformata discreta non è sempre invertibile poichè le wavelets della famiglia $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ non sono, in generale, linearmente indipendenti. Una formula di ricostruzione "esatta", che permette cioè di ricostruire f a partire dai coefficienti wavelet, può essere ottenuta nello spazio funzionale $L_2(\mathbb{R})$ scelti $a_0 = 1/2$ e $b_0 = 1$. La famiglia di wavelets

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

costituisce infatti un sistema ortonormale (vedi Appendice A) e si ha quindi la seguente formula di ricostruzione

$$f(x) = \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi_{j,k}(x) \quad \forall f \in L_2(\mathbb{R}) \tag{1.8}$$

nota anche come *serie wavelet*. Le basi wavelet ortonormali rivestono un ruolo particolarmente importante sia nella teoria che nelle applicazioni. Nei capitoli successivi tali basi saranno lo strumento utilizzato per le applicazioni statistiche descritte.

Una breve rassegna bibliografica è qui di seguito proposta con lo scopo di fornire al lettore indicazioni utili per l'approfondimento della teoria sulle wavelets. Sebbene le wavelets costituiscano una recente scoperta, la letteratura sull'argomento è già vasta. Come evidenziato all'inizio di questa sezione, i primi contributi sono stati dati da Grossmann e Morlet, [55], con la definizione della trasformata continua wavelet e da Mallat, [75], e Meyer, [78], con la sistematizzazione del caso discreto. Importante è stato il contributo di Ingrid Daubechies, [27], che ha costruito famiglie di funzioni wavelet particolarmente utili in ambito applicativo. Oltre a questi, altri testi che permettono una adeguata comprensione della teoria wavelet sono: *Ten Lectures on Wavelets* di Ingrid Daubechies, [29], che contiene gli aspetti matematici e che è il testo più conosciuto e *An Introduction to Wavelets* di Charles Chui, [21], che fornisce una introduzione all'argomento dal punto di vista della teoria dell'approssimazione. Si segnala anche *Wavelet Theory and Application* di Randy Young, [131], per una trattazione dell'argomento con particolare riferimento alla teoria del *signal processing* (elaborazione di segnali). Altri volumi, che raccolgono contributi di vari autori sia alla teoria che alle applicazioni delle wavelets, sono riportati in bibliografia.

Tra gli articoli di carattere divulgativo e introduttivo si segnalano *An Overview of Wavelet based Multiresolution Analyses* di B. Jawerth e W. Sweldens, [65], *Wavelets and Signal Processing* di O. Rioul e M. Vetterli, [98], e *Wavelets and Dilation Equations: a brief introduction* di G. Strang, [108].

Esistono poi monografie dedicate a specifici campi di applicazione delle wavelets, quali *Wavelet Applications in Chemical Engineering*, [84], *Wavelets in Geophysics*, [48] e *Wavelets in Statistics*, [5], e numeri speciali di riviste quali *IEEE Transactions on Information Theory*, [1], e *IEEE Transactions on Signal Processing*, [2]. Si segnala infine la rivista elettronica *Wavelet Digest* curata dal Dipartimento di Matematica della Università della Carolina del Sud (WAVELET@MATH.SCAROLINA.EDU) che rappresenta un ottimo e veloce canale di informazione.

Capitolo 2

Wavelets e Stima Non Parametrica

2.1 Premessa

Basi di funzioni wavelet possono essere utilizzate per costruire stimatori wavelet non parametrici di funzioni.

Particolare interesse è stato rivolto, fino ad oggi, alla stima di densità di probabilità e di funzioni di regressione. Nel primo caso, dato $(X_1, \cdots, X_n)$ un campione casuale estratto da una funzione di densità incognita f, si tratta di stimare f senza alcuna assunzione sulla sua forma. Nel caso invece di un modello non parametrico di regressione, dato un campione (X_i, Y_i), $i = 1, \cdots, n$ con X e Y variabili aleatorie, lo scopo è stimare il valore atteso di Y dato $X = x$, cioè la funzione di regressione $r(x) = E(Y|X = x)$.

Il capitolo è strutturato come segue: la prima parte contiene una breve rassegna sugli stimatori statistici classici maggiormente utilizzati per la stima di funzioni di densità e di regressione. Saranno poi introdotti i concetti fondamentali relativi alla teoria delle basi wavelet, strumento essenziale per la comprensione degli stimatori wavelet descritti subito dopo. Nell'ambito della stima di densità, gli stimatori wavelet saranno presentati nelle due versioni, lineare e a soglia, proposte in letteratura. Particolare attenzione sarà rivolta al problema della scelta del valore del parametro di scala al quale troncare lo sviluppo in serie wavelet della funzione di densità. Per questo problema, rilevante da un punto di vista applicativo, sarà proposto, nel caso lineare, un nuovo metodo che utilizza una stima wavelet dell'informazione di Fisher come funzione di penalità.

Gli stimatori saranno inoltre estesi al caso multivariato. La validità degli stimatori wavelet di densità sarà confermata dai risultati ottenuti nelle applicazioni a insiemi di dati di varia natura.

L'ultima parte del capitolo è dedicata alla costruzione di stimatori wavelet di funzioni di regressione di cui sarà realizzata una applicazione ad un insieme di dati economici confrontando i risultati ottenuti con quelli relativi all'utilizzo di stimatori classici.

Riferimenti bibliografici saranno dati nel corso del testo.

2.2 Breve Rassegna di Metodi Classici

Prima di introdurre gli stimatori wavelet di funzioni di densità e di regressione è utile ricordare brevemente quali sono gli stimatori classici più utilizzati e le loro principali proprietà.

2.2.1 Stimatori di Funzioni di Densità

Dato un campione casuale $(X_1, \cdots, X_n)$ estratto da una funzione di densità di probabilità incognita f, il problema è quello di stimare f. Molti stimatori sono stati proposti in letteratura; questa sezione contiene una breve rassegna di quelli che saranno più utili per il confronto con gli stimatori wavelet. Per una più completa esposizione vedi [64] [105].

Prima di procedere con la rassegna si ricorda che, indicato con $\hat{f}$ lo stimatore di una funzione f, $\hat{f}$ è detto *consistente in senso debole (forte)* se

$$\hat{f}(x) \to f(x) \quad \text{per} \quad n \to \infty$$

in probabilità (quasi certamente) per ogni x. Se f appartiene allo spazio $L_2(\mathbb{R})$ la bontà dello stimatore può essere misurata mediante misure puntuali, quale l'*errore quadratico medio*

$$EQM(x) = E_f[\hat{f}(x) - f(x)]^2 = var[\hat{f}(x)] + (bias[\hat{f}(x)])^2$$

dove $var[\hat{f}(x)] = E_f[\hat{f}(x) - E_f[\hat{f}(x)]]^2$ e $bias[\hat{f}(x)] = E_f[\hat{f}(x)] - f(x)$, o mediante misure globali, quali l'*errore quadratico medio integrale*

$$EQMI = \int E_f[\hat{f}(x) - f(x)]^2 dx$$

e l'*errore quadratico integrale*, cioè la norma L_2,

$$EQI = \int [\hat{f}(x) - f(x)]^2 dx.$$

Lo stimatore $\hat{f}$ è detto *consistente in media quadratica* se

$$EQM(x) \to 0 \quad \text{per} \quad n \to \infty$$

per ogni x, e *consistente in media quadratica integrale* se

$$EQMI \to 0 \quad \text{per} \quad n \to \infty.$$

Stimatori come Successioni Delta

Una classe molto ampia di stimatori non parametrici è quella degli stimatori espressi come *successione delta*

$$\hat{f}_m(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_m(x, X_i)$$

dove $\{\delta_m(x, y)\}$ è una successione delta, cioè una successione di funzioni tale che

$$\delta_m(x, y) \to \delta(x - y), \quad \text{per } m \to \infty.$$

Questi stimatori sono consistenti in media quadratica; risultati sul tasso di convergenza sono noti per la classe delle funzioni di densità di probabilità appartenenti a spazi di Sobolev e per funzioni soddisfacenti la condizione di Lipschitz, [128], (vedi Appendice A per le definizioni di spazi di Sobolev e di condizione di Lipschitz).

La maggior parte degli stimatori non parametrici utilizzati in statistica possono essere considerati casi particolari di questa classe. In particolare, è possibile ottenere istogrammi, stimatori con nucleo e stimatori espressi come serie di funzioni ortogonali semplicemente definendo in modo opportuno la successione delta. In questo contesto, particolare attenzione sarà rivolta ad istogrammi e stimatori in serie ortogonale, relativamente al caso di funzioni di densità appartenenti allo spazio $L_2(\mathbb{R})$ e con supporto compatto.

Istogrammi

La classe di stimatori più semplice è quella degli *istogrammi*. Sia $[a, b]$ il supporto di f, ovvero la chiusura dell'insieme in cui la funzione è diversa da zero. Tali stimatori possono essere ottenuti definendo

$$\delta_m(x, X_i) = \sum_{j=1}^{m}(t_{j+1} - t_j)^{-1} I_{T_j}(x) I_{T_j}(X_i)$$

dove $t_1, \cdots, t_m$ è una partizione equispaziata di $[a, b]$, $T_i = [t_i, t_{i+1})$ e I_{T_i} è la funzione indicatrice dell'i-esimo intervallo della partizione. Un istogramma è quindi dato da

$$\hat{f}(x) = \frac{1}{nh}\sum_{1}^{m} N_i I_{T_i}(x)$$

dove $h = (t_{i+1} - t_i)$ è la ampiezza comune degli intervalli della partizione ed N_i è il numero di valori campionari che cadono in T_i.

Si riporta anche l'estensione al caso d-dimensionale: se f ha supporto in $[a_1, b_1] \times \cdots \times [a_d, b_d]$, scegliendo una partizione di iper-rettangoli di ampiezze

$h_1 \times \cdots \times h_d$, lo stimatore diventa

$$\hat{f}(x) = \frac{1}{nh_1 \cdots h_d} \sum_{i_1, \cdots, i_d} N_{i_1,\ldots,i_d} I_{T_{1,i} \times \cdots \times T_{d,i}}(x)$$

dove $T_{j,i} = [t_{j,i}, t_{j,i+1})$, $j = 1, \cdots, d$ e $N_{i_1,\ldots,i_d}$ è il numero di valori campionari che cadono in $T_{i,d}$.

Gli istogrammi sono stimatori EQMI consistenti e l'EQMI può essere asintoticamente minimizzato scegliendo h; il valore asintotico ottimale di h dipende però dalla norma di f, quindi valori approssimati devono essere considerati a fini applicativi. Un valore approssimato tra i più utilizzati è quello suggerito da Scott, [104]

$$\hat{h}_k = 3.5 s_k n^{-1/(2+d)} \tag{2.1}$$

dove s_k è la deviazione standard campionaria della k-esima variabile. Questa scelta è ottimale se il campione è Gaussiano e determina una sovra-stima (*oversmoothing*) altrimenti.

Stimatori in Serie Ortogonale

Gli *stimatori in serie ortogonale* o *stimatori di proiezione ortogonale* furono introdotti nel 1962 da Čencov, [19]; la densità incognita è rappresentata in questo caso da uno sviluppo in serie del tipo

$$f(x) = \sum_{j \in \mathbb{Z}} a_j \psi_j(x),$$

dove $\{\psi_j,\ j \in \mathbb{Z}\}$ è un sistema ortogonale completo in $L_2(\mathbb{D})$, $\mathbb{D} \subset \mathbb{R}$ (vedi Appendice A). I coefficienti a_j sono definiti come

$$\begin{aligned} a_j &= \langle f, \psi_j \rangle \\ &= \int \psi_j(x) f(x) dx \\ &= E[\psi_j(X)] \end{aligned}$$

e possono essere stimati, in maniera non distorta, da

$$\hat{a}_j = \frac{1}{n} \sum_{i=1}^{n} \psi_j(X_i).$$

Lo stimatore di $f(x)$ è quindi

$$\hat{f}(x) = \sum_{j \in \mathbb{Z}} \hat{a}_j \psi_j(x).$$

Questo stimatore può non essere ben definito per un insieme infinito di indici poichè ha varianza infinita ed è EQI non consistente, cioè $\|\hat{f} - f\|_2$ non converge a zero per $n \to \infty$. La procedura più comune per ovviare a questo problema è quella di *tapering* (restringimento), che consiste nel considerare lo stimatore nella forma

$$\hat{f}(x) = \sum_{j \in \mathbb{Z}} b_j \hat{a}_j \psi_j(x)$$

dove $0 < b_j < 1$ è un peso simmetrico che comprime $\hat{a}_j$ verso l'origine. Per esempio, un numero finito di coefficienti $\hat{a}_j$ può essere selezionato scegliendo $b_j = 1$ per $j \in J$, J insieme finito, e 0 altrimenti ottenendo lo stimatore

$$\hat{f}(x) = \sum_{j \in J} \hat{a}_j \psi_j(x). \tag{2.2}$$

Se si suppone $J = [-m, m]$ questi stimatori possono essere ottenuti come casi particolari di stimatori espressi come successioni delta scegliendo

$$\delta_m(x, X_i) = \sum_{k=-m}^{m} \psi_k(x) \psi_k(X_i).$$

Naturalmente il grado di bontà di questi stimatori dipende dalla scelta di J, ovvero dal numero di termini nello sviluppo. Una *regola di arresto ottimale termine a termine* è stata proposta da Knonmal e Tarter, [73], e consiste nello scegliere J minimizzando l'EQMI stimato; altri autori hanno però dimostrato che questo porta a stime "oversmoothed" e che i risultati ottenuti empiricamente nel caso di distribuzioni multimodali non sono accettabili.

Alcuni dei sistemi ortogonali più utilizzati sono quelli di Fourier per funzioni con supporto compatto e quelli di Hermite in $\mathbb{R}$. Se f è una funzione di densità di probabilità continua e a supporto compatto in $[-\pi, \pi]$ allora lo stimatore in serie di Fourier di f è dato da

$$\hat{f}(x) = \sum_{k=-m}^{m} c_k e^{ikx}$$

con

$$e^{ikx} = cos(kx) + isin(kx)$$

e

$$c_k = \frac{1}{2\pi n} \sum_{i=1}^{n} e^{-ikX_i}.$$

Si ha la seguente

Proposizione 2.2.1 *Sia $f \in C^\alpha$ con $\alpha > 1$ in $[-\pi, \pi]$ e tale che $f^{(k)}(\pm\pi) = 0,\ k = 0, 1, \cdots, \alpha - 1$, allora*

$$E[\hat{f}(x) - f(x)]^2 = O(n^{-1+1/(2\alpha-1)}), \quad m = [n^{1/(2\alpha-1)}].$$

Lo stimatore in serie di Hermite per funzioni f appartenenti ad $L_2(\mathbb{R})$ è dato invece da

$$\hat{f}(x) = \sum_{k=-m}^{m} c_k h_k(x)$$

con

$$c_k = \frac{1}{n} \sum_{i=1}^{n} h_k(X_i)$$

e dove $h_k(x)$ sono le funzioni di Hermite

$$h_n(x) = c_n(x) H_n(x),\ n = 0, 1, 2, \cdots,$$

$$c_n(x) = \frac{e^{-x^2/2}}{(2^n n! \pi^{1/2})^{1/2}} \quad \text{e} \quad H_n(x) = (-1)^n e^{-x^2/2} \frac{d^n}{dx^n}(e^{-x^2}),$$

che formano un sistema ortonormale completo in $L_2(\mathbb{R})$. $H_n(x)$ è detto n-esimo polinomio di Hermite. Si ha

Proposizione 2.2.2 *Sia f tale che $(x - D)^r f \in L_2(\mathbb{R})$ per $r > 1$; allora*

$$E[\hat{f}(x) - f(x)]^2 = O(n^{-1+6/(6r+1)}), \quad m = n^{6/(6r+1)}$$

uniformemente per $x \in \mathbb{R}$.

Si osservi che per entrambi gli stimatori il tasso di convergenza si avvicina a $O(n^{-1})$ per $f \in C^\infty$ e che invece nel caso di funzioni f che siano solo continue la convergenza di EQM a zero non è assicurata. Infine, uno svantaggio presentato da questo tipo di stimatori è costituito dalla scarsa sensibilità nello stimare le proprietà locali di f.

2.2.2 Stimatori di Funzioni di Regressione

Siano X, Y variabili casuali con funzione di densità di probabilità congiunta $f(x, y)$ e sia $f_x(x)$ la funzione di densità di probabilità di X. Il valore atteso di Y dato $X = x$

$$r(x) = E(Y|X = x) = \int \frac{y f(x, y)}{f_x(x)} dy$$

è detto *funzione di regressione*. Dato un campione (X_i, Y_i) $i = 1, \cdots, n$ lo scopo è quello di costruire uno stimatore di r. Naturalmente, ogni stimatore di funzioni di densità di probabilità su $\mathbb{R}^d$ può essere utilizzato.

Uno degli stimatori classici più utilizzati è lo stimatore con nucleo moltiplicativo di Nadaraya e Watson, [87] [130],

$$\hat{r}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i K_{h_x}(x - X_i)}{\frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - X_i)} \tag{2.3}$$

dove $K_h(x - X_i)$ è una funzione nucleo tale che

$$\int_{\mathbb{R}^d} K_h(x)dx = 1$$

e dove il parametro h è detto ampiezza di banda. Questo stimatore è ottenuto considerando che

$$\hat{f}(x, y) = \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - X_i)K_{h_y}(y - Y_i)$$

e che, poichè $\int K_h(t)dt = 1$,

$$\int \hat{f}(x, y)dy = \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - X_i)\int K_{h_y}(y - Y_i)dy = \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - X_i).$$

Scegliendo K_h nucleo di ordine 2, cioè tale che $\int tK_h(t)dt = 0$, si ha

$$\int yK_h(y - Y_i)dy = Y_i$$

e quindi lo stimatore (2.3).

Il valore dell'ampiezza di banda può essere selezionato mediante validazione incrociata, ovvero h è scelto in modo tale da minimizzare

$$CV(h) = n^{-1}\sum_{1}^{n}(Y_i - \hat{r}_{(i)}(X_i))^2$$

dove $\hat{r}_{(i)}(x)$ è lo stimatore *leave one out* ottenuto calcolando $\hat{r}$, come funzione di h e x, con l'i-esimo valore campionario rimosso.

Uno stimatore alternativo a quello kernel è lo stimatore *k-Nearest Neighbor* (k-vicini-più prossimi) che può essere espresso come

$$\hat{f}(x) = \sum_{i=1}^{n} y_i w_i(x, k)$$

dove $w_i(x, k)$ è una successione di pesi determinati in base ad una funzione di probabilità e che assumono valore zero se $i > k$. Tale successione stabilisce il contributo dei valori y_i sulla base del grado di vicinanza, misurato secondo una opportuna metrica, tra x e i punti x_j del campione.

2.3 Sistemi Wavelet Ortonormali

Questa sezione contiene una breve descrizione delle basi wavelet ortonormali, strumento necessario per la comprensione degli stimatori wavelet di densità e di regressione che saranno in seguito presentati. Come già evidenziato nella breve introduzione alle wavelets riportata nel primo capitolo, la famiglia di funzioni

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),$$

ottenute come traslazioni e dilatazioni di una wavelet madre ψ, costituisce una base ortonormale in $L_2(\mathbb{R})$. La costruzione di tali basi è stata rigorosamente sistematizzata nel contesto generale della teoria della analisi a più livelli di risoluzione. La sezione seguente illustra tale teoria. Di seguito, saranno brevemente descritte le wavelet di Daubechies che sono, tra le famiglie di wavelets proposte in letteratura, le più utilizzate in ambito applicativo.

2.3.1 Analisi a più Livelli di Risoluzione

Una *analisi a più livelli di risoluzione* (multiresolution analysis), in seguito indicata con l'acronimo MRA, è un approccio formale generale alla costruzione di basi ortonormali in $L_2(\mathbb{R})$. L'idea è quella di rappresentare una funzione f come limite di successive approssimazioni ognuna delle quali è una versione più "regolare" di f. Tali approssimazioni corrispondono a differenti livelli di risoluzione.

Formalmente, una MRA, [75] [76], di $L_2(\mathbb{R})$ è una successione di sottospazi lineari e chiusi $\{V_j, j \in \mathbb{Z}\}$ di $L_2(\mathbb{R})$ tale che

$$(i) \qquad V_j \subset V_{j+1}, \quad j \in \mathbb{Z}$$

$$(ii) \qquad \overline{\cup_j V_j} = L_2(\mathbb{R})$$

$$(iii) \qquad \cap_j V_j = \{0\}$$

$$(iv) \quad f(x) \in V_j \iff f(2x) \in V_{j+1}, \quad f(x) \in V_j \Rightarrow f(x+k) \in V_j$$

(v) Esiste una funzione $\phi \in V_0$ tale che la famiglia $\{\phi(x-k), k \in \mathbb{Z}\}$ è una base ortonormale in V_0.

La proprietà (iv) è particolarmente utile per capire la struttura della successione di spazi V_j. Innanzi tutto, essa può essere riformulata come

$$f(x) \in V_0 \iff f(2^j x) \in V_j.$$

Quindi, ciascuno degli spazi V_j è ottenuto da V_0 come versione "scalata" secondo la *scala* (o *livello*) j. Poichè $V_j \subset V_{j+1}$, valori più grandi di j corrisponderanno a scale più fini e valori più bassi a scale più grossolane. Il fattore 2^j, che permette il passaggio da V_0 a V_j, è detto invece *risoluzione*. Definite le funzioni

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \tag{2.4}$$

allora, per le proprietà (iv) e (v), la famiglia $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ è una base ortonormale in V_j. La costante $2^{j/2}$ è quella di normalizzazione.

Poichè $\phi \in V_0$ e $V_0 \subset V_1$, si ha

$$\phi(x) = \sqrt{2}\sum_l h_l \phi(2x - l). \tag{2.5}$$

La funzione ϕ è detta *funzione di scala*. L'equazione (2.5) è detta *equazione di dilatazione* e i coefficienti h_l, dati da $h_l = \langle \phi, \phi_{1,l}\rangle$, sono detti *coefficienti di filtro*. Data $f \in L_2(\mathbb{R})$, un'approssimazione di f alla risoluzione 2^j (o alla scala j) è definita come la proiezione ortogonale P_j di f in V_j

$$P_j f(x) = \sum_k \langle f, \phi_{j,k}\rangle \phi_{j,k}(x). \tag{2.6}$$

La proprietà (ii) assicura che $\lim_{j\to\infty} P_j f = f$ per ogni $f \in L_2(\mathbb{R})$. Ogni funzione $f \in L_2(\mathbb{R})$ viene quindi considerata come limite di approssimazioni successive con risoluzione 2^j. Poichè $V_j \subset V_{j+1}$, quando la scala aumenta da j a $j+1$, l'approssimazione al livello più fine $j+1$ contiene tutte le informazioni su f che sono disponibili al livello inferiore j ed è, in particolare, ottenuta aggiungendo a $P_j f$ informazioni aggiuntive su f dette *dettagli*; tali dettagli possono essere modellati come complemento ortogonale di V_j in V_{j+1}. Sia W_j tale complemento ortogonale, ovvero $V_j \oplus W_j = V_{j+1}$ e $W_j \perp W_{j'}$ per $j \neq j'$ (il simbolo $\oplus$ è quello di somma diretta, vedi Appendice A). E' possibile dimostrare che

$$L_2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j. \tag{2.7}$$

Si definisca adesso *wavelet madre* una funzione $\psi \in W_0$ tale che la famiglia $\{\psi(x-k), k \in \mathbb{Z}\}$ sia una base ortonormale in W_0. Allora, date le funzioni

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),$$

la famiglia $\{\psi_{j,k}(x), k \in \mathbb{Z}\}$ costituisce una base ortonormale per W_j e, per la (2.7), la famiglia

$$\{\psi_{j,k}(x), j, k \in \mathbb{Z}\} \tag{2.8}$$

una base ortonormale per $L_2(\mathbb{R})$. Quindi, nel passare dalla scala j alla $j+1$, si ha

$$P_{j+1}f = P_j f + \sum_k \langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

Inoltre, poichè $W_0 \subset V_1$, la wavelet madre ψ può essere espressa come

$$\psi(x) = \sqrt{2}\sum_l g_l \phi(2x - l) \tag{2.9}$$

dove i coefficienti g_l sono scelti come $g_l = (-1)^l \overline{h_{1-l}}$.

Si osservi che una MRA può essere costruita a partire dalla scelta della funzione di scala ϕ. Tale scelta determina lo spazio V_0 come generato da $\{\phi(\cdot - k)\}_{k\in\mathbb{Z}}$, gli spazi V_j, generati da V_0, ed infine la wavelet madre ψ dalla (2.9).

L'importanza di una struttura di MRA è costituita dal fatto che, utilizzando le basi ortonormali wavelet, ogni funzione $f \in L_2(\mathbb{R})$ può essere rappresentata in *serie wavelet* come

$$f(x) = \sum_{j,k\in\mathbb{Z}} d_{jk}\psi_{jk}(x). \tag{2.10}$$

Data la struttura degli spazi V_j e W_j, è possibile esprimere la (2.10) come

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0 k}\phi_{jk}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x). \tag{2.11}$$

Quest'ultima relazione può essere vista come una proiezione ortogonale di f in V_{j_0}, cioè una approssimazione di f, più un insieme di proiezioni di f in W_j, per $j > j_0$, che aggiungono dettagli su f a questa approssimazione.

I coefficienti di scala $\{c_{j,k}\}$ e quelli wavelet $\{d_{j,k}\}$ sono definiti come

$$c_{j,k} = \langle f, \phi_{j,k}\rangle = \int f(y)\overline{\phi_{j,k}(y)}dy \tag{2.12}$$

$$d_{j,k} = \langle f, \psi_{j,k}\rangle = \int f(y)\overline{\psi_{j,k}(y)}dy. \tag{2.13}$$

I coefficienti $c_{j,k}$ sono detti *coefficienti di scala* e quelli $d_{j,k}$ *coefficienti wavelet*. L'identità di Parseval connette tali coefficienti all'energia di f, cioè alla norma L_2 della funzione:

$$\int |f(x)|^2 dx = \sum_k |c_{j_0,k}|^2 + \sum_{j_0}^{+\infty}\sum_k |d_{j,k}|^2.$$

Si osservi che, poichè $V_{\bar{j}} = \bigoplus_{i=-\infty}^{\bar{j}-1} W_i$, troncare la serie wavelet (2.10) ad una scala $\bar{j}-1$ fissata equivale a considerare la proiezione di f in $V_{\bar{j}}$, cioè

$$\sum_{j=-\infty}^{\bar{j}-1} \sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x) = \sum_{k\in\mathbb{Z}} c_{\bar{j},k}\phi_{\bar{j}k}(x) = P_{\bar{j}}f(x).$$

In altre parole, la somma delle proiezioni della funzione f negli spazi W_j per $j = -\infty, \cdots, \bar{j}-1$ equivale alla determinazione della proiezione di f nello spazio $V_{\bar{j}}$.

Particolarmente utile nelle applicazioni è il concetto di MRA r-regolare: una funzione g in $\mathbb{R}$ è detta *r-regolare* se è r-volte differenziabile e soddisfa

$$\exists C_{k,p} > 0 \;:\; |g(x)^{(k)}| \leq C_{k,p}(1+|x|)^{-p}, \quad k = 1,\cdots,r;\; p \in \mathbb{Z}$$

cioè $g \in S_r$ (vedi Appendice A). Una MRA è detta r-regolare se e soltanto se ϕ è r-regolare. In tal caso anche ψ appartiene a S_r. L'utilità di questo tipo di MRA è dovuta al fatto che, data una MRA r-regolare, la wavelet madre ψ possiede r *momenti nulli*, ovvero è ortogonale a tutti i polinomi di grado minore o uguale ad r

$$\int_{\mathbb{R}} x^k\psi(x)dx = 0, \quad k = 0,1,\cdots,r$$

e che i coefficienti wavelet $d_{j,k}$ hanno un decadimento del tipo

$$d_{j,k} = \int_{\mathbb{R}} f(x)\psi_{j,k}(x)dx \leq C2^{-jr}. \tag{2.14}$$

Come già precedentemente osservato, una MRA può essere definita a partire dalla scelta della funzione di scala ϕ, come soluzione dell'equazione di dilatazione (2.5), e determinando poi la wavelet madre ψ attraverso l'equazione (2.9). Vari sono i metodi proposti per la costruzione di ϕ e, conseguentemente, di ψ, [108]. Qui di seguito è riportata una breve descrizione del metodo che sarà poi utilizzato nelle applicazioni degli stimatori wavelet e che è noto con il nome di *algoritmo a cascata* (cascade algorithm).

Utilizzando l'equazione di dilatazione (2.5) è possibile determinare ricorsivamente i valori che ϕ assume in un insieme finito di punti. Più precisamente, si consideri una successione finita di coefficienti h_l, $l = 0,\cdots,2N-1$, N fissato. La scelta di un numero finito di coefficienti implica che ϕ sia nulla al di fuori dell'intervallo $(0, 2N-1)$ e che, in particolare, $\phi(0) = \phi(2N-1) = 0$. L'equazione di dilatazione (2.5) costituisce, per $x = 1,2,\cdots,2N-2$, un sistema lineare omogeneo la cui risoluzione determina i valori di ϕ negli interi $0,1,\cdots,2N-1$. Tali valori possono essere utilizzati per calcolare quelli nei punti $1/2, 3/2, \cdots$, poi $1/4, 3/4, \cdots$, eccetera, direttamente dalla equazione di dilatazione. Ottenuti in tal modo i valori di ϕ nell'insieme finito di punti diadici $intero/2^n$, $n = 0, 1, \cdots, N$, N fissato, i corrispondenti valori di ψ possono essere calcolati usando l'equazione (2.9). Dal punto di vista applicativo, l'implementazione dell'algoritmo a cascata è, rispetto ad altri metodi utilizzabili, particolarmente semplice e veloce.

E' importante osservare che questo metodo non determina una forma analitica di ϕ e ψ nè una loro approssimazione; noti i valori che ϕ assume in un insieme finito di punti diadici, il valore in un punto arbitrario non compreso in tale insieme può essere calcolato usando metodi di interpolazione.

Le basi ortonormali wavelet sono facilmente estendibili allo spazio a d dimensioni $L_2(\mathbb{R}^d)$. Vari sono i metodi proposti. Il tipo di costruzione al quale verrà fatto riferimento nelle sezioni successive è una estensione del caso univariato concettualmente semplice che presenta tuttavia una qualche complessità di calcolo. Per non appesantire l'esposizione, si ricorda soltanto che, mentre nel caso unidimensionale una MRA è completamente determinata da una funzione di scala ϕ e da una wavelet madre ψ, la generalizzazione al caso bivariato coinvolge una funzione di scala definita utilizzando la ϕ univariata come

$$\phi(x_1, x_2) = \phi(x_1)\phi(x_2)$$

e 3 wavelet madri definite come

$$\psi^{(1)}(x_1, x_2) = \phi(x_1)\psi(x_2)$$

$$\psi^{(2)}(x_1, x_2) = \psi(x_1)\phi(x_2)$$

e

$$\psi^{(3)}(x_1, x_2) = \psi(x_1)\psi(x_2)$$

dove ψ è la wavelet madre unidimensionale associata a ϕ.

La generalizzazione al caso d-dimensionale coinvolge una funzione di scala e $2^d - 1$ wavelet madri Particolarmente utile sarà la rappresentazione in serie wavelet d-dimensionale: data una funzione f in $L_2(\mathbb{R}^d)$, la sviluppo wavelet di f è dato da

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{+\infty} \sum_{\mathbf{k}} \sum_{l=1}^{2^d-1} d_{j,\mathbf{k}}^{(l)}\psi_{j,\mathbf{k}}^{(l)}(\mathbf{x}),$$

dove $\mathbf{k} \in \mathbb{Z}^d$, $\mathbf{k} = (k_1, ..., k_d)$ e

$$\phi_{j_0,\mathbf{k}}(\mathbf{x}) = 2^{j_0 d/2}\phi(2^{j_0}x_1 - k_1, ..., 2^{j_0}x_d - k_d)$$

$$\psi_{j,\mathbf{k}}^{(l)}(\mathbf{x}) = 2^{jd/2}\psi^{(l)}(2^j x_1 - k_1, ..., 2^j x_d - k_d), \quad l = 1, \cdots, 2^d - 1.$$

2.3.2 Wavelets di Daubechies

Molte sono le famiglie di funzioni wavelet proposte in letteratura. Questa sezione riporta una breve descrizione di quelle costruite da Ingrid Daubechies, [27], che, per la loro caratteristica di supporto compatto, sono le più utilizzate in ambito applicativo. Il supporto compatto assicura infatti una buona localizzazione nel tempo.

Tali famiglie saranno indicate con il simbolo $Daub\#N$, $N \in \mathbb{N}$. La wavelet madre ψ della famiglia $Daub\#N$ ha supporto compatto di ampiezza $2N-1$ ed ha N momenti nulli. Per N sufficientemente grande, ϕ e ψ appartengono allo spazio $C^{\mu N}$ con $\mu \approx 0.2$ [1]; al crescere di N si hanno quindi funzioni wavelet con supporto maggiore ma più regolari. Inoltre, per la (2.14), un alto numero di momenti nulli implica che i coefficienti wavelet decadano a zero per alti valori del parametro di scala; questo consente di rappresentare caratteristiche locali di funzioni utilizzando un minor numero di coefficienti ed assicura quindi una alta compressibilità. Infine, in [29] Daubechies dimostra con un esempio come questa proprietà assicuri che i coefficienti wavelet alle scale piu fini assumono valori grandi in corrispondenza di singolarità della funzione o delle sue derivate.

Una particolarità di queste famiglie di wavelets è data dal fatto che, tranne che per il caso $N=1$, non esiste forma analitica di queste funzioni. Le proprietà di ortogonalità e quella dei momenti nulli possono infatti essere formulate utilizzando i coefficienti di filtro h_l in (2.5) come

$$\sum_l h_l = \sqrt{2} \qquad \sum_l (-1)^l h_l = 0 \qquad \text{ortogonalità}$$

$$\sum_l h_l (-1)^k l = 0, \qquad k = 0, 1, \cdots, r \qquad r \text{ momenti nulli.}$$

Daubechies ha definito le wavelets $Daub\#N$ assegnando i valori dei coefficienti h_l in modo tale che tali proprietà siano soddisfatte. In particolare, le wavelets della famiglia $Daub\#N$ sono caratterizzate da $2N$ coefficienti $h_0, \cdots, h_{2N-1}$. Dati tali coefficienti, i valori di ϕ e ψ in un punto arbitrario possono essere calcolati utilizzando l'algoritmo a cascata precedentemente descritto. Il numero finito di coefficienti h_l assicura il supporto compatto; in particolare $Supp(\phi) = [0, 2N-1]$ e $Supp(\psi) = [1-N, N]$. La figura 2.1 riporta i grafici della funzione di scala e della wavelet madre per alcune delle famiglie $Daub\#N$.

La famiglia di wavelets ottenuta per $N=1$ costituisce una eccezione: la funzione di scala è determinata dai coefficienti di filtro $h_0 = h_1 = \frac{1}{\sqrt{2}}$ ed è la

[1] Per $\alpha = n + \beta$, $n \in \mathbb{N}$, $0 < \beta < 1$, una funzione f appartiene allo spazio C^α se ha n derivate continue e se la derivata di ordine n è Holder continua con esponente β, cioè tale che $|f^{(n)}(x) - f^{(n)}(x+t)| \leq C|t^\beta| \; \forall x, t$

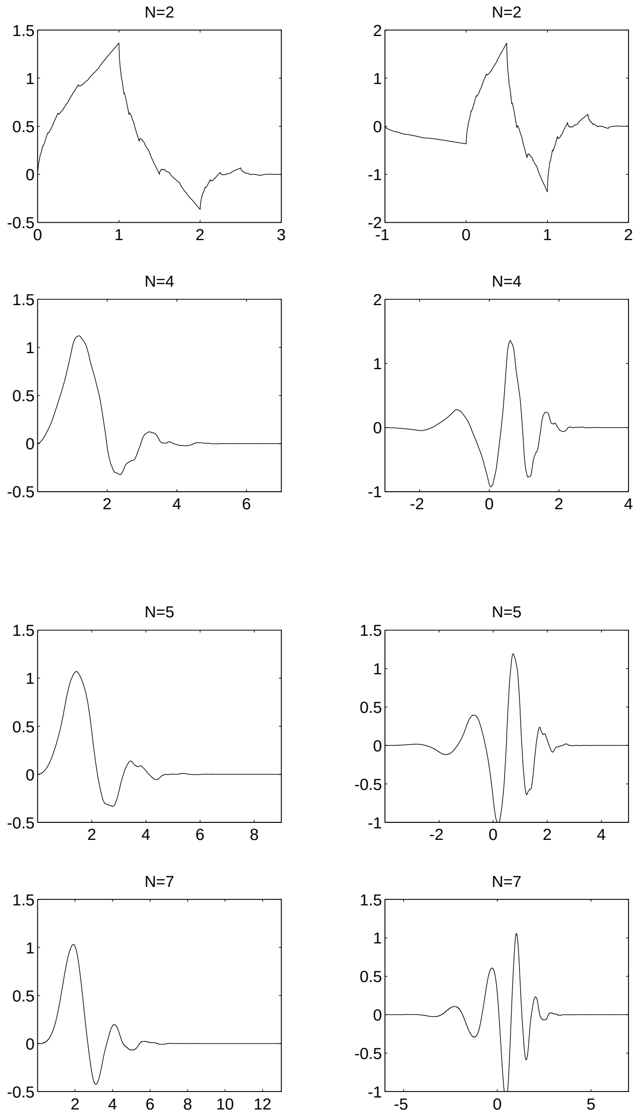

Figura 2.1: Funzione di scala ϕ (prima colonna) e wavelet madre ψ (seconda colonna) per le wavelets di Daubechies $Daub\#N$, $N = 2, 4, 5, 7$.

funzione indicatrice dell'intervallo $[0,1)$

$$\phi(x) = I_{[0,1)}(x).$$

La famiglia $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ è quindi costituita da funzioni costanti a tratti (in particolare $\phi_{j,k}$ è costante in $[2^{-j}k, 2^{-j}(k+1))$). La wavelet madre è data da

$$\psi(x) = \begin{cases} 1 & 0 \leq x < 1/2 \\ -1 & 1/2 \leq x < 1 \\ 0 & \text{atrimenti} \end{cases}$$

e si ha $g_0 = -g_1 = \frac{1}{\sqrt{2}}$. Questa famiglia di funzioni, note come wavelets di Haar, era già nota prima della formulazione della teoria sulle wavelets, [56], ed è considerata un prototipo. Le wavelets di Haar sono la famiglia più semplice che si possa costruire. La MRA associata a tale famiglia non è regolare.

Esistono altre famiglie di funzioni wavelet costruite da Ingrid Daubechies. Una classe particolarmente importante è quella delle *Coiflet*, famiglie ortonormali di funzioni wavelet con supporto compatto dove anche ϕ (non solo ψ) ha N momenti nulli, [29] [31]. La costruzione di queste wavelets è stata richiesta da Ronald Coifman, da cui il nome Coiflet. Tali wavelets saranno indicate con il simbolo $Coiflet\#N$.

2.4 Stimatori Wavelet di Funzioni di Densità

Le basi wavelet introdotte nella sezione precedente possono essere utilizzate per costruire stimatori di funzioni di densità di probabilità. Nel seguito, introdotta la forma generale dello stimatore wavelet, ne verranno presentate due diverse versioni, una lineare ed una a soglia. L'applicazione dello stimatore wavelet alla radice quadrata della funzione di densità determinerà stimatori positivi con integrale uguale ad uno.

Da un punto di vista applicativo, è necessario determinare criteri che permettano di calcolare il valore ottimale del parametro di scala al quale troncare la rappresentazione in serie wavelet della funzione di densità. A tal scopo sarà proposto un nuovo metodo che utilizza una stima wavelet dell'informazione di Fisher come funzione di penalità. Gli stimatori wavelet saranno poi estesi al caso di funzioni di densità multivariate. Infine, saranno presentate alcune applicazioni realizzate su insiemi di dati di diversa natura.

2.4.1 Stimatori Lineari e Stimatori a Soglia

Sia $(X_1, \cdots, X_n)$ un campione casuale estratto da una funzione di densità $f \in L_2(\mathbb{R})$. Utilizzando le basi ortogonali wavelet la rappresentazione di f

in serie ortogonale di funzioni wavelet è

$$f(x) = \sum_{j,k} d_{j,k}\psi_{j,k}(x) = \sum_k c_{j_0,k}\phi_{j_0,k}(x) + \sum_{j_0}^{+\infty}\sum_k d_{j,k}\psi_{j,k}(x) \tag{2.15}$$

dove i coefficienti $\{c_{j,k}\}$ e $\{d_{j,k}\}$ sono definiti come in (2.12) e (2.13). Poichè f è una funzione di densità tali coefficienti possono essere stimati in maniera non distorta, come nel caso generale degli stimatori espressi come serie ortogonale, considerando i coefficienti empirici

$$\hat{c}_{j_0,k} = n^{-1}\sum_{i=1}^{n}\phi_{j_0,k}(X_i) \tag{2.16}$$

$$\hat{d}_{j,k} = n^{-1}\sum_{i=1}^{n}\psi_{j,k}(X_i). \tag{2.17}$$

Lo stimatore wavelet di f è dunque

$$\hat{f}(x) = \sum_k \hat{c}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j_0}^{j_1}\sum_k \hat{d}_{j,k}\psi_{j,k}(x) \tag{2.18}$$

dove j_0 e j_1 devono essere opportunamente scelti. Due diverse versioni di questo stimatore sono state proposte in letteratura:

Stimatori Lineari

Lo stimatore (2.18) è una funzione che appartiene allo spazio

$$V_{j_1+1} = V_{j_0} \oplus W_{j_0} \oplus \cdots \oplus W_{j_1}.$$

E' quindi possibile esprimere tale stimatore utilizzando la base in V_{j_1+1} data da $\{\phi_{j_1+1,k}\ ;\ k \in \mathbb{Z}\}$ come

$$\hat{f}(x) = \sum_k \hat{c}_{\bar{j},k}\phi_{\bar{j},k}(x), \quad \bar{j} = j_1 + 1 \tag{2.19}$$

ovvero come proiezione lineare ortogonale di f in $V_{\bar{j}}$. Il valore $\bar{j}$ deve essere scelto con un opportuno criterio. Lo stimatore (2.19) può essere espresso come

$$\hat{f}(x) = \sum_k n^{-1}\sum_{i=1}^{n}\phi_{\bar{j},k}(X_i)\phi_{\bar{j},k}(x) = n^{-1}\sum_{i=1}^{n}\sum_k \phi_{\bar{j},k}(x)\phi_{\bar{j},k}(X_i),$$

ovvero

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} q_{\bar{j}}(x, X_i)$$

dove

$$q_{\bar{\jmath}}(x, X_i) = \sum_k \phi_{\bar{\jmath},k}(x)\phi_{\bar{\jmath},k}(X_i). \tag{2.20}$$

La funzione $q_{\bar{\jmath}}(x, y)$ è detta *nucleo autoriproducente* (reproducing kernel), [8], in $V_{\bar{\jmath}}$ e verifica, [126],

$$q_{\bar{\jmath}}(x, y) = 2^{\bar{\jmath}} q_0(2^{\bar{\jmath}} x, 2^{\bar{\jmath}} y)$$

con

$$q_0(x, y) = \sum_k \phi(x - k)\overline{\phi(y - k)}$$

nucleo autoriproducente in V_0 tale che

$$q_0(x + k, y + k) = q_0(x, y)$$

e

$$\int_{\mathbb{R}} q_0(x, y)dy = 1.$$

E' interessante osservare che considerando la base wavelet di Haar, dove $\phi(x) \equiv I_{[0,1)}(x)$, lo stimatore wavelet (2.19) è un istogramma con nodi $2^{-\bar{\jmath}}k$, k intero. Infatti

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} q_{\bar{\jmath}}(x, X_i) = \sum_k \hat{c}_{\bar{\jmath},k} 2^{\bar{\jmath}/2} \phi(2^{\bar{\jmath}} x - k)$$

dove

$$\hat{c}_{\bar{\jmath},k} = n^{-1} \sum_{i=1}^{n} 2^{\bar{\jmath}/2} \phi(2^{\bar{\jmath}} X_i - k),$$

e poichè ϕ è la funzione indicatrice dell'intervallo $[0, 1)$ si ha

$$\hat{c}_{\bar{\jmath},k} = \frac{2^{\bar{\jmath}/2}}{n} \#\{X_i \in [k2^{-\bar{\jmath}}, (k + 1)2^{-\bar{\jmath}})\}$$

e

$$\hat{f}(x) = \frac{2^{\bar{\jmath}}}{n} \sum_k n_k I_{T_k}(x)$$

dove $T_k = [k2^{-\bar{\jmath}}, (k + 1)2^{-\bar{\jmath}})$ e n_k è il numero di valori campionari che cadono in T_k.

Lo stimatore (2.19) è stato studiato da diversi autori, quali Walter, [125] [126] [127], Kerkyacharian e Picard, [70] [71], Antoniadis e Carmona, [3]. Poichè $\{q_{\bar{\jmath}}(y, x);\ \bar{\jmath} \in \mathbb{Z}\}$ è una successione delta, tale stimatore rientra nella classe degli stimatori espressi come successione delta introdotti all'inizio del capitolo.

Si osservi inoltre che la famiglia di funzioni $\{\phi_{\bar{\jmath},k}(x)\}_{k\in\mathbb{Z}}$ costituisce una base ortonormale in $V_{\bar{\jmath}}$ e che i coefficienti $\hat{c}_{\bar{\jmath},k}$ in (2.19) non dipendono da x; lo stimatore wavelet lineare rientra quindi, più specificatamente, nella classe degli stimatori espressi come serie di funzioni ortogonali. Tuttavia, le proprietà di localizzazione nel tempo e nella frequenza, tipiche delle basi wavelet, rendono lo stimatore wavelet "migliore" di stimatori ortogonali che utilizzano basi come quella di Fourier: dovendo rappresentare funzioni irregolari, lo stimatore wavelet richiede un numero di coefficienti minore mostrando quindi una sensibilità maggiore nella stima delle caratteristiche locali.

Lo stimatore wavelet è asintoticamente non distorto poichè

$$E(\hat{f}(x)) - f(x) = \int q_{\bar{\jmath}}(x,y)f(y)dy - f(x)$$

converge a zero per funzioni f continue.

Per quanto concerne le proprietà di consistenza, si hanno i seguenti risultati, [126]

Proposizione 2.4.1 *Sia $\phi \in S_r$ (vedi Appendice A) e tale che per qualche $\lambda \geq 1$ si ha*

$$\mathcal{F}_\phi(\omega) = 1 + O(|\omega|^\lambda), \quad per \;\; \omega \to 0$$

e

$$(Z\phi)(t,\omega) = e^{-i\omega t}(1 + O(|\omega|^\lambda) \quad uniformemente \; per \; \omega \to 0$$

dove $Z_\phi(t,\omega) = \sum_k e^{-i\omega k}\phi(t-k)$ è la trasformata di Zak di ϕ, [66]. Sia inoltre f una funzione di densità di probabilità continua e limitata. Allora

$$(i) \qquad E|\hat{f}(x) - f(x)|^2 \to 0,$$

uniformemente in sottoinsiemi compatti per $\bar{\jmath} \to \infty$ e $\bar{\jmath} = O(logn)$

$$(ii) \qquad se \; f \in H^\alpha, \; \alpha > \lambda + 1/2, \; \bar{\jmath} \approx (\log 2 \log n)/(2\lambda + 1), \quad allora$$

$$E|\hat{f}(x) - f(x)|^2 = O(n^{-2\lambda/(2\lambda+1)}) \quad e \quad \int E|\hat{f}(x) - f(x)|^2 dx = O(n^{-2\lambda/(2\lambda+1)}).$$

(iii) se $r \geq 1$, allora $\hat{f}$ è consistente in senso forte per $\bar{\jmath} \to \infty$, $\bar{\jmath} = O(\log n)$ e

$$\hat{f}(x) - f(x) = O\Big((n^{-\lambda/(\lambda+1)} \log\log n)^{1/2}\Big) \qquad q.c.$$

per $\bar{\jmath} = O(\log n/2(\lambda+1)\log 2)$.

Tali proprietà discendono da quelle relative alla convergenza della serie wavelet (2.15); per una ottima esposizione di queste ed un confronto con quelle relative a serie più classiche, quali quella di Fourier e di Hermite, vedi [125] [126].

I tassi di convergenza in (ii) sono ottenuti considerando separatamente varianza e distorsione dello stimatore; in particolare

$$var[\hat{f}(x)] = O(2^{2\bar{j}}/n) \quad \text{e} \quad bias[\hat{f}(x)] = O(2^{-\bar{j}\lambda}). \tag{2.21}$$

Si osservi che le proprietà di consistenza dello stimatore lineare wavelet sono più forti di quelle relative agli stimatori in serie classici prima introdotti. In particolare, lo stimatore wavelet è consistente in media quadratica nella classe della funzioni di densità continue ed il tasso di convergenza aumenta per densità C^{∞}.

Le proprietà dello stimatore (2.19) sono state studiate anche nel caso di funzioni f appartenenti a spazi di Besov, [70] [71]. Per non appesantire l'esposizione tali proprietà sono riportate nella Appendice B.

Stimatori a Soglia

Si consideri ancora lo stimatore wavelet nella forma (2.18). Come già osservato nella sezione 2.3.1, tale espressione può essere vista come una approssimazione grossolana di f alla scala j_0 più dettagli ad alcuni livelli più fini che migliorano l'approssimazione. In termini più statistici, tale stimatore consente una analisi della funzione che è allo stesso tempo globale (in termini delle funzioni di scala $\phi_{j_0,k}$) e locale (le wavelets $\psi_{j,k},\ j = j_0, \cdots, j_1$).

Donoho, Johnstone, Kerkyacharian e Picard, [44], propongono uno stimatore wavelet a soglia ottenuto ponendo a zero alcuni dei coefficienti wavelet $\hat{d}_{j,k}$ più piccoli. L'idea sottostante è che tali coefficienti catturano dettagli su f che possono essere trascurati. Questa procedura può essere considerata un particolare tipo di tapering (restringimento). Lo stimatore wavelet a soglia è ottenuto scegliendo un particolare valore di soglia λ e ponendo a zero i coefficienti che sono in valore assoluto minore di λ (hard threshold, soglia forte). Una seconda tecnica proposta prevede anche un avvicinamento dei coefficienti più grandi, cioè una loro compressione, verso lo zero (soft threshold, soglia dolce). Gli autori propongono un valore di soglia dipendente da j dato da $\lambda_j = C\sqrt{j/n}$ e dimostrano che lo stimatore wavelet ottenuto considerando i coefficienti wavelet "compressi" mediante le tecniche di threshold sopra descritte è in certi casi "ottimo" (vedi Appendice B).

Lo stimatore a soglia è stato studiato anche da altri autori che hanno essenzialmente proposto tecniche di threshold alternative. Tra questi, Delyon e Juditsky, [35], propongono ancora threshold di tipo hard e soft ma con parametro di soglia $\lambda_j = C\sqrt{max(j - j_0, 0)/n}$. Pinheiro e Vidakovic, [95], utilizzano invece una tecnica di threshold proposta da Goel e Vidakovic, [53]: analizzando la curva di Lorentz relativa ai coefficienti wavelet empirici, quelli con energia più piccola di C volte l'energia media vengono posti uguali a zero.

Le tecniche di threshold dei coefficienti wavelet saranno ampiamente utilizzate nel prossimo capitolo per lo smoothing (lisciamento) di serie storiche.

Si osservi che

$$V_{j_1+1} = V_{j_0} \oplus \bigoplus_{j_0}^{j_1} W_j,$$

quindi lo stimatore lineare (2.19) per $\bar{\jmath} = j_1 + 1$ è equivalente a quello a soglia qualora si consideri una approssimazione grossolana di f in V_{j_0} e si aggiungano dettagli fino al livello j_1 *senza* modificare nessuno dei coefficienti wavelet. Lo stimatore threshold è quindi, in generale, più parsimonioso di quello lineare da un punto di vista computazionale.

2.4.2 Stimatori Bona Fide

Gli stimatori wavelet presentati possono assumere a volte valori negativi, soprattutto in corrispondenza di valori che si trovano nelle code delle distribuzioni o nelle parti di supporto per le quali non ci sono osservazioni disponibili. Questo è un problema abbastanza ricorrente nel campo della stima nonparametrica ed è generalmente risolto troncando la stima alla parte positiva e ri-normalizzando oppure stimando una trasformata di f, per esempio $\log f$ o $f^{1/2}$, e poi trasformando all'indietro la stima ottenuta.

In [95] Pinheriro e Vidakovic applicano lo stimatore wavelet alla radice quadrata della funzione di densità; un tale approccio richiede un pre-stimatore di f allo scopo di calcolare i coefficienti wavelet campionari. Infatti

$$\sqrt{f(x)} = \sum_k c_{j_0,k}\phi_{j_0,k}(x) + \sum_{j_0}^{+\infty}\sum_k d_{j,k}\psi_{j,k}(x)$$

con

$$d_{j,k} = \langle \sqrt{f}, \psi_{j,k} \rangle = \int \sqrt{f(x)}\psi_{j,k}(x)dx = \int \frac{\psi_{j,k}(x)}{\sqrt{f(x)}} f(x)dx$$

e

$$c_{j,k} = \langle \sqrt{f}, \phi_{j,k} \rangle = \int \sqrt{f(x)}\phi_{j,k}(x)dx = \int \frac{\phi_{j,k}(x)}{\sqrt{f(x)}} f(x)dx.$$

Quindi, sotto l'assunzione che f sia diversa da zero nel proprio dominio, si ha

$$\hat{f}(x) = \left(\sum_k \hat{c}_{j_0,k}\phi_{j_0,k}(x) + \sum_{j_0}^{j_1} \sum_k \hat{d}_{j,k}\psi_{j,k}(x) \right)^2 \tag{2.22}$$

con coefficienti campionari

$$\hat{d}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \frac{\psi_{j,k}(X_i)}{\sqrt{\hat{f}_n(X_i)}} \quad \text{e} \quad \hat{c}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \frac{\phi_{j,k}(X_i)}{\sqrt{\hat{f}_n(X_i)}} \tag{2.23}$$

e dove $\hat{f}_n(x)$ è un pre-stimatore della densità incognita f.

Naturalmente, questo tipo di approccio è giustificato solo se i coefficienti campionari dipendono da un pre-stimatore semplice e che non richiede eccessivi calcoli. Pinheiro e Vidakovic propongono pre-stimatori del tipo

$$\hat{f}_n(X_i) = \#\{X_j \ : \ X_i - r \le X_j \le X_i + r, \quad j = 1, \cdots, n\} \tag{2.24}$$

con $r \in \mathbb{R}^+$, che, oltre alla semplicità di calcolo, assicurano $\sqrt{\hat{f}_n(\cdot)} \neq 0$.

Questo approccio consente, inoltre, di avere stimatori che hanno integrale uguale ad uno; infatti

$$\|\sqrt{f}\|_2^2 = \langle \sqrt{f}, \sqrt{f} \rangle = \int f(x)dx$$

e, dall'identità di Parseval, ri-normalizzando i coefficienti campionari si ottiene

$$1 = \sum_k |\hat{c}_{j_0,k}|^2 + \sum_{j_0}^{j_1} \sum_k |\hat{d}_{j,k}|^2 = \|\sqrt{\hat{f}}\|_2^2$$

cioè

$$\int \hat{f}(x)dx = 1.$$

Stimatori wavelet non negativi di funzioni di densità sono stati studiati anche da Penev e Dechevsky, [93]. In questo caso sembra che il pre-stimatore sia espresso utilizzando le statistiche d'ordine $X_{n,1} \le X_{n,2} \le \cdots \le X_{n,n}$ come

$$\hat{f}_n(X_{n,i}) = \frac{C}{n(X_{n,i+1} - X_{n,i})}.$$

I due autori dimostrano risultati di convergenza dello stimatore wavelet così ottenuto nel caso di funzioni di densità appartenenti a spazi di Besov e soddisfacenti particolari condizioni.

2.4.3 Scelta della Scala

Un altro problema nella costruzione di stimatori wavelet di funzioni di densità, tipico degli stimatori espressi come serie ortogonale, è la scelta del parametro di scala, ovvero del numero di coefficienti. Tale scelta è necessaria al fine di evitare stimatori wavelet troppo regolari (che non catturano, quindi, le caratteristiche essenziali della funzione) ottenuti per valori del parametro di scala troppo piccoli, o troppo irregolari (che sovrastimano la funzione) nel caso di valori troppo grandi. In particolare, l'espressione (2.21) per la varianza e la distorsione dello stimatore lineare mette in evidenza come al crescere del parametro di scala la distorsione diminuisca mentre la varianza cresce. Come fenomeno generale, scale molto grandi portano a stimatori wavelet simili a funzioni delta concentrate nelle singole osservazioni, cioè a stimatori del tipo $\sum_{i=1}^{n} \delta(x - X_i)$.

Naturalmente, nel caso di stimatori con soglia occorre scegliere opportunamente i due parametri j_0 e j_1, mentre, nel caso lineare, la scelta si riduce al solo parametro $\bar{j}$. Si osservi come, intuitivamente, il problema della scelta del parametro di scala sia più determinante nel caso lineare: utilizzando stimatori a soglia, le tecniche di threshold applicate ai coefficienti wavelet consentono, in un certo qual modo, di eliminare una qualche "ridondanza" dovuta ad una scelta di j_1 troppo alta.

Questa sezione contiene una breve introduzione alle tecniche suggerite nei due diversi casi. Un nuovo metodo, che utilizza una stima wavelet dell'informazione di Fisher come funzione di penalità, sarà poi proposto per il caso lineare.

Nel caso di stimatori con soglia:

a) Donoho et al., [44], suggeriscono di scegliere j_0 secondo il grado di regolarità di ψ e j_1 come $\lfloor \log_2 n - \log_2(\log n) \rfloor$.

b) Pinheiro e Vidakovic, [95], utilizzano invece la funzione *spettro wavelet* di f definita come
$$h_w(j) = \sum_k |\langle f, \psi_{jk} \rangle|^2.$$
La funzione spettro wavelet descrive la distribuzione dell'energia totale di f, cioè la norma $||f||_2^2$, rispetto al parametro j. Questa funzione sarà descritta più ampiamente nel prossimo capitolo. I due autori dimostrano che, se f ha supporto compatto, $h_w(j)$ è sufficientemente regolare per piccoli valori di j mentre cresce in modo esponenziale per j grande (vedi [95] per i dettagli matematici). Dunque, a partire da un livello relativamente grossolano j_0, lo spettro wavelet empirico
$$\hat{h}_w(j) = \sum_k \hat{d}_{jk}^2$$

può essere utilizzato allo scopo di determinare il livello più fine j_1. Gli autori raccomandano di selezionare tale valore come il massimo livello dopo il quale $\hat{h}_w$ inizia la sua crescita esponenziale.

Indicato con $[a, b]$ il campo di variazione dei valori campionari, si può invece scegliere j_0 tale che

$$supp(\phi_{j_0+1,k}) \subset [a, b] \subset supp(\phi_{j_0,k})$$

per qualche k.

Nel caso di stimatori lineari:

a) Poichè lo stimatore wavelet costruito utilizzando la famiglia di wavelets Haar è un istogramma, il valore del parametro di scala per tale stimatore può essere determinato attraverso la regola di Scott (2.1)

$$\bar{\jmath} = 3^{-1} \log_2 n - 2 - \log_2 \sigma.$$

La deviazione standard σ può essere stimata utilizzando quella campionaria, [127]. Inoltre, le wavelets Haar, nell'insieme delle wavelets Daubechies $Daub\#N$, costituiscono la famiglia più semplice, che non ha cioè particolari caratteristiche di regolarità e che determina quindi, in generale, la "peggiore" approssimazione di funzioni $f \in L_2(\mathbb{R})$ sufficientemente regolari. Quindi, da un punto di vista intuitivo, il valore ottimale del parametro di scala determinato per lo stimatore Haar può essere considerato come un limite superiore per gli stimatori costruiti utilizzando le famiglie $Daub\#N$, $N \neq 1$.

b) Walter, [126], suggerisce di scegliere il parametro di scala in modo tale da minimizzare l'errore quadratico medio integrale (EQMI) stimato. Infatti, indicato l'EQMI con e_j (in modo tale da evidenziare la dipendenza di $\hat{f}(x)$ da j)

$$e_j = \int_{-\infty}^{+\infty} E[\hat{f}(x) - f(x)]^2 dx,$$

è possibile dimostrare che l'incremento $e_j - e_{j-1}$ è dato da

$$e_j - e_{j-1} = \frac{1}{n} \int 2^j q(2^j x, 2^j x) f(x) dx - \frac{n+1}{n} \sum_k d_{j-1,k}^2$$

dove l'integrale può essere stimato mediante $\frac{1}{n} 2^j \sum_1^n q(2^j X_i, 2^j X_i)$. Inoltre, nel caso si utilizzi la famiglia Haar, si ha

$$e_j - e_{j-1} = \frac{2^{j-1}}{n} - \frac{n+1}{n} \sum_k d_{j-1,k}^2.$$

Walter suggerisce dunque di partire da un livello fissato, che dovrà essere scelto non più fine di quello necessario per avere ogni X_i in un diverso intervallo, ed usare i coefficienti wavelet empirici per stimare gli incrementi di errore ai livelli successivi più piccoli fino al livello per il quale l'errore aumenta in modo considerevole. Il livello raggiunto è quello ottimale.

Lo scopo principale di questa sezione è quello di proporre un nuovo metodo per la scelta del valore ottimale del parametro di scala nel caso di stimatori wavelet lineari. Tale metodo utilizza l'informazione di Fisher come funzione di penalità. Data f una funzione di densità di probabilità derivabile, l'informazione di Fisher di f è definita come

$$F(f) = \int_{\mathbb{R}} \frac{1}{f(x)} \Big[\frac{d}{dx} f(x)\Big]^2 dx.$$

Una stima dell'informazione di Fisher dello stimatore wavelet $\hat{f}$ può essere ottenuta nel seguente modo: si indichi con T l'operatore derivata d/dx e con P_j l'usuale operatore di proiezione in V_j. Seguendo Beylkin, [12], la rappresentazione wavelet di $T_j = P_j T P_j$ che utilizza le wavelets $Daub\#N$ è

$$T_j f(x) = \sum_k r_{j,k} \phi_{j,k}(x) \tag{2.25}$$

dove

$$r_{j,k} = 2^j \sum_l r_l c_{j,k-l}$$

e $c_{j,k-l} = \langle f, \phi_{j,k-l} \rangle$. L'operatore T_j è un'approssimazione di T alla scala j e si ha $T_j \to T$ per $j \to \infty$. I valori dei coefficienti (pesi) $\{r_l\}$ sono tabulati in [12]; tali valori sono ottenuti risolvendo il sistema

$$r_l = 2\Big[r_{2l} + \frac{1}{2} \sum_{k=1}^{L/2} a_{2k-1}(r_{2l-2k+1} + r_{2l+2k-1})\Big],$$

$$\sum_l l r_l = -1,$$

dove $a_n = 2\sum_{i=0}^{L-1-n} h_i h_{i+n}$ e $n = 1, \cdots, L-1$ con $L = 2N$.

Si ricorda che lo stimatore wavelet lineare di densità, $\hat{f}$, è stato ottenuto dalla proiezione di f in V_j utilizzando i coefficienti empirici come stime dei $c_{j,k}$. L'espressione (2.25) determina pertanto uno stimatore della derivata di $\hat{f}$ del tipo

$$\widehat{\frac{d\hat{f}}{dx}}(x) = \sum_k r_{j,k} \phi_{j,k}(x)$$

dove

$$r_{j,k} = 2^j \sum_l r_l \hat{c}_{j,k-l}$$

e con coefficienti $\hat{c}_{j,k}$ dati da (2.16).

Stimatori a valori positivi del tipo (2.22) sono poi stati ottenuti applicando lo stimatore wavelet alla radice quadrata di f. In tal caso si ha

$$\widehat{\frac{d\sqrt{\hat{f}}}{dx}}(x) = \sum_k r_{j,k}\phi_{j,k}(x) \tag{2.26}$$

dove

$$r_{j,k} = 2^j \sum_l r_l \hat{c}_{j,k-l}$$

e con coefficienti $\hat{c}_{j,k}$ dati adesso dalla (2.23).

Applicando l'identità di Parseval alla (2.26) si ottiene

$$\|\widehat{\frac{d\sqrt{\hat{f}}}{dx}}\|_2^2 = \sum_k |r_{j,k}|^2. \tag{2.27}$$

Si osservi adesso che

$$\|\frac{d\sqrt{f}}{dx}\|_2^2 = \frac{1}{4}\int \frac{f'^2(x)}{f(x)}dx$$

ovvero la norma al quadrato della derivata di $\sqrt{f}$ è proporzionale all'informazione di Fisher di f. L'espressione (2.27) può essere quindi scritta come

$$\frac{1}{4}\int \frac{\widehat{\frac{d\hat{f}}{dx}}^2(x)}{\hat{f}(x)}dx = \sum_k |r_{j,k}|^2$$

e determina una *stima wavelet* dell'informazione di Fisher di $\hat{f}$. L'aggettivo wavelet sta ad indicare che questa particolare stima coinvolge la rappresentazione wavelet della derivata di $\hat{f}$. Tale stima dipende dal parametro di scala j e sarà indicata con $F(j)$; si ha quindi

$$F(j) = \sum_k |r_{j,k}|^2. \tag{2.28}$$

Fissato j, $F(j)$ è la stima wavelet dell'informazione di Fisher dello stimatore di densità (2.22) ottenuto per quel particolare valore del parametro.

A questo punto è utile ricordare che l'informazione di Fisher è stata utilizzata come funzione di penalità da Good e Gaskin, [54], nell'ambito della stima di massima verosimiglianza. In tale contesto, dato $(X_1, \cdots, X_n)$ un campione estratto da una densità f, lo stimatore di massima verosimiglianza di f è definito come la soluzione del problema di ottimizzazione vincolato

$$max_f[L(f)]$$

$$\int f(x)dx = 1 \text{ e } f(x) \geq 0 \ \forall x,$$

dove il funzionale $L(f)$ è la log-verosimiglianza

$$L(f) = \sum_{i=1}^{n} \log f(X_i).$$

Lo stimatore ottenuto come soluzione di tale problema è tuttavia "troppo" irregolare e risulta, in alcuni casi, simile ad una media di funzioni delta centrate nelle singole osservazioni. Il problema è dunque simile a quello relativo all'utilizzo di stimatori wavelet a scale troppo fini. Nel contesto della stima di massima verosimiglianza, la pratica comune è quella di scegliere una funzione di penalità $F(f)$ che misuri in qualche modo il grado di regolarità (smoothness) di f e massimizzare il funzionale

$$\tilde{L}(f) = \sum_{i=1}^{n} \log f(x_i) - \alpha F(f)$$

dove α è il parametro che regola lo *smoothness*. La verosimiglianza penalizzata può essere vista come un modo per quantificare il necessario compromesso tra la bontà di adattamento delle stime ai dati, misurata dalla log-verosimiglianza, e il grado di smoothness della curva stimata. Il parametro α controlla tale compromesso: piccoli valori di α determinano, ad esempio, stimatori molto irregolari ma con un buon adattamento ai dati. Una ottima esposizione degli stimatori di massima verosimiglianza penalizzata è data da Tapia e Thompson in [111].

L'irregolarità di una funzione può essere misurata dalla facilità con cui è possibile discriminare la funzione da se stessa qualora si considerino piccole traslazioni di questa. Good e Gaskins suggeriscono di utilizzare l'informazione di Fisher come possibile scelta per $F(f)$

$$F(f) = \int_{\mathbb{R}} \frac{1}{f(x)} \Big[\frac{d}{dx} f(x)\Big]^2 dx.$$

L'informazione di Fisher è invariante per traslazioni ed è quindi una misura di irregolarità nel senso sopra indicato.

L'idea di Good e Gaskins può essere applicata anche al caso degli stimatori wavelet ortonormali. Come già osservato, valori del parametro di scala troppo piccoli determinano stimatori wavelet troppo regolari e valori grandi stimatori irregolari. Nella pratica, solo un piccolo numero di valori (spesso tre o quattro) determina stimatori che risultano "accettabili", come può essere facilmente verificato da una analisi dei grafici. Empiricamente, è stato osservato che la funzione (2.28) presenta valori crescenti al crescere di j e tende a zero per valori decrescenti; questo ne incoraggia l'utilizzo come funzione che misura l'irregolarità degli stimatori wavelet ottenuti al variare di j con lo scopo di selezionare, nell'insieme degli stimatori accettabili, quello "ottimale" per il problema. A tale scopo è utile la seguente

Proposizione 2.4.2 *Nella famiglia delle funzioni di densità di probabilità f con supporto compatto $[a, b]$ e tali che $f(a) = f(b) = 0$, la funzione che minimizza l'informazione di Fisher è*

$$\tilde{f}(x) = \frac{1}{m}\cos^2\frac{\pi x}{2m},$$

dove $m = (b-a)/2$. L'informazione minima raggiunta da $\tilde{f}$ è

$$F(\tilde{f}) = \frac{\pi^2}{m^2}. \tag{2.29}$$

Si riporta un breve cenno alla dimostrazione: il problema variazionale

$$min \int_a^b F(x, y, y')dx$$

sotto il vincolo

$$\int_a^b \phi(x, y, y')dx = C$$

ha equazione di Eulero, [49],

$$\frac{\partial}{\partial y}(F - \lambda\phi) - \frac{\partial}{\partial x}(\frac{\partial}{\partial y'}(F - \lambda\phi)) = 0.$$

Posto $F(x, y, y') = \frac{(y')^2}{y}$ e $\phi(x, y, y') = y$ si ha

$$2\Big(\frac{y'}{y}\Big)' + \Big(\frac{y'}{y}\Big)^2 = -C.$$

Sostituendo $z = \sqrt{y}$ si ottiene

$$(z)'' + \frac{C}{4}z = 0$$

ovvero

$$(\sqrt{y})'' + \frac{C}{4}\sqrt{y} = 0.$$

La soluzione positiva che soddisfa le condizioni $\int_a^b y dx = 1$ e $y(a) = y(b) = 0$ è

$$y = \frac{1}{m}\cos^2\frac{\pi x}{2m},$$

dove $m = (b-a)/2$. Integrando si ha

$$F(y) = \int_{-m}^{m}\frac{(y')^2}{y}dx = \frac{\pi^2}{m^2}.$$

Allora, il metodo proposto per la determinazione del valore ottimale del parametro di scala è il seguente, [118]: sia $[a, b]$ un intervallo chiuso contenente i valori campionari $X_1, \cdots, X_n$ e sia FI_{min} l'informazione di Fisher minima ottenibile in tale intervallo calcolata secondo la (2.29). Il valore ottimale del parametro di scala è quello che permette di ottenere uno stimatore la cui stima wavelet dell'informazione di Fisher è più vicina a FI_{min}. In altre parole, si seleziona il più piccolo valore $j = \bar{\jmath}$ tale che

$$F(\bar{\jmath}) \geq \frac{\pi^2}{4m^2}. \tag{2.30}$$

La funzione di densità deve soddisfare le condizioni richieste nella proposizione 2.4.2.

2.4.4 Caso Multivariato

La maggior parte dei risultati relativi allo stimatore wavelet univariato possono essere facilmente estesi al caso multivariato, [127] [117]. Si ricorda infatti che la rappresentazione wavelet di una funzione f in $L_2(\mathbb{R}^d)$ è

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}(x) + \sum_{j=j_0}^{+\infty}\sum_{\mathbf{k}}\sum_{l=1}^{2^d-1} d_{j,\mathbf{k}}^{(i)}\psi_{j,\mathbf{k}}^{(l)}(x)$$

dove $\mathbf{k} \in \mathbb{Z}^d$, $\mathbf{k} = (k_1, \cdots, k_d)$ e

$$\phi_{j_0,\mathbf{k}}(x) = 2^{jd/2}\phi(2^j x_1 - k_1, \cdots, 2^j x_d - k_d)$$

$$\psi_{j,\mathbf{k}}^{(l)}(x) = 2^{jd/2}\psi^{(i)}(2^j x_1 - k_1, \cdots, 2^j x_d - k_d).$$

Allora, dato $\mathbf{X}_i, i = 1, \cdots, n$ un campione estratto da una funzione di densità di probabilità f, lo stimatore wavelet di f è costruito, come nel caso univariato, considerando i coefficienti empirici e troncando la serie ad un fissato livello j_1. Ovvero

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k}} \hat{c}_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{j_1}\sum_{\mathbf{k}}\sum_{l=1}^{2^d-1} \hat{d}_{j,\mathbf{k}}^{(j)}\psi_{j,\mathbf{k}}^{(l)}(\mathbf{x}). \tag{2.31}$$

Ancora, è possibile considerare tale stimatore nella versione lineare come proiezione ortogonale di f in $V_{\bar{j}}$

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k}} \hat{c}_{\bar{j},\mathbf{k}}\phi_{\bar{j},\mathbf{k}}(\mathbf{x}), \quad \bar{j} = j_1 + 1 \tag{2.32}$$

dove

$$\hat{c}_{\bar{j},\mathbf{k}} = n^{-1}\sum_{i=1}^{n} \phi_{\bar{j},\mathbf{k}}(\mathbf{X}_i).$$

Questo stimatore essere espresso utilizzando il nucleo autoriproducente in $V_{\bar{j}}$ come

$$\hat{f}(\mathbf{x}) = n^{-1}\sum_{i=1}^{n} q_{\bar{j}}(\mathbf{X}_i, \mathbf{x})$$

dove

$$q_{\bar{j}}(\mathbf{t}, \mathbf{x}) = \prod_{s=1}^{d} q_{\bar{j}}(t_s, x_s),$$

e $q_{\bar{j}}(t_s, x_s)$ dato dalla (2.20).

Di seguito il caso bivariato è elaborato in dettaglio. Sia $(X_i, Y_i), i = 1, \cdots, n$ un campione estratto da (X, Y) con funzione di densità di probabilità $f(x, y)$. Si ha

$$\hat{f}(x, y) = \sum_{k_1}\sum_{k_2} \hat{c}_{j,k_1,k_2}\phi_{j,k_1,k_2}(x, y)$$

dove

$$\hat{c}_{j,k_1,k_2} = n^{-1}\sum_{i=1}^{n} \phi_{j,k_1,k_2}(X_i, Y_i)$$

e

$$\phi_{j,k_1,k_2}(x, y) = \phi_{j,k_1}(x)\phi_{j,k_2}(y).$$

Inoltre,

$$\begin{aligned}\hat{f}(x, y) &= \sum_{k_1}\sum_{k_2}(n^{-1}\sum_{i=1}^{n} \phi_{j,k_1,k_2}(X_i, Y_i))\phi_{j,k_1,k_2}(x, y) \\ &= n^{-1}\sum_{i=1}^{n}(\sum_{k_1}\sum_{k_2} \phi_{j,k_1,k_2}(X_i, Y_i)\phi_{j,k_1,k_2}(x, y))\end{aligned}$$

$$= n^{-1}\sum_{i=1}^{n}(\sum_{k_1}\sum_{k_2}\phi_{j,k_1}(X_i)\phi_{j,k_2}(Y_i)\phi_{j,k_1}(x)\phi_{j,k_2}(y))$$
$$= n^{-1}\sum_{i=1}^{n}(\sum_{k_1}(\sum_{k_2}\phi_{j,k_2}(Y_i)\phi_{j,k_2}(y))\phi_{j,k_1}(X_i)\phi_{j,k_1}(x))$$

Posto

$$q_j(X_i,x) = \sum_{k_1}\phi_{j,k_1}(X_i)\phi_{j,k_1}(x)$$
$$q_j(Y_i,y) = \sum_{k_2}\phi_{j,k_2}(Y_i)\phi_{j,k_2}(y)$$

allora

$$\hat{f}(x,y) = n^{-1}\sum_{i=1}^{n} q_j(X_i,x)q_j(Y_i,y).$$

I risultati di consistenza visti nel caso univariato possono essere estesi a quello multivariato. In particolare, [127],

Proposizione 2.4.3 *Sia f una funzione di densità di probabilità continua e limitata. Per $\bar{\jmath} \to \infty$ si ha*

$$(i) \quad E|\hat{f}(x) - f(x)|^2 \to 0, \ \textit{uniformemente in sottoinsiemi compatti}$$

$$(ii) \quad \textit{se } f \in H^{-\alpha} \textit{ allora } E|\hat{f}(x) - f(x)|^2_{-\alpha} \to 0.$$

Lo stimatore wavelet a soglia multivariato è ottenuto considerando lo stimatore nella forma (2.31) e applicando opportune tecniche di threshold ai coefficienti wavelet. Si osserva che, come nel caso univariato,

$$V_{j_1+1} = V_{j_0} \oplus_{i=1}^{2^d-1} \bigoplus_{j_0}^{j_1} W_j^{(i)}$$

Quindi, per $\bar{\jmath} = j_1 + 1$, stimatore lineare e stimatore a soglia sono equivalenti se non si modifica alcuno dei coefficienti wavelet.

Stimatori wavelet Bona Fide multivariati possono essere ottenuti stimando ancora la radice quadrata di f. Si ha

$$\hat{f}(\mathbf{x}) = \left(\sum_{\mathbf{k}}\hat{c}_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{j_1}\sum_{\mathbf{k}}\sum_{l=1}^{2^d-1}\hat{d}_{j,\mathbf{k}}^{(j)}\psi_{j,\mathbf{k}}^{(l)}(\mathbf{x})\right)^2 \tag{2.33}$$

dove

$$\hat{c}_{j_0,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\phi_{j_0,\mathbf{k}}(\mathbf{X}_i)}{\sqrt{\hat{f}_n(\mathbf{X}_i)}} \quad \text{e} \quad \hat{d}_{j,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\psi_{j,\mathbf{k}}(\mathbf{X}_i)}{\sqrt{\hat{f}_n(\mathbf{X}_i)}}.$$

Ri-normalizzando i coefficienti campionari lo stimatore ha integrale uguale ad uno. Pre-stimatori, semplice generalizzazione di quelli (2.24) proposti per il caso univariato, possono essere utilizzati. Precisamente

$$\hat{f}_n(\mathbf{X}_i) = \#\{\mathbf{X}_j \in B(\mathbf{X}_i, r), \quad j = 1, \cdots, n\} \tag{2.34}$$

con $B(\mathbf{X}_i, r)$ una palla centrata in $\mathbf{X}_i$ con raggio r opportunamente scelto.

Per quanto riguarda la scelta dei parametri di scala j_0 e j_1, si osservi, innanzi tutto, che, nel caso di istogrammi multidimensionali, la regola (2.1) suggerita da Scott determina partizioni di ampiezze diverse per ogni dimensione poichè $\hat{h}_k$ dipende dalla deviazione standard campionaria della k-esima variabile. Tale regola non è quindi applicabile allo stimatore wavelet lineare (2.32) che, nel caso si utilizzi la base di Haar, è ancora un istogramma che ha però ampiezza di banda $2^{-\bar{j}/d}$ per ciascuna dimensione.

L'estensione multivariata del metodo proposto nel caso univariato per stimatori wavelet lineari e che utilizza l'informazione di Fisher come funzione di penalità è tuttora in fase di studio. Si riporta inoltre un cenno, di carattere puramente bibliografico, ad un metodo suggerito da Tribouley, [116], che propone una procedura di validazione incrociata per la determinazione del valore ottimale $\bar{j}$, apparso durante la stesura di questo lavoro.

Per quanto concerne invece gli stimatori wavelet a soglia, la tecnica proposta da Pinheiro e Vidakovic nel caso univariato può essere facilmente estesa: lo spettro wavelet empirico in d dimensioni può essere definito come

$$\hat{h}_w(j) = \sum_{i=1}^{2^d-1} \sum_{\mathbf{k}} [\hat{d}_{j\mathbf{k}}^{(i)}]^2$$

dove

$$\hat{d}_{j,\mathbf{k}}^{(i)} = n^{-1} \sum_{i=1}^{n} \psi_{j,\mathbf{k}}^{(i)}(\mathbf{X}_i)$$

e il livello più fine j_1 può essere calcolato in base al grafico di tale funzione secondo il criterio suggerito per il caso univariato. Inoltre, se i valori campionari sono contenuti in $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$, j_0 può essere selezionato in modo tale che

$$supp(\phi_{j_0+1,\mathbf{k}}) \subset [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d] \subset supp(\phi_{j_0,\mathbf{k}})$$

per qualche $\mathbf{k}$. Ovvero

$$supp(\phi_{j_0+1,k_l}) \subset [a_l, b_l] \subset supp(\phi_{j_0,k_l}), \quad l = 1, \cdots, d.$$

Si osservi che scalare i dati nello stesso intervallo semplifica notevolmente tali operazioni.

2.4.5 Applicazioni

Allo scopo di verificarne la validità, gli stimatori wavelet di densità sono stati utilizzati su un insieme di dati simulati da una distribuzione normale e su un insieme bivariato di dati reali. Tali applicazioni sono state realizzate utilizzando le wavelets di Daubechies, $Daub\#N$ e $Coiflet\#N$, presentate nella sezione 2.3.2.

I problemi computazionali che si sono presentati riguardavano soprattutto la determinazione del campo di variazione di k per $\phi_{j,k}$ e $\psi_{j,k}$ in (2.18) o (2.19). A tale scopo si osservi che, poichè le wavelets di Daubechies sono caratterizzate dall'avere supporto compatto, il campo di variazione di k è finito e può essere calcolato nel seguente modo: nel caso si utilizzi la famiglia di wavelets $Daub\#N$ il supporto di ϕ è $[0, 2N-1]$ e poichè $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ si ha

$$0 \leq 2^j x - k \leq 2N - 1$$

cioè

$$\frac{k}{2^j} \leq x \leq \frac{2N-1+k}{2^j}.$$

Se il campo di variazione dei valori campionari è l'intervallo $[a, b]$, sarà sufficiente calcolare i valori di k per i quali il supporto delle corrispondenti funzioni $\phi_{j,k}$ interseca $[a, b]$

$$\Big[\frac{k}{2^j}, \frac{2N-1+k}{2^j}\Big] \bigcap [a, b] \neq \emptyset.$$

Dunque

$$\frac{2N-1+k}{2^j} \geq a$$
$$\frac{k}{2^j} \leq b$$

cioè

$$\lceil a2^j \rceil - 2N + 1 \leq k \leq \lfloor b2^j \rfloor.$$

Per esempio, per $[a, b] = [0, 1]$ e $j = 0$ il campo di variazione di k è dato da $-2N + 1 \leq k \leq 1$. In modo analogo, il supporto di ψ è $[1 - N, N]$ e le $\psi_{j,k}$ possono essere scelte in modo tale che

$$\lceil a2^j \rceil - N \leq k \leq \lfloor b2^j \rfloor + N - 1.$$

Nel caso invece si utilizzi la famiglia di wavelets $Coiflet\#N$, si ha che la funzione di scala ϕ_N^c e la wavelet madre ψ_N^c hanno come supporto intervalli di ampiezza $(6N - 1)$. In particolare, *supp* $\phi_N^c = [-2N, 4N - 1]$, quindi

$$-2N \leq 2^j x - k \leq 4N - 1$$

cioè

$$\frac{-2N+k}{2^j} \leq x \leq \frac{4N-1+k}{2^j}$$

e i valori di k per i quali il supporto delle corrispondenti $\phi^c_{j,k}$ interseca $[a,b]$ sono dati da

$$a2^j - 4N + 1 \leq k \leq b2^j + 2N.$$

Analogamente per ψ^c_N.

Queste considerazioni possono essere facilmente estese al caso multivariato. Se il campo di variazione dei valori campionari è dato da $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$, nel caso vengano utilizzate le wavelets $Daub\#N$, si ha $\phi_{j,\mathbf{k}}$ con $\mathbf{k} = (k_1, \cdots, k_d)$ tale che

$$\lceil a_i 2^j \rceil - 2N + 1 \leq k_i \leq \lfloor b_i 2^j \rfloor$$

e $\psi_{j,\mathbf{k}}$ con $\mathbf{k} = (k_1, \cdots, k_d)$ tale che

$$\lceil a_i 2^j \rceil - N \leq k_i \leq \lfloor b_i 2^j \rfloor + N - 1.$$

La procedura è analoga nel caso si usino le wavelets $Coiflet\#N$.

Si osservi inoltre che il calcolo degli stimatori wavelet richiede la scelta di un algoritmo che permetta di determinare i valori della funzione di scala e della wavelet madre in un punto arbitrario. L'algoritmo utilizzato nelle applicazioni presentate è quello a cascata descritto nella sezione 2.3.1.

I programmi sono stati realizzati in Matlab utilizzando il pacchetto *Wavbox*[2] *4.2* per ottenere i coefficienti di filtro delle wavelets Daubechies e per calcolarne i valori in punti arbitrari mediante l'algoritmo a cascata.

Gli stimatori wavelet che hanno dato i risultati migliori sono stati quelli lineari. Allo scopo di evitare valori negativi tali stimatori sono stati applicati alla radice quadrata della funzione di densità. I pre-stimatori utilizzati sono quello (2.24) nel caso univariato e quello (2.34) in quello bivariato. I coefficienti campionari sono stati ri-normalizzati allo scopo di ottenere stimatori con integrale uguale ad uno.

Il valore ottimale del parametro di scala è stato determinato, nel caso univariato, con il metodo che usa la stima wavelet dell'informazione di Fisher come funzione di penalità. Si osservi come, utilizzando il pre-stimatore (2.24), la funzione $F(j)$ definita in (2.28) dipende non solo da j ma anche dal raggio r utilizzato per il pre-stimatore. I risultati numerici presentati dimostreranno tuttavia che il metodo utilizzato seleziona lo stesso valore del parametro di scala per valori diversi di r. In altre parole, il metodo proposto appare robusto rispetto alla scelta del raggio.

[2]Copyright (c) 1994, Carl Taswell.

Inoltre, se i valori campionari sono contenuti in un intervallo di ampiezza s, il raggio r varierà al più in $[0, s]$. Valori in $[0, s/2]$ possono essere considerati scelte ragionevoli.

Campione Casuale Normale

Come primo esempio si consideri un campione casuale di 500 osservazioni simulato da una densità normale $N(0, 1)$.

Poichè i valori campionari ottenuti sono contenuti in $[-3, 4]$, si ha, con riferimento alla proposizione 2.4.2, $m = 3.5$ e

$$\frac{\pi^2}{4m^2} = 0.2014.$$

Varie famiglie wavelet $Daub\#N$ sono state utilizzate. Quella che ha dato i risultati migliori è la $Daub\#7$.

La tabella 2.1 contiene i valori della funzione $F(j)$ in (2.28) per diversi valori di j e per diverse scelte del raggio r. Dall'esame di questa tabella risulta evidente che il più piccolo valore di j che soddisfa la condizione (2.30) è $j = -1$ per ogni scelta di r.

Raggio	Livello 2	Livello 1	Livello 0	Livello -1	Livello -2
0	1.6359	0.5304	0.4641	0.3112	0.1023
0.3	1.6894	0.3407	0.2574	0.2254	0.0958
0.5	2.0498	0.3726	0.2676	0.2307	0.0963
1	1.7271	0.4469	0.3090	0.2496	0.0978
1.5	1.6107	0.4482	0.3561	0.2719	0.0995
2	1.6352	0.4726	0.3962	0.2877	0.1007
2.5	1.6278	0.5007	0.4278	0.2986	0.1014
3	1.6323	0.5161	0.4478	0.3052	0.1019

Tabella 2.1: Campione normale. Wavelets $Daub\#7$. Valori di $F(j)$ per diverse scelte della scala j e del raggio r.

La figura 2.2 mostra l'andamento della funzione $F(j)$ per $j = -3, \cdots, 3$ e raggio 0.3. Si osservi come la funzione decada a zero per $j \to -\infty$ e cresca invece esponenzialmente per $j \to +\infty$. La figure 2.3 *a)* mostra lo stimatore wavelet (2.22) con raggio $r = 0.3$ e scala $j = -1$.

Data la regolarità della funzione di densità Normale, risultati migliori sono stati ottenuti con le wavelets $Coiflet\#N$ che sono, in generale, più regolari delle $Daub\#N$. La figura 2.3 *b)* mostra lo stimatore wavelet (2.22) che utilizza la famiglia $Coiflet\#5$. Si osservi che, in questo caso, il metodo che utilizza la stima

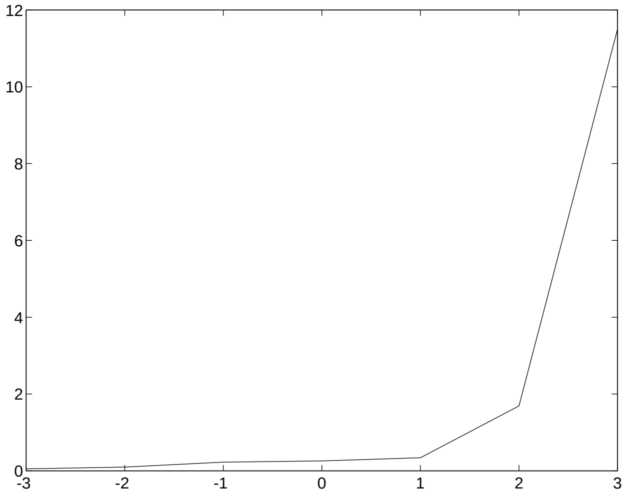

Figura 2.2: Campione normale. Wavelets $Daub\#7$ e raggio $r = 0.3$. Grafico della funzione $F(j),\ j = -3, \cdots, 3$.

wavelet dell'informazione di Fisher per la determinazione del valore ottimale del parametro di scala non può essere utilizzato. Tuttavia, poichè si conosce la forma della densità, la determinazione di tale valore può avvenire semplicemente guardando i grafici degli stimatori wavelet ottenuti a scale diverse.

Il set di dati Geyser

L'insieme di dati *Old Faithful Geyser* ha due componenti, il *tempo di attesa tra successive eruzioni* del geyser Old Faithful nello Yellowstone National Park, Wyoming, USA, e il *tempo di durata delle eruzioni.* Il tempo di attesa è l'intervallo di tempo tra l'inizio di successive eruzioni ed il tempo di durata è la durata della eruzione susseguente (cioè che segue all'attesa). I tempi sono misurati in minuti. I dati sono stati collezionati nel periodo 1/15 agosto 1985, senza interruzioni, per un totale di 299 coppie di osservazioni. Alcune misurazioni relative alla variabile tempo di durata, prese durante la notte, erano state originariamente registrate come S (breve), M (media) e L (lunga); questi valori sono poi stati codificati come 2, 3 e 4 minuti rispettivamente. Questo insieme di dati è stato studiato da Azzalini and Bowman, [9]. Quella che segue è una breve descrizione delle ipotesi relative al funzionamento del geyser riportata da Rinehart, [97].

> L'acqua entra nel tubo geyser ed è riscaldata dalle rocce circostanti.
> L'acqua è portata ad una temperatura al di sopra di quella di ebol-

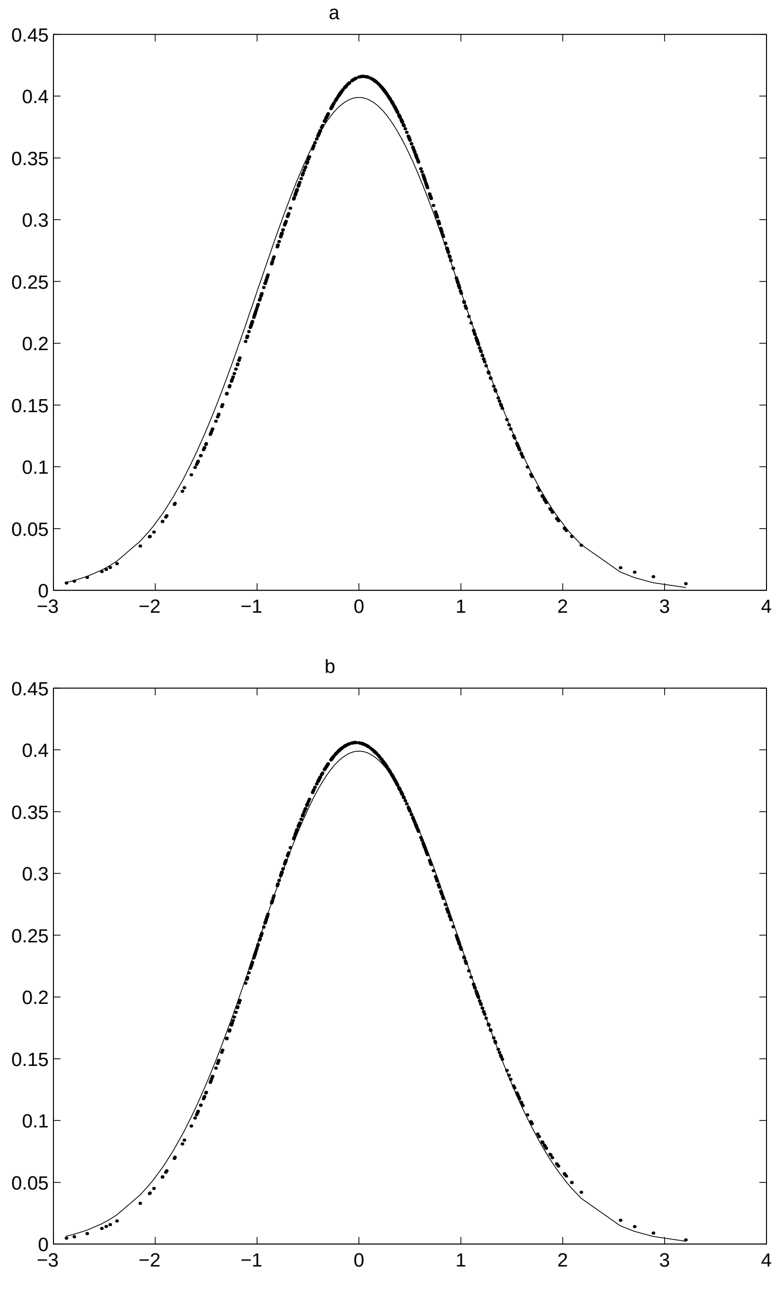

Figura 2.3: Campione normale (linea continua). Stimatori wavelet di densità (linea tratteggiata) con raggio $r = 0.3$ e scala $j = -1$. *a)* Wavelets $Daub\#7$. *b)* Wavelets $Coiflet\#5$.

lizione a causa della pressione crescente dovuta alla massa di acqua che si trova in alto: più profonda è l'acqua più alta è la temperatura richiesta per l'ebollizione. Quando l'acqua sulla cima raggiunge tale temperatura, diventa vapore e si muove verso la superficie. La pressione sul fondo allora scende rapidamente al livello normale e, per effetto di induzione, l'acqua sul fondo diventa rapidamente vapore. Questo meccanismo a cascata si ripete molte volte: non appena l'acqua è convertita in vapore, la pressione sull'acqua sottostante decresce causando la produzione di più vapore e sostenendo l'eruzione.

caso a) Se, ad un certo punto di tale meccanismo a cascata, la temperatura nelle zone più profonde è più bassa di quella prevista, la cascata si ferma e l'eruzione è breve.

caso b) Altrimenti, quando la temperatura è sufficientemente alta anche nelle zone più profonde, il meccanismo a cascata continua e l'eruzione è più lunga.

A questo punto il tubo è completamente o parzialmente vuoto, pronto per essere riempito ancora con nuova acqua.

La descrizione sopra riportata suggerisce tempi di attesa per l'eruzione successiva sia brevi che lunghi; nel caso a) il sistema inizia un nuovo ciclo ma il tubo è parzialmente pieno di acqua calda, quindi il tempo di riscaldamento successivo è più breve. Inoltre, figura 2.4, un'occhiata ai dati relativi alla durata delle eruzioni mostra un comportamento oscillatorio tra due livelli, uno basso ed uno alto, ed il grafico scatter delle variabili tempo di attesa e tempo di durata dell'eruzione precedente mostra che brevi durate possono essere associate a brevi attese e lunghe durate a lunghe attese.

Azzalini e Bowman hanno tentato una modellazione fisica del fenomeno presentando una catena markoviana a due stati come modello per le oscillazioni tra alti e bassi livelli con le variabili tempo di durata dell'eruzione precedente e tempo di attesa come indicatori equivalenti dello stato del sistema.

Questo insieme di dati viene spesso considerato un banco di prova per stimatori di funzioni di densità. Grafici bidimensionali della funzione di densità di probabilità stimata relativa a questo insieme di dati, ottenuti con l'istogramma *traslato in media* (average shifted) proposto da Scott, [104], e con uno stimatore con nucleo quartico bidimensionale, sono presentati in [58] e [119]. Lo stimatore wavelet è stato utilizzato su questo insieme di dati per stimare la funzione di densità di probabilità univariata della variabile aleatoria tempo di durata delle eruzioni e per stimare la densità bivariata.

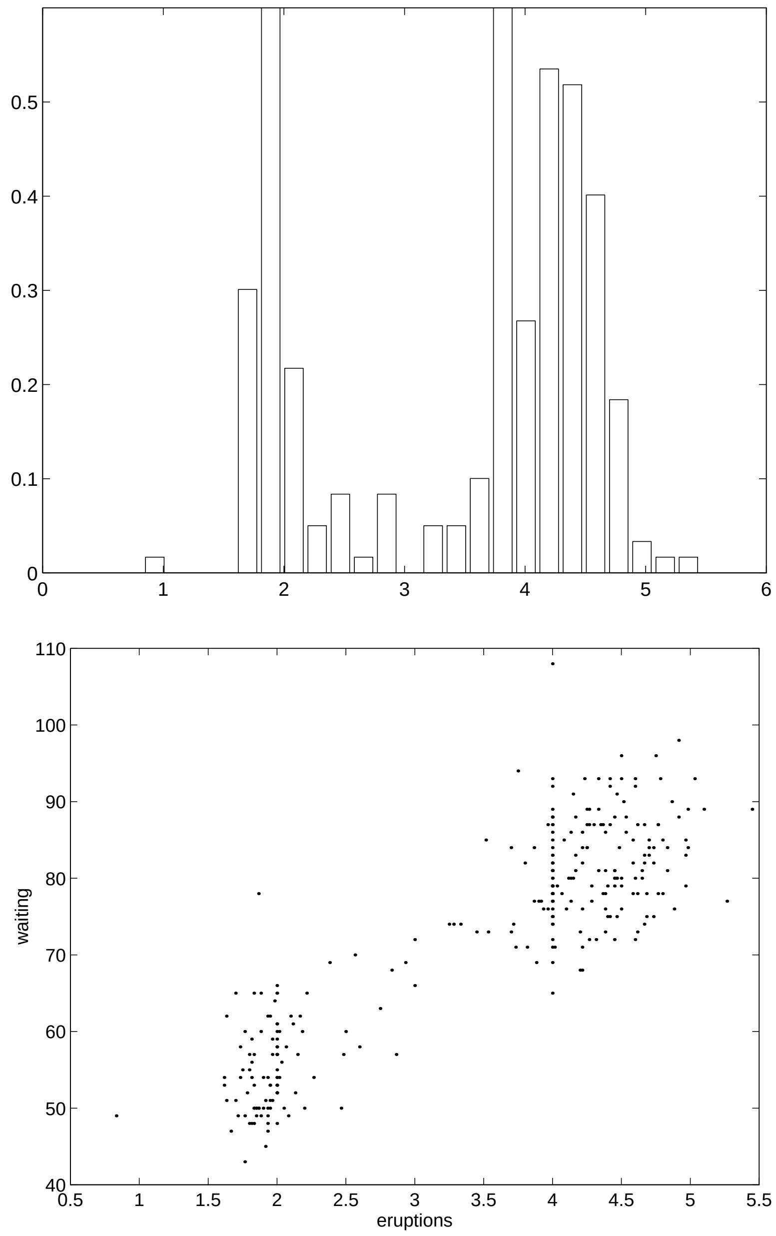

Figura 2.4: Dataset *Geyser*. Istogramma della componente *tempo di durata delle eruzioni* e grafico scatter delle variabili *tempo di attesa* e *tempo di durata dell'eruzione precedente*.

Nel caso univariato, poichè i dati sono contenuti nell'intervallo $[0, 6]$, si ha $m = 3$ e

$$\frac{\pi^2}{4m^2} = 0.2742.$$

La famiglia wavelet che ha dato i risultati migliori è la $Daub\#7$. La tabella 2.2 riporta i valori di $F(j)$ per diverse scelte di j e del raggio r. Il livello selezionato è $j = 0$. Anche in questo caso, si osservi come tale scelta non dipende dal valore di r.

Raggio	Livello 2	Livello 1	Livello 0	Livello -1	Livello -2
0	7.2377	3.2491	1.1175	0.2591	0.0965
0.2	8.4493	2.3024	0.8745	0.2623	0.0967
0.5	12.8244	2.9571	1.0001	0.2498	0.0954
1	13.5481	3.8524	1.2119	0.2569	0.0955
1.5	13.6173	3.9924	1.2609	0.2581	0.0956
2	13.1021	3.9169	1.2808	0.2591	0.0960
2.5	12.3629	3.6583	1.2578	0.2691	0.0971

Tabella 2.2: Dati Geyser. Componente: durata delle eruzioni. Wavelets $Daub\#7$. Valori di $F(j)$ per diverse scelte della scala j e del raggio r.

La figura 2.5 mostra lo stimatore wavelet (2.22) con raggio $r = 0.2$ e scala $j = 0$.

Infine, lo stimatore wavelet è stato applicato al caso bivariato. La figura 2.6 mostra lo stimatore wavelet lineare nella forma (2.33) con wavelet madre $Daub\#7$ e pre-stimatore (2.34) con raggio $r = 0.25$. Il valore del parametro di scala, $j = 0$, è stato determinato in base ai grafici dello stimatore relativi a diversi valori di tale parametro.

2.5 Stimatori Wavelet di Funzioni di Regressione

Nel seguito le basi ortonormali wavelet saranno utilizzate per costruire uno stimatore non parametrico di funzioni di regressione. Tale stimatore sarà, in particolare, costruito utilizzando lo stimatore wavelet lineare di funzioni di densità (2.19). Lo stimatore ottenuto è consistente in media quadratica. Infine, sarà presentata una applicazione ad una serie storica di tassi di cambio; i risultati ottenuti saranno confrontati con quelli relativi a stimatori più classici.

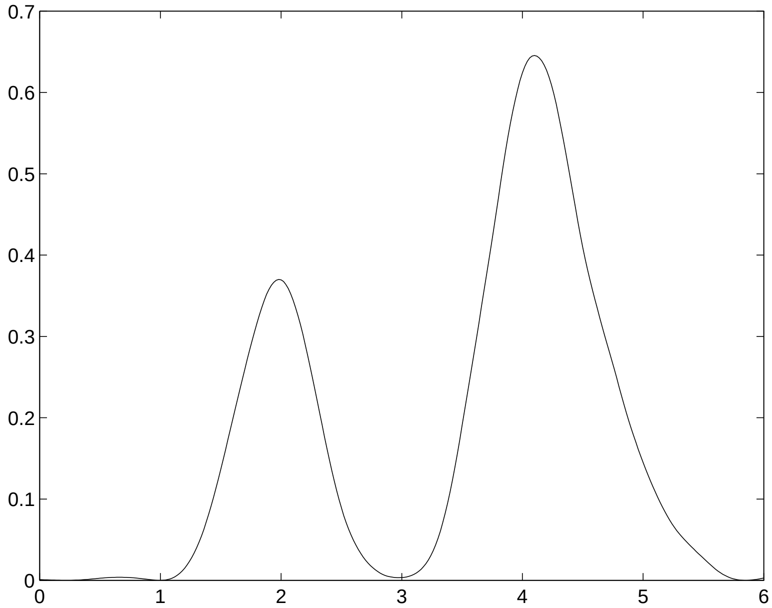

Figura 2.5: Dataset *Geyser*, componente *tempo di durata delle eruzioni*. Stimatore wavelet di densità con *Daub#7*, raggio $r = 0.2$ e scala $j = 0$.

2.5.1 Stimatore

In questa sezione verrà fatto riferimento allo stimatore wavelet di funzioni di densità di probabilità nella forma (2.19), ossia

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} q_{\bar{j}}(x, X_i)$$

dove

$$q_{\bar{j}}(x, X_i) = \sum_k \phi_{\bar{j},k}(x)\phi_{\bar{j},k}(X_i)$$

Questo stimatore può essere espresso nel caso bivariato come

$$\hat{f}(x, y) = n^{-1} \sum_{i=1}^{n} q_{\bar{j}}(x, X_i) q_{\bar{j}}(y, Y_i)$$

dove

$$q_{\bar{j}}(x, X_i) = \sum_{k_1} \phi_{\bar{j},k_1}(x)\phi_{\bar{j},k_1}(X_i)$$

$$q_{\bar{j}_1}(y, Y_i) = \sum_{k_2} \phi_{\bar{j},k_2}(y)\phi_{\bar{j},k_2}(Y_i).$$

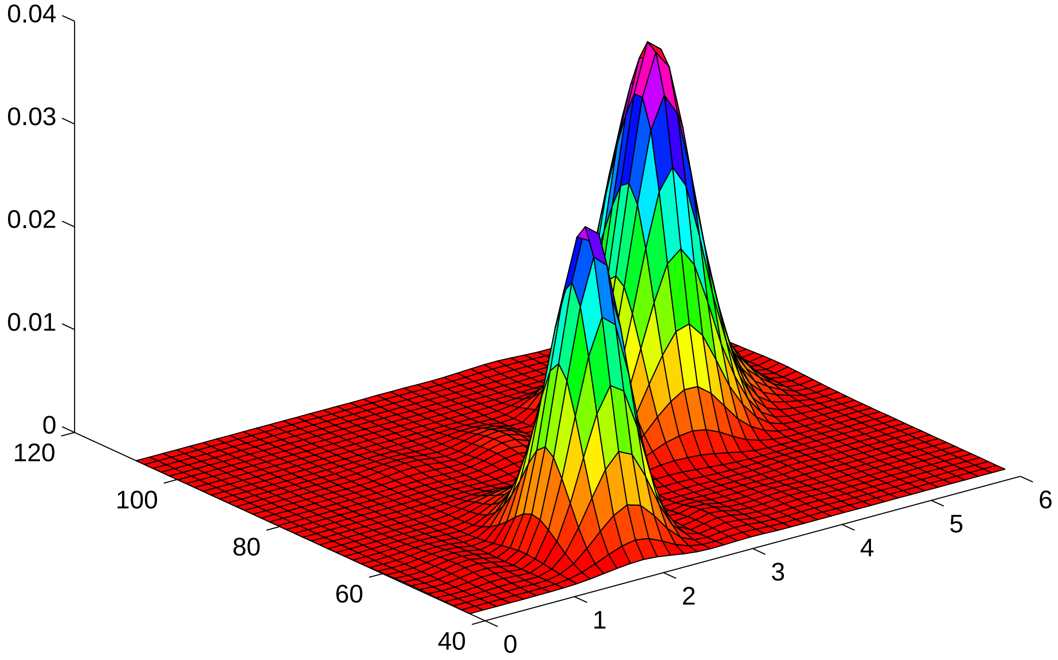

Figura 2.6: Dataset *Geyser*. Stimatore wavelet di densità bivariato, *Daub*#7, raggio $r = 0.25$ e scala $j = 0$.

Date (X,Y) variabili aleatorie, si può quindi costruire uno stimatore wavelet per la funzione di regressione $r(x) = \int \frac{yf(x,y)}{f_x(x)}dy$. Poichè

$$\int q_{\bar{\jmath}}(y, Y_i) y dy = Y_i$$

tale stimatore può essere espresso come

$$\hat{r}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i q_{\bar{\jmath}}(x, X_i)}{\frac{1}{n}\sum_{i=1}^{n} q_{\bar{\jmath}}(x, X_i)}. \tag{2.35}$$

Questo stimatore è stato proposto da Antoniadis et al., [4], ed è una versione wavelet dello stimatore con nucleo di Nadaraya e Watson; la scala $\bar{\jmath}$ gioca il ruolo della ampiezza di banda. La figura 2.7 mostra i grafici di $q_{\bar{\jmath}}(x_0, y)$ per $y \in [0,1]$, $x_0 = 0.3, 0.5, 0.8$, $\bar{\jmath} = 3$ e wavelets $Daub\#8$. Si osservi come tali grafici siano centrati in $x = x_0$. Alcune proprietà della funzione nucleo sono riportate nella sezione 2.4.1.

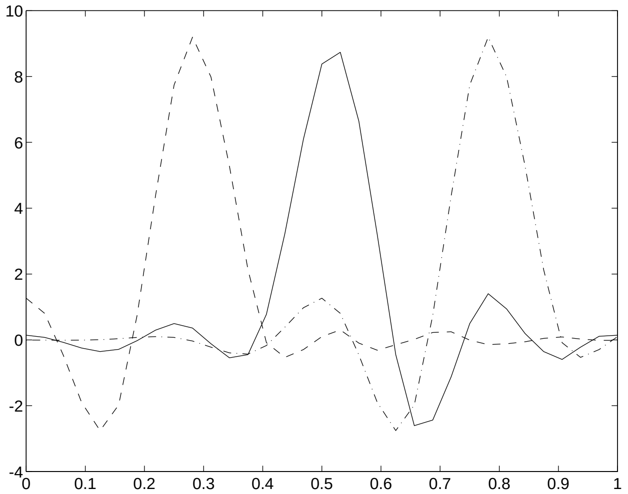

Figura 2.7: Grafici di $q_{\bar{\jmath}}(x_0, y)$ per $y \in [0,1]$, $x_0 = 0.3$ (linea tratteggiata), 0.5 (linea continua), 0.8 (linea a punti e tratti), $\bar{\jmath} = 3$ e wavelets $Daub\#8$.

Si ha la seguente proprietà di consistenza:

Proposizione 2.5.1 *Se $\bar{\jmath} \to \infty$, $n2^{-\bar{\jmath}} \to \infty$ e $E(Y^2|X = x)$ è limitata in un intorno di x, allora $\hat{r}(x)$ è consistente in media quadratica.*

In [4] gli autori dimostrano inoltre la normalità asintotica di una versione dello stimatore (2.35) opportunamente modificato allo scopo di stabilizzarne la varianza.

Il valore del parametro di scala $\bar{\jmath}$ può essere selezionato utilizzando il metodo di validazione incrociata.

2.5.2 Una Applicazione: Tassi di Cambio

Lo stimatore wavelet di funzioni di regressione è stato utilizzato in un modello nonparametrico autoregressivo con un ritardo $Y_i = r(Y_{i-1}) + \epsilon_i$ adattato alle differenze logaritmiche di tassi di cambio giornalieri, [82].

Si ricorda solamente che un tasso di cambio è il prezzo a cui una valuta nazionale può essere scambiata per un'altra divisa e che il tasso utilizzato più comunemente è quello bilaterale nominale che consiste nel numero di unità di una certa valuta offerte per una unità dell'altra. I tassi utilizzati per questa applicazione sono quelli relativi a Franco francese rispetto a Dollaro USA, registrati al mercato cambi di Londra nel periodo 3 gennaio 1989-14 gennaio 1994 per un totale di 1276 osservazioni. I dati sono stati trasformati per evitare problemi di non stazionarietà. La figura 2.8 mostra il grafico dei dati.

I risultati ottenuti sono stati confrontati con quelli, [91], relativi a stimatori con nucleo del tipo Nadaraya e Watson (nucleo Gaussiano e diversi valori dell'ampiezza di banda ottenuti moltiplicando la deviazione standard delle osservazioni per una costante) e stimatori k-vicini-più prossimi (pesi quadratici e numero variabile di vicini) descritti nella sezione 2.2.2. La bontà di adattamento delle stime ai dati è stata misurata mediante l'errore quadratico medio campionario

$$\widehat{EQM} = \frac{1}{n}\sum_{i}^{n}(Y_i - \hat{Y}_i)^2$$

e quello medio assoluto campionario

$$\widehat{EAM} = \frac{1}{n}\sum_{i}^{n}|Y_i - \hat{Y}_i|$$

dove $\hat{Y}_i$ è il valore stimato.

Le wavelets che hanno dato i risultati migliori sono state quelle $Daub\#N$. Come già osservato per lo stimatore wavelet di funzioni di densità, il supporto compatto delle wavelets di tali famiglie assicura che il campo di variazione dell'indice k_1 in $q_{\bar{\jmath}}(x, X_i)$ sia finito mentre l'algoritmo a cascata può essere utilizzato per determinare i valori della funzione di scala e della wavelet madre in un punto arbitrario.

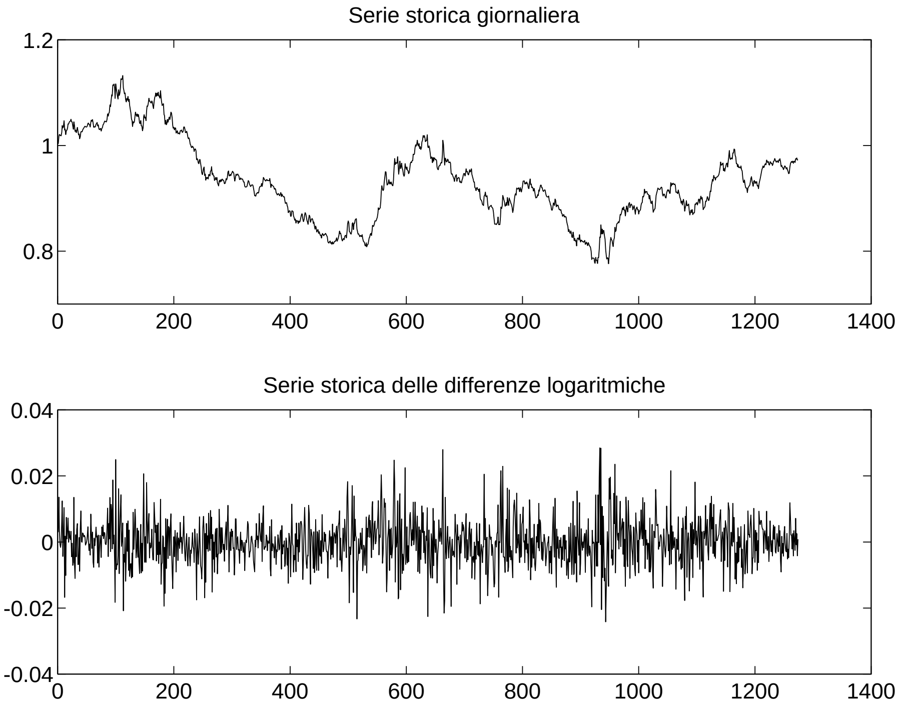

Figura 2.8: Tassi di cambio giornalieri Fr/$.

Per quanto riguarda la scelta del parametro $\bar{\jmath}$, è stato osservato durante l'applicazione che gli errori $\widehat{EQM}$ e $\widehat{EAM}$ migliorano al crescere di $\bar{\jmath}$ e che la funzione di validazione incrociata

$$CV(j) = n^{-1} \sum_{1}^{n} (Y_i - \hat{r}_{(i,i-1)}(X_i))^2,$$

dove $\hat{r}_{(i,i-1)}(x)$ è lo stimatore ottenuto rimuovendo dal campione la i-esima e la $(i-1)$-esima osservazione, coincide con l'errore $\widehat{EQM}$. Valori crescenti del parametro di scala comportano purtroppo intervalli di variazione di k_1 crescenti e quindi costi computazionali onerosi. Tuttavia, per valori di $\bar{\jmath}$ al di sopra di 20/22, i valori di $\widehat{EQM}$ e $\widehat{EAM}$ migliorano di quantità trascurabili.

La tabella 2.3 mostra i risultati ottenuti; NN(k) indica lo stimatore con k-vicini-più prossimi e NWG(h) lo stimatore con nucleo di Nadaraya e Watson con nucleo Gaussiano e ampiezza di banda h. Questi valori sono stati ottenuti in [91]. $Daub\#N(j)$ indica invece lo stimatore wavelet con wavelet madre $Daub\#N$ alla scala j più fine raggiunta.

Modello	$\widehat{EQM}$	$\widehat{EAM}$
NN(1)	$9.7999 \cdot 10^{-5}$	$7.7313 \cdot 10^{-3}$
NN(5)	$6.4130 \cdot 10^{-5}$	$6.2056 \cdot 10^{-3}$
NN(30)	$5.2610 \cdot 10^{-5}$	$5.5017 \cdot 10^{-3}$
NWG($0.05 \times \hat{\sigma}$)	$4.5257 \cdot 10^{-5}$	$5.1026 \cdot 10^{-3}$
NWG($0.2 \times \hat{\sigma}$)	$4.7868 \cdot 10^{-5}$	$5.2225 \cdot 10^{-3}$
NWG($0.9 \times \hat{\sigma}$)	$4.9123 \cdot 10^{-5}$	$5.2661 \cdot 10^{-3}$
Daub#1(24)	$1.4320 \cdot 10^{-6}$	$1.6707 \cdot 10^{-4}$
Daub#3(23)	$1.5700 \cdot 10^{-6}$	$1.9206 \cdot 10^{-4}$
Daub#8(23)	$1.4716 \cdot 10^{-6}$	$1.8992 \cdot 10^{-4}$

Tabella 2.3: Fr/$, modello AR(1) non parametrico.

La figura 2.9 (grafico b)) mostra lo stimatore wavelet con wavelets $Daub\#1$ alla scala $\bar{\jmath} = 24$.

Si osservi, infine, che la trasformazione dei dati allo scopo di evitare problemi di non stazionarietà è necessaria per l'applicazione degli stimatori classici riportati nella tabella 2.3 ma non per quella degli stimatori costruiti utilizzando le basi wavelet. Questi ultimi infatti non richiedono la stazionarietà . Gli stimatori wavelet sono stati applicati quindi anche ai dati originali (non trasformati). I risultati ottenuti sono paragonabili a quelli precedenti relativi ai dati trasformati: gli errori $\widehat{EQM}$ e $\widehat{EAM}$ raggiungono un ordine di grandezza rispettivamente di 10^{-6} e 10^{-4} intorno alla scala $j = 20$. Al di sopra di tale livello i risultati tendono

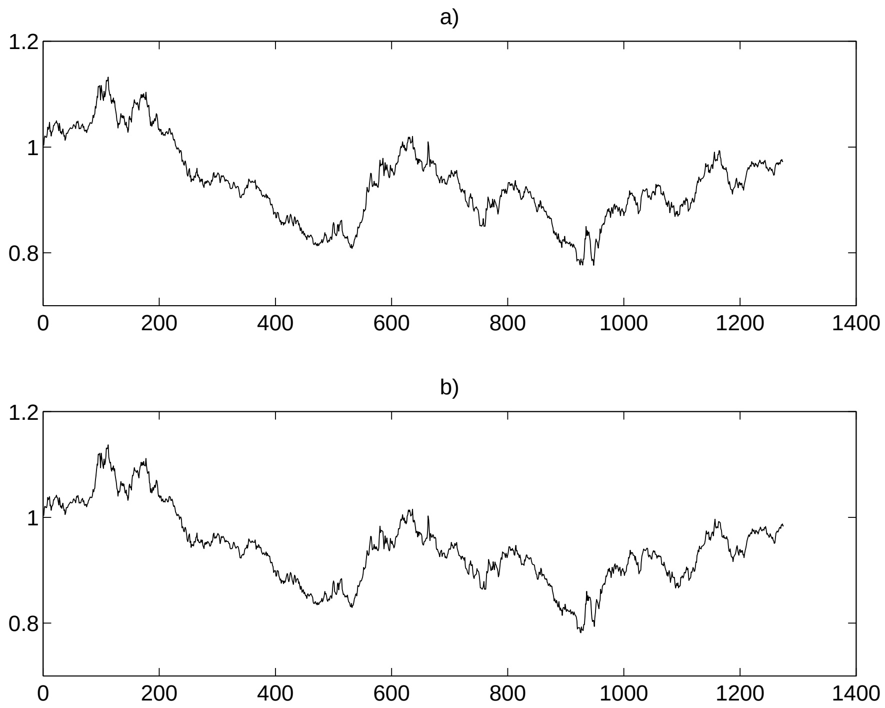

Figura 2.9: a) Tassi di cambio giornalieri Fr/\$. b) Stimatore wavelet con $Daub\#1$ e scala $j = 24$ calcolato nelle osservazioni.

poi a migliorare di quantità trascurabili.

I risultati ottenuti in questa applicazione confermano la validità delle wavelets come strumento di analisi. Il confronto con i metodi classici evidenzia poi come questo strumento possa essere competitivo e, nel caso di fenomeni non stazionari, addirittura migliore.

Capitolo 3

Wavelet Shrinkage e altre Applicazioni

3.1 Premessa

Un campo di applicazione delle wavelets che ha riscosso notevole successo è quello della *regressione non parametrica non casuale* (fixed nonparametric regression) dove il problema è quello di rimuovere un rumore da una serie di dati osservati modellabili come $y_i = f(t_i) + \epsilon_i$. Si tratta dunque, secondo una terminologia più statistica, di *smoothing* (lisciamento) di serie storiche. La tecnica che utilizza le wavelets è nota come *wavelet shrinkage*: determinata la scomposizione della serie storica in un insieme di coefficienti wavelet, il rumore viene rimosso scegliendo un valore di soglia e "comprimendo" i coefficienti wavelet verso lo zero.

Nella prima parte del capitolo saranno brevemente introdotte la trasformata discreta wavelet e la sua inversa, strumenti teorici essenziali per la comprensione della tecnica di wavelet shrinkage che sarà presentata subito dopo. Descritte le diverse proposte relative alla scelta del parametro di soglia e a quella della tecnica di compressione dei coefficienti, particolare enfasi sarà poi data ad un approccio che utilizza un modello Bayesiano per i coefficienti wavelet. L'applicazione della analisi Bayesiana in questo contesto rappresenta un nuovo tema di ricerca che sta riscuotendo particolare attenzione; i risultati ottenuti dalla applicazione del modello presentato ad una serie storica di prezzi azionari incoraggiano ulteriori approfondimenti.

La seconda parte del capitolo è invece dedicata alla descrizione di altre applicazioni delle wavelets ritenute di interesse statistico. In particolare, la tecnica di wavelet shrinkage verrà utilizzata per la stima di log-spettri. Una versione modificata di tale tecnica sarà invece adattata al caso di outliers in una serie storica. Infine, verranno introdotte, in analogia con l'analisi spettrale di Fourier, le nozioni di spettro wavelet e di scalogramma: determinata la scomposizione di una funzione f nell'insieme dei coefficienti wavelet, queste funzioni permettono di descrivere la distribuzione dell'energia di f (la norma L_2) al variare dei parametri di scala e di traslazione. L'applicazione di questi strumenti nell'ambito dell'analisi di serie storiche è una delle più recenti applicazioni statistiche delle wavelets e promette interessanti risultati.

Riferimenti bibliografici saranno dati nel corso del testo.

3.2 Trasformata Discreta Wavelet e Trasformata Inversa

Nel capitolo precedente sono state introdotte le basi wavelet ortonormali: determinata la funzione di scala ϕ come soluzione della equazione di dilatazione

$$\phi(x) = \sqrt{2}\sum_l h_l \phi(2x - l)$$

e la wavelet madre ψ come soluzione di

$$\psi(x) = \sqrt{2}\sum_l g_l \phi(2x - l), \quad g_l = (-1)^l \overline{h_{1-l}}$$

un sistema ortonormale wavelet è una collezione di funzioni ottenute come traslazioni e dilatazioni di ϕ e ψ

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

tali che la famiglia $\{\psi_{j,k}(x), j, k \in \mathbb{Z}\}$ è una base ortonormale per $L_2(\mathbb{R})$. Data $f \in L_2(\mathbb{R})$, tali basi consentono di sviluppare f in serie wavelet.

Una delle proprietà più interessanti delle basi ortonormali wavelet è costituita dal fatto che esistono relazioni ricorsive che coinvolgono coefficienti di scala, $c_{j,k} = \langle f, \phi_{j,k}\rangle$, e coefficienti wavelet, $d_{j,k} = \langle f, \psi_{j,k}\rangle$, relativi a differenti valori del parametro j. E' possibile infatti dimostrare, [17], che

$$c_{j,k} = \sum_m \overline{h_{m-2k}} c_{j+1,m}, \tag{3.1}$$

che rappresenta una relazione ricorsiva tra i coefficienti di scala al livello j e quelli al successivo livello $j+1$, e che

$$d_{j,k} = \sum_m \overline{g_{m-2k}} c_{j+1,m}, \tag{3.2}$$

che è una relazione ricorsiva tra coefficienti ai livelli j e $j+1$. Utilizzando queste relazioni, coefficienti di scala e coefficienti wavelet relativi al livello j possono essere calcolati direttamente dai coefficienti di scala del livello più fine $j+1$.

Nella direzione opposta, i coefficienti di scala al livello $j+1$ possono essere ricostruiti a partire da coefficienti di scala e coefficienti wavelet al livello più grossolano j come

$$c_{j+1,k} = \sum_m h_{k-2m} c_{j,m} + \sum_m g_{k-2m} d_{j,m}. \tag{3.3}$$

Queste relazioni possono essere espresse in una forma più semplice se si considerano gli operatori lineari

$$H \ : \ (Ha)_k = \sum_n \overline{h_{n-2k}} a_n$$

$$G \ : \ (Ga)_k = \sum_n \overline{g_{n-2k}} a_n$$

dove $a_n \in l_2(\mathbb{Z})$. Le relazioni (3.1) e (3.2) diventano infatti

$$c^{(j)} = Hc^{(j+1)}$$

e

$$d^{(j)} = Gc^{(j+1)}$$

dove $c^{(j)}$ è il vettore che ha come componenti i coefficienti di scala al livello j (e analogamente per $d^{(j)}$). La relazione (3.3) può invece essere riformulata come

$$c^{(j+1)} = H^* c^{(j)} + G^* d^{(j)}$$

dove H^* e G^* sono gli operatori aggiunti

$$H^* \ : \ (H^* a)_n = \sum_k h_{n-2k} a_k$$

$$G^* \ : \ (G^* a)_n = \sum_k g_{n-2k} a_k.$$

La relazione (3.1) è una operazione di convoluzione seguita da una di decimazione (*down sampling*) che consiste nel considerare solo i termini con indice pari, mentre la (3.3) è una operazione di *up sampling*, che consiste nell'inserimento di uno zero tra ogni coppia di valori, seguita da una operazione di convoluzione.

Queste operazioni di convoluzione hanno una interessante interpretazione in termini di scomposizione dello spettro di frequenza: secondo la terminologia ingegneristica, [120], H è infatti un filtro *passa basso*, che consente cioè il passaggio delle componenti con basse frequenze, per l'intervallo $[0, \pi/2]$, mentre G è un filtro *passa alto*, che consente cioè il passaggio delle componenti con alte frequenze, per l'intervallo $[\pi/2, \pi]$.

L'utilità delle relazioni ricorsive (3.1) e (3.3) risiede nel fatto che possono essere utilizzate per costruire algoritmi veloci per la scomposizione di una funzione $f \in L_2(\mathbb{R})$ in un insieme di coefficienti e per la ricostruzione di f a partire da questi, [108] [122].

Si consideri $c^{(m)} = (c_1, ..., c_n)$, $n = 2^m$, m fissato, un vettore di osservazioni relative ai valori che una funzione f assume in n punti equispaziati. Tale vettore può essere considerato come la proiezione di f nello spazio V_m della MRA (analisi a più livelli di risoluzione, vedi sezione 2.3.1) generata dalla famiglia di wavelets e quindi come una approssimazione di f alla scala m. Si osservi inoltre che, per un valore del parametro di scala sufficientemente alto, le funzioni di scala corrispondenti agiscono come funzioni delta di Dirac poichè hanno valore costante in un intervallo molto piccolo. I coefficienti di scala a tale livello rappresentano allora un campione su f poichè (vedi [17] per i dettagli matematici)

$$c_{j,k} = \langle f, \phi_{j,k} \rangle = 2^{j/2} f(2^j k).$$

Non è quindi riduttivo supporre che il campione $c^{(m)}$ sia in realtà, per m sufficientemente grande, il vettore dei coefficienti di scala relativi ad f al livello m. A questo punto, le relazioni (3.1) e (3.2) possono essere applicate al vettore delle osservazioni per ottenere il vettore dei coefficienti di scala e quello dei coefficienti wavelet al livello inferiore $m-1$. La procedura può essere ripetuta a tale livello per ottenere i coefficienti al livello $m-2$, eccetera. Ad ogni livello la lunghezza dei vettori sarà dimezzata per effetto dell'operazione di decimazione; la procedura quindi terminerà (nel senso che gli operatori H e G non possono essere ulteriormente applicati) quando i due vettori di coefficienti hanno lunghezza uno. In formule, tale algoritmo può essere rappresentato come

$$c^{(m)} \longrightarrow (d^{(m-1)}, d^{(m-2)}, \cdots, d^{(0)}, c^{(0)})$$

cioè

$$c^{(m)} \longrightarrow (Gc^{(m)}, GHc^{(m)}, GH^2 c^{(m)}, \cdots, GH^{m-1} c^{(m)}, GH^m c^{(m)}).$$

Questo algoritmo permette di scomporre il vettore delle osservazioni in un altro vettore che ha ancora lunghezza 2^m e che è costituito da una approssimazione di f ad un livello più grossolano e da "dettagli" relativi ai livelli successivi più fini $0, \cdots, m-1$. Tale procedura prende nome *Trasformata Wavelet Discreta* e sarà indicata con il simbolo DWT. L'algoritmo (inverso) di ricostruzione di f, è ottenuto applicando la (3.3) e si ha

$$(d^{(m-1)}, d^{(m-2)}, \cdots, d^{(0)}, c^{(0)}) \longrightarrow c^{(m)} = \sum_{j=0}^{n-1} (H^*)^j G^* d^{(j)} + (H^*)^n c^{(0)}.$$

$$
\begin{array}{ccccccc}
c^{(m)} & \longrightarrow & c^{(m-1)} & \longrightarrow & \cdots & \longrightarrow & c^{(m-r)} \\
 & \searrow & & \searrow & & \searrow & \\
 & & d^{(m-1)} & & \cdots & & d^{(m-r)}
\end{array}
$$

Figura 3.1: Trasformata Wavelet Discreta (DWT)

$$
\begin{array}{ccccccc}
c^{(m-r)} & \longrightarrow & c^{(m-r+1)} & \longrightarrow & \cdots & \longrightarrow & c^{(m)} \\
 & \nearrow & & \nearrow & & \nearrow & \\
d^{(m-r)} & & d^{(m-r+1)} & & \cdots & &
\end{array}
$$

Figura 3.2: Trasformata Wavelet Inversa (IWT)

Questo ha nome *Trasformata Wavelet Inversa* e sarà indicato con IWT. La coppia di algoritmi DWT e IWT è nota anche come *algoritmo piramidale di Mallat*. Le figure 3.1 e 3.2 illustrano graficamente i due algoritmi. L'ordine di complessità di ciascuno, cioè il numero di operazioni da compiere, è proporzionale ad n.

Si osservi che, per effetto dei filtri H e G, il primo passo dell'algoritmo DWT divide lo spettro di $c^{(m)}$ in due bande, una alta e una bassa. Al secondo passo la banda bassa è ulteriormente divisa in una banda bassa ed una alta, eccetera. Quindi il primo passo divide lo spettro in due metà uguali, il secondo divide la prima metà in quarti, eccetera. Il risultato della DWT è quindi una successione di ampiezze di banda in scala logaritmica.

3.3 Wavelet Shrinkage

La tecnica di *wavelet shrinkage* è stata proposta nel 1992 da Donoho e Johnstone ed è finalizzata alla eliminazione di un rumore Gaussiano da una serie di dati osservati. Tale procedura coinvolge gli algoritmi di scomposizione e ricostruzione descritti nella sezione precedente.

3.3.1 Procedura

Sia $y_i, i = 1, ..., n$ una successione di osservazioni modellabili come

$$y_i = f(t_i) + \epsilon_i$$

dove $f(t_i)$, $i = 1, ..., n$ sono i valori che una funzione f assume in n punti equispaziati $t_i, i = 1, ..., n$ con $n = 2^m$, m fissato e dove le ϵ_i sono supposte variabili aleatorie i.i.d. normali con media zero e varianza σ^2. Lo scopo è quello di rimuovere la componente erratica (rumore) ϵ dai dati. Donoho e Johnstone, [40] [41] [43] [45], propongono una tecnica che utilizza le wavelets nota come *wavelet shrinkage*. Essa può essere descritta come una procedura che consiste di tre passi:

(i) Scomposizione wavelet dei dati utilizzando la DWT.

Questa trasformazione determina la scomposizione del vettore delle osservazioni y in un insieme di coefficienti

$$y \longrightarrow (d^{(m-1)}, d^{(m-2)}, \cdots, d^{(0)}, c^{(0)}).$$

Questa trasformazione può essere indicata in forma matriciale [1] come

$$y_{dwt} = Wy$$

dove y_{dwt} è il vettore dei coefficienti e W è una matrice opportunamente costruita utilizzando i coefficienti di filtro h_l che compaiono nella equazione di dilatazione che definisce la funzione di scala ϕ. Poichè W è una matrice ortogonale si ha

$$y_{dwt} = Wy = Wf + W\epsilon$$

dove $W\epsilon$ è ancora un rumore Gaussiano con media zero e varianza $\sigma^2 I$. La trasformata wavelet, quindi, non altera la struttura dei dati

$$y = f + \epsilon \longrightarrow y_{dwt} = f_{dwt} + \epsilon_{dwt}. \tag{3.4}$$

(ii) Shrinkage (compressione) dei coefficienti wavelet verso lo zero

$$y_{dwt} \longrightarrow \eta(y_{dwt})$$

[1] Tale notazione viene utilizzata solo a scopo semplificativo. Nelle applicazioni, infatti, la DWT non è realizzata come prodotto matriciale ma utilizzando le relazioni ricorsive (3.1) e (3.2).

allo scopo di rimuovere il rumore dai coefficienti y_{dwt}.

Fissato un valore di soglia, tale compressione avviene sulla base di una procedura che consente di porre a zero i coefficienti che non superano la soglia e diminuire gli altri. Molte sono le proposte per la scelta del parametro di soglia e per quella della tecnica di shrinkage. Le più importanti saranno descritte in seguito.

(iii) Ricostruzione dei dati utilizzando la IWT

$$\hat{f} = W^{-1}\eta(y_{dwt}).$$

Il vettore $\hat{f}$ è il vettore dei dati dove, adesso, il rumore è stato parzialmente rimosso ed è quindi una stima di f.

La motivazione di questa tecnica deriva dal fatto che i coefficienti wavelet ottenuti dalla DWT catturano "dettagli" e quindi, qualora assumano piccoli valori, possono essere omessi senza perdere informazioni significative. L'idea è quindi quella che questi piccoli (non significativi) dettagli possono essere considerati rumore.

Un esempio può essere utile per capire meglio questa procedura. Si considerino i quattro segnali *Blocks, Bumps, HeaviSine, Doppler* i cui grafici sono riportati in figura 3.3. Le espressioni matematiche che ne consentono la riproduzione sono riportate in [43]. I nomi assegnati sono un chiaro riferimento alle caratteristiche principali dei segnali che, come si può osservare dai grafici, sono molto diverse tra loro. Questi segnali possono quindi rappresentare una vasta gamma di fenomeni. In letteratura, essi sono stati utilizzati come banco di prova per la tecnica di smoothing prima descritta.

Di ciascuno di questi segnali si considerino 1024 osservazioni, $f_i = f(t_i)$, $t_i = (i-1)/n$, $i = 1, \cdots, n$, $n = 2^{10} = 1024$. Un rumore Gaussiano a media zero e varianza uno viene aggiunto in un secondo tempo ai dati secondo il modello $y_i = f_i + \epsilon_i$. La figura 3.4 mostra il grafico dei dati con rumore. Ai dati originali f_i e a quelli con rumore y_i viene applicata la DWT.

La figura 3.5 riporta i coefficienti wavelet $d_{j,k}$ relativi alla trasformata wavelet dei dati f_i calcolata utilizzando la famiglia di wavelets Daubechies *Daub#*8 (vedi sezione 2.3.2). In tali grafici ciascuna riga corrisponde ad un livello della DWT. La riga più in alto corrisponde al livello più fine ($j = 9$ in questo esempio). Le altre righe contengono invece i coefficienti relativi ai livelli più grossolani successivi. Si ricorda che il numero di coefficienti wavelet si dimezza ad ogni livello. I coefficienti sono rappresentati in ciascuna riga mediante linee verticali che si estendono dallo zero in modo proporzionale ai valori assunti dai coefficienti

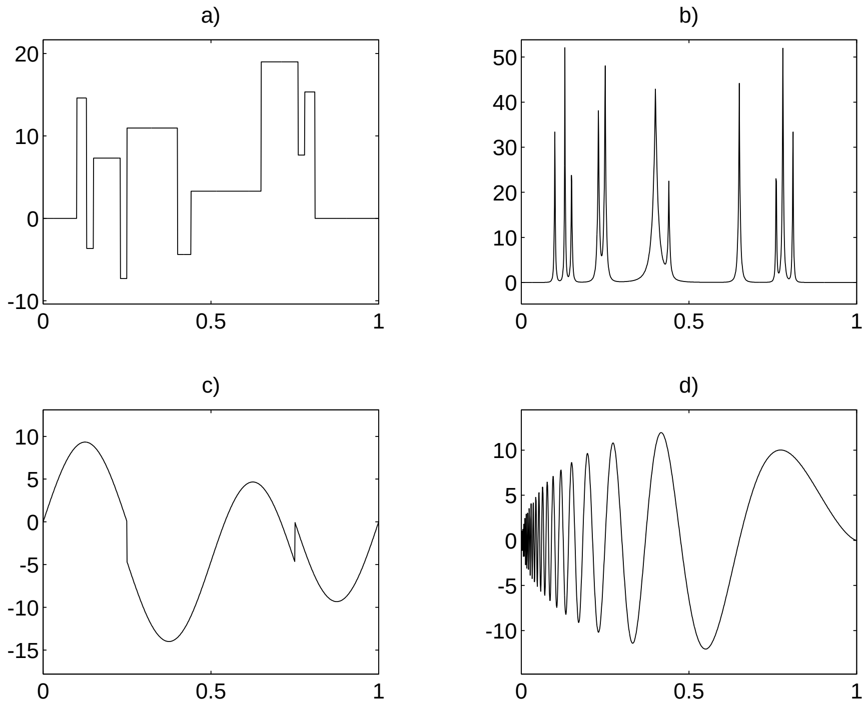

Figura 3.3: Grafici delle funzioni *a)Blocks, b)Bumps, c)HeaviSine, d)Doppler*.

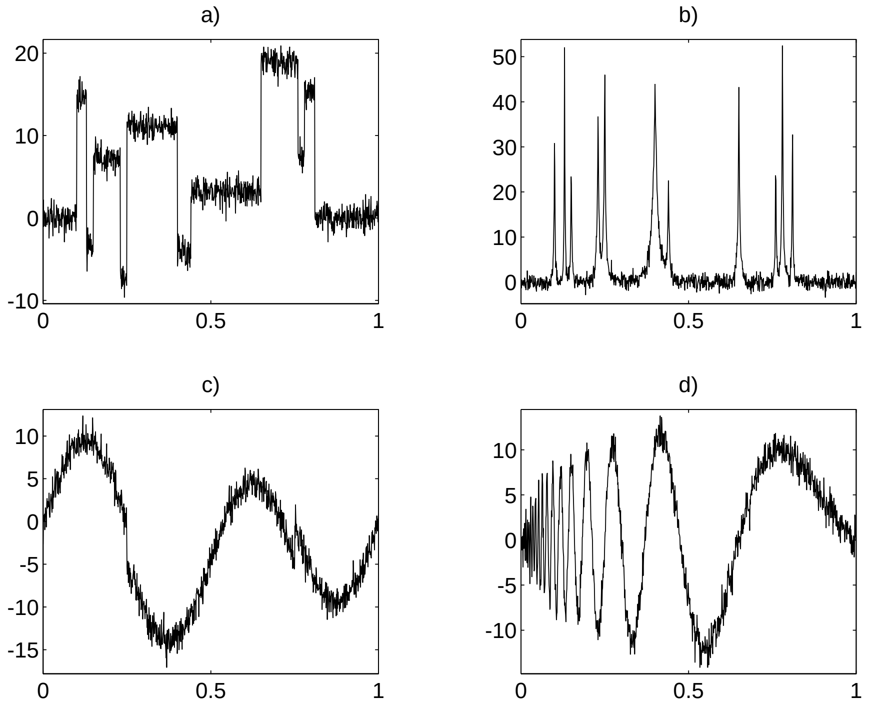

Figura 3.4: Grafici delle funzioni *a)Blocks, b)Bumps, c)HeaviSine, d)Doppler* corrotte da un rumore Gaussiano.

stessi. Si ricorda che un coefficiente wavelet $d_{j,k}$ è definito come

$$d_{j,k} = \int \psi_{j,k}(x) f(x) dx$$

e fornisce quindi informazioni su f in un intorno del punto $2^{-j}k$ e alla scala 2^j. Nei grafici 3.5 il generico coefficiente $d_{j,k}$ occupa, nella riga relativa alla scala j, la posizione $2^j(k + 1/2)$. C'è quindi, da un punto di vista empirico, una corrispondenza tra coefficienti wavelet ad una determinata posizione temporale e le parti delle funzioni corrispondenti a tale posizione. Si osservi come i coefficienti ai livelli più bassi, che catturano le caratteristiche più regolari delle funzioni (quelle alle basse frequenze), assumano valori grandi mentre quelli ai livelli più fini, che catturano ulteriori dettagli (alte frequenze), sono più piccoli. Si osservi anche come questi ultimi si addensano attorno ai picchi (singolarità) delle funzioni.

Se si considera adesso la figura 3.6, che riporta invece i coefficienti wavelet relativi alla trasformata wavelet dei dati con rumore, y_i, si può osservare come il rumore alteri in generale tutti i coefficienti ma sia particolarmente evidente ai livelli più alti.

Riassumendo, la trasformata wavelet comprime l'energia, cioè la norma L_2, di una funzione f in un piccolo numero di coefficienti (quelli ai livelli più grossolani). Questo significa che le caratteristiche più importanti della funzione sono catturate da pochi (grandi) coefficienti. Dall'altra parte, poichè la trasformata wavelet è ortogonale, nella scomposizione wavelet dei dati la componente erratica preserva la sua struttura di rumore bianco ed influenza tutti i coefficienti. Quindi, porre a zero i coefficienti più piccoli e comprimere verso lo zero gli altri permette di rimuovere il rumore e preservare le caratteristiche principali di f.

Sui coefficienti wavelet ottenuti dalla DWT dei quattro segnali con rumore prima simulati viene dunque eseguita una opportuna operazione di shrinkage. La figura 3.7 mostra i dati ricostruiti mediante la IWT.

Le elaborazioni sono state realizzate con il pacchetto Matlab *Wavbox 4.2*. Per una esposizione più completa di questo esempio vedi [43].

La tecnica di shrinkage ora illustrata presenta notevoli proprietà di efficacia e flessibilità. Donoho e Johnstone dimostrano inoltre come i risultati ottenuti con questa tecnica su gli esempi simulati prima presentati siano migliori di quelli ottenibili con tecniche di smoothing più tradizionali, quali quelle che utilizzano funzioni spline interpolanti, funzioni kernel o serie di Fourier con opportuna scelta del parametro di troncamento. Queste ultime infatti riescono, in generale, ad eliminare una parte consistente del rumore solo sacrificando anche caratteristiche significative della funzione.

Si osservi, inoltre, che il vettore $\hat{f}$ ha la stessa lunghezza del vettore y. La DWT infatti scompone la serie in un numero di coefficienti che è pari al numero

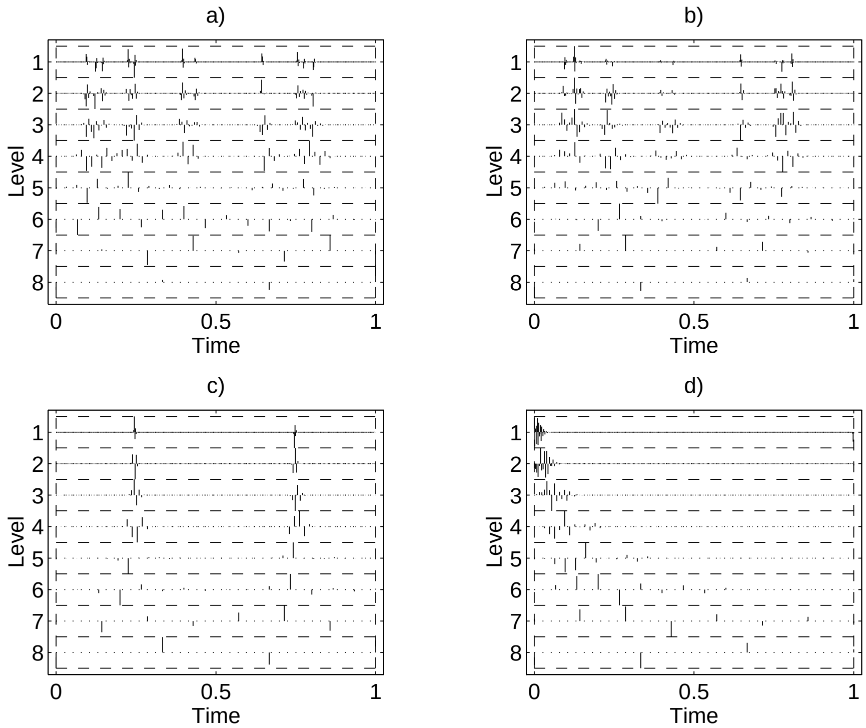

Figura 3.5: DWT delle funzioni *a)Blocks, b)Bumps, c)HeaviSine, d)Doppler.*

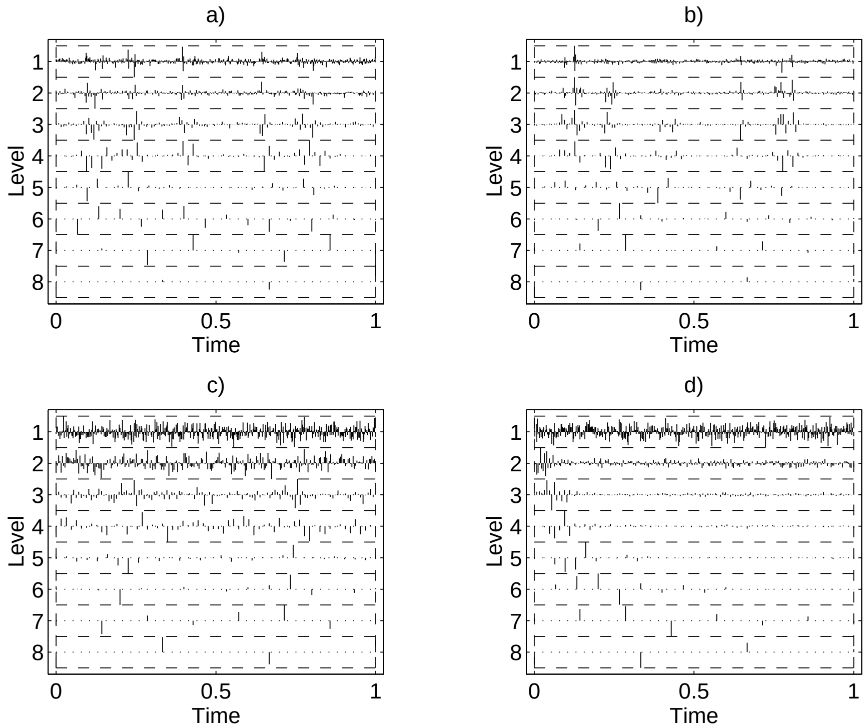

Figura 3.6: DWT delle funzioni *a)Blocks, b)Bumps, c)HeaviSine, d)Doppler* corrotte da un rumore Gaussiano.

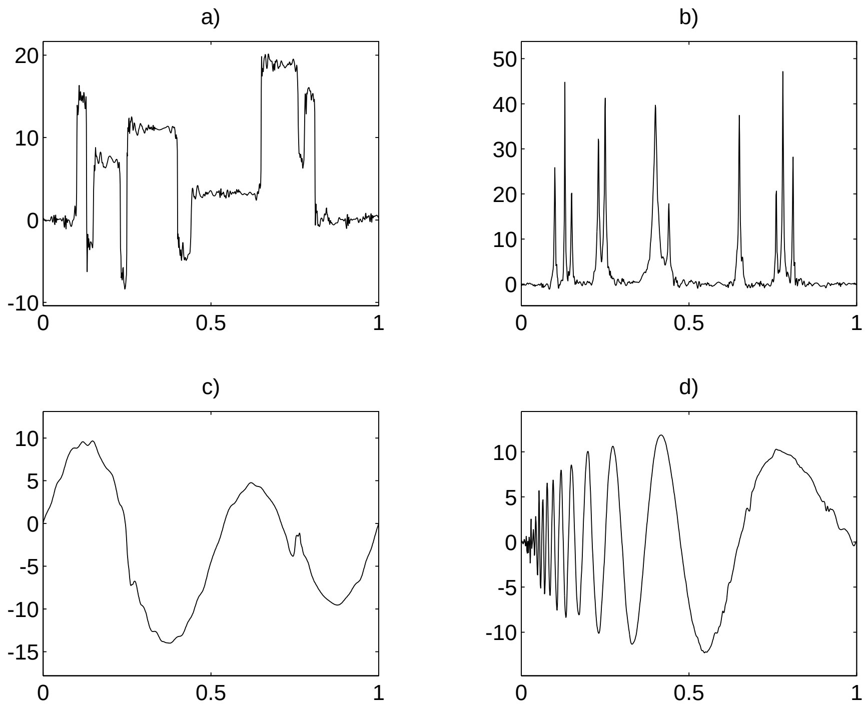

Figura 3.7: Funzioni *a)Blocks, b)Bumps, c)HeaviSine, d)Doppler* ricostruite con la IWT.

delle osservazioni. Questo significa che la tecnica di smoothing qui presentata non comporta alcuna perdita di informazioni come invece accade con altre tecniche più classiche, quali quelle che utilizzano medie mobili di dati.

Un'ottima rassegna della tecnica di wavelet shrinkage e delle sue applicazioni statistiche, alcune delle quali saranno successivamente descritte, è data da Donoho in [39].

3.3.2 Tecniche di Shrinkage

Rimane ancora da chiarire come scegliere il parametro di soglia e come comprimere, in base a tale valore, i coefficienti wavelet. Queste scelte determinano non solo la buona riuscita della tecnica di smoothing ma anche le proprietà teoriche della stima ottenuta. Vari sono i metodi proposti. Questa sezione riporta una descrizione dei principali.

Le tecniche di shrinkage più note sono senz'altro quelle proposte da Donoho e Johnstone. La più utilizzata è quella *soft threshold* (soglia dolce), secondo la quale il generico coefficiente wavelet $d_{j,k}$ viene sostituito da

$$\eta(d_{j,k}) = sign(d_{j,k})(|d_{j,k}| - \lambda)_+$$

dove $x_+ = max(x, 0)$. Questa scelta pone a zero i coefficienti che sono in valore assoluto minori di λ e comprime gli altri verso lo zero. Altra tecnica è quella di *hard threshold* (soglia forte) dove

$$\eta(d_{j,k}) = d_{j,k} I_{(|d_{j,k}| \geq \lambda)}.$$

Questa scelta pone a zero i coefficienti che sono in valore assoluto minori di λ senza comprimere gli altri. Poichè nella scomposizione wavelet di un segnale il rumore influenza tutti i coefficienti, la tecnica soft sembra essere, nella maggior parte dei casi, la più adatta.

Varie sono le proposte per la scelta del parametro di soglia λ. Si ricorda innanzi tutto che la applicazione della trasformata wavelet discreta non modifica la struttura dei dati nel senso che

$$y \rightarrow y_{dwt} = f_{dwt} + \epsilon_{dwt}$$

con ϵ_{dwt} rumore Gaussiano a media zero e varianza $\sigma^2 I$. Il problema della eliminazione del rumore dai dati osservati è quindi ricondotto a quello della eliminazione del rumore ϵ_{dwt} dai coefficienti wavelet.

Una proposta per la scelta del parametro di soglia è

$$\lambda = \sigma\sqrt{2log(n)}$$

che è nota con il nome di *soglia universale*, [45], e dove σ, deviazione standard del rumore, può essere stimata, qualora non sia nota, utilizzando i coefficienti wavelet al livello più fine. La scelta di tale soglia è motivata dal fatto che, dati $z_i,\ i = 1, \cdots, n$ i.i.d. normali con media zero e varianza σ^2, la probabilità che $max|z_i|$ superi il valore di soglia universale tende a zero per $n \to \infty$.

Altra proposta è quella di determinare valori di soglia che siano dipendenti dal livello di risoluzione. Tra questi metodi rientra il cosiddetto *metodo SURE* (Stein's Unbiased Risk Estimator). Prima di descrivere questo metodo si ricorda che, dato uno spazio funzionale F, una funzione di perdita $d(\cdot, \cdot)$ e un insieme di stimatori V, si definisce *rischio* di uno stimatore $\hat{f} \in V$ di $f \in F$ la quantità

$$R(\hat{f}, f) = E(d(\hat{f}, f)) \tag{3.5}$$

e *rischio minimax* la quantità

$$Inf_{\hat{f} \in V} Sup_{f \in F} R(\hat{f}, f).$$

Ad ogni livello j della DWT i coefficienti wavelet soddisfano il modello

$$d_{j,k} = w_{j,k} + z_{j,k} \tag{3.6}$$

con $z \sim N(0, \sigma^2)$. Si indichi con $d^{(j)}$ il vettore dei coefficienti wavelet al livello j. Lo scopo è quello di eliminare il rumore e determinare $\eta(d^{(j)})$ come stima del vettore w. Nel 1981 Stein, [106], propose per modelli del tipo (3.6) un metodo non distorto di stima del rischio (3.5) nel caso $d(\hat{f}, f) = \|\hat{f} - f\|_2^2$. Il metodo $SURE$, proposto ancora da Donoho e Johnstone, [43], consiste nel determinare ad ogni livello j della DWT un valore di soglia $\sigma\lambda_j$ dove λ_j è tale da minimizzare la stima del rischio $E(\|\eta(d^{(j)}) - w\|_2^2)$ proposta da Stein. Ovvero

$$\lambda_j = argmin_{t \geq 0} SURE(d^{(j)}, t)$$

dove

$$SURE(d^{(j)}, t) = K - 2\sum_{k=1}^{K} I_{|d_{j,k}| \leq t\sigma} + \sum_{k=1}^{K} min((d_{j,k}/\sigma)^2, t^2),$$

con K uguale alla dimensione del vettore $d^{(j)}$, è la stima non distorta del rischio proposta da Stein.

Quando i coefficienti wavelet sono sparsi, cioè quando molti dei coefficienti sono zero, è preferibile utilizzare il valore di soglia universale. Donoho e Johnstone propongono una misura della quantità di zeri ed un metodo "ibrido" che ad ogni singolo livello determina, in base a tale misura, un valore di soglia pari alla soglia universale o a quella SURE. La tecnica di shrinkage utilizzata sui dati simulati nella sezione precedente è proprio quella di soft threshold con valori di soglia determinati, ai quattro livelli più fini, mediante il metodo "ibrido".

Donoho e Johnstone determinano la ottimalità asintotica degli stimatori wavelet $\hat{f}_{DJ}$ di f ottenuti con le tecniche shrinkage da loro proposte dimostrando che questi si avvicinano al rischio minimax

$$R(\hat{f}_{DJ}, f) \approx Inf_{\hat{f} \in V} Sup_{f \in F} R(\hat{f}, f)$$

per una ampia varietà di funzioni di perdita e di spazi funzionali F.

Si riporta solo un breve cenno ad altre tecniche di shrinkage dei coefficienti wavelet proposte. Nason, [88], propone un metodo di validazione incrociata (*cross validation*) per la selezione del valore di soglia nel caso si utilizzi la tecnica di soft threshold. L'autore dimostra che, indicato con $\hat{f}$ lo stimatore wavelet di f ottenuto con tale tecnica, l'*errore quadratico integrale*

$$EQI = \int [\hat{f}(x) - f(x)]^2 dx,$$

considerato come funzione del parametro di soglia, ha quasi sempre (con probabilità alta) un minimo. L'EQI può essere stimato, in generale, con una procedura di validazione incrociata che consiste nel calcolare lo stimatore rimuovendo dal campione l'i-esima osservazione e valutare le differenze tra lo stimatore ottenuto calcolato nell'i-esima osservazione e l'osservazione stessa. Nason adatta tale procedura al caso dello stimatore wavelet separando le osservazioni con indice pari da quelle con indice dispari. Il valore di soglia viene poi determinato minimizzando l'EQI stimato.

Goel e Vidakovic, [53], propongono invece un diverso approccio. Data una variabile aleatoria X con $E(X) = \mu$ e con funzione di ripartizione strettamente crescente e continua in μ, si definisce *curva di Lorentz* di X la funzione

$$L(p) = \frac{1}{\mu} \int_0^{\xi_p} x dF(x)$$

dove ξ_p è il p-esimo quantile. Motivati dal fatto che la maggior parte della energia di f è catturata da un piccolo numero di (grandi) coefficienti, i due autori analizzano la curva di Lorentz relativa ai coefficienti wavelet empirici e considerano solo quelli con energia più grande di C volte l'energia media, ovvero solo i coefficienti $d_{j,k}$ tali che $d_{j,k}^2 > C\overline{d}^2$, dove $\overline{d}^2$ è la media dei quadrati dei coefficienti. Questo consente di conservare la maggior parte dell'energia di f. Per una esposizione più completa vedi [53].

3.3.3 Approccio Bayesiano

Questa sezione descrive un approccio di tipo Bayesiano alla tecnica di wavelet shrinkage: determinato un modello per i coefficienti wavelet, lo shrinkage avviene sostituendo ai coefficienti stessi stime a posteriori risultato di una analisi Bayesiana. Questo approccio costituisce un nuovo tema di ricerca e sta riscuotendo particolare attenzione. Il modello descritto in questa sezione è attualmente oggetto di studio, Vidakovic e Müller, [123].

Si ricorda, ancora una volta, che la trasformata wavelet è lineare ed ortogonale e che quindi i coefficienti wavelet ottenuti applicando ai dati la DWT soddisfano un modello del tipo

$$y_{dwt} = f_{dwt} + \epsilon_{dwt} \tag{3.7}$$

dove ϵ_{dwt} è un rumore Gaussiano a media zero e varianza $\sigma^2 I$. Si può dunque assumere

$$[y_{dwt}|f_{dwt}, \sigma^2] \sim \mathcal{MVN}(f_{dwt}, \sigma^2 I)$$

dove il simbolo $\mathcal{MVN}(m, \Sigma)$ indica una distribuzione normale multivariata con media m e matrice di covarianza Σ. Il problema dell'eliminazione del rumore è ricondotto quindi a quello della stima del valore medio f_{dwt}.

Un primo approccio Bayesiano alla tecnica di wavelet shrinkage è stato proposto da Vidakovic in [121]. Poichè le regole Bayesiane sono, in generale, degli "shrinkers" (nel senso che tendono a comprimere i valori), la "compressione" dei coefficienti wavelet avviene, in questo caso, sostituendo ai coefficienti stessi le stime di Bayes a posteriori. L'autore assume tuttavia l'indipendenza delle a priori, ipotesi questa non sempre ragionevole.

In questo contesto verrà fatto riferimento ad un'altra scelta della a priori. Si consideri un generico modello lineare del tipo

$$y = X\beta + \epsilon \tag{3.8}$$

dove gli elementi del vettore ϵ sono incorrelati con media zero e varianza comune σ^2. In [90] O'Hagan propone per tale modello la distribuzione a priori *normal-inverse-gamma*

$$f(\beta, \sigma^2] \sim \mathcal{NIG}(\alpha, \delta, m, \Sigma)$$

la cui funzione di densità è data da

$$f(\beta, \sigma^2|\alpha, \delta, m, \Sigma) = C(\sigma^2)^{\frac{\delta+p+2}{2}} \exp[-\{(\beta - m)'\Sigma^{-1}(\beta - m) + \alpha\}/(2\sigma^2)]$$

dove p è la dimensione del vettore dei parametri β.

La funzione di verosimiglianza del modello (3.8) è

$$f(y,|\beta,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp[-(y - X\beta)'(y - X\beta)/(2\sigma^2)]$$

e la $\mathcal{NIG}$ è una famiglia coniugata di distribuzioni a priori. La distribuzione marginale a priori di β è

$$f(\beta) = C_1(1 + (\beta - m)'(\alpha\Sigma)^{-1}(\beta - m))^{-(\delta+p)/2}$$

ed è simmetrica rispetto a m con media $E(\beta) = m$ e varianza finita $var(\beta) = (\alpha/(\delta - 2))\Sigma$ se $\delta > 2$. La distribuzione marginale a priori di σ^2 è

$$f(\sigma^2) = C_2(\sigma^2)^{-(\delta+2)/2} exp(-\alpha/(2\sigma^2))$$

con media $E(\sigma^2) = \alpha/(\delta-2)$ se $\delta > 2$ e varianza $var(\sigma^2) = 2\alpha^2/((\delta-2)^2(\delta-4))$ se $\delta > 4$. La funzione di densità $\mathcal{NIG}$ è propria se $\alpha > 0$, $\delta > 0$ e se Σ è definita positiva.

Nel caso del modello (3.7) è ragionevole considerare una distribuzione a priori del tipo

$$[f_{dwt}, \sigma^2] \sim \mathcal{NIG}(\alpha, \delta, 0, \Sigma).$$

La scelta $m = 0$ è motivata dal fatto che i coefficienti devono essere compressi in modo simmetrico rispetto allo zero. La compressione è determinata dalla scelta della matrice Σ. Una possibile struttura è: si assuma $y_{dwt} = (c^{(j_0)}, d^{(j_0)}, \cdots, d^{(j_1)})$ dove j_0 è il livello più piccolo (grossolano) raggiunto nella DWT e j_1 quello più alto (fine), si ha

$$\Sigma = \lambda_{j_0,\phi}\Sigma_{j_0,\phi} \oplus \bigoplus_{j=j_0}^{j_1} \lambda_{j,\psi}\Sigma_{j,\psi}.$$

Le matrici $\Sigma_{j_0,\phi}$ e $\Sigma_{j,\psi}$, $j = j_0, \cdots, j_1$ descrivono la struttura di covarianza rispettivamente dei coefficienti di scala al livello j_0 e di quelli wavelet ai livelli j, $j = j_0, \cdots, j_1$. La matrice Σ, costruita come somma diretta $\oplus$ tra tali matrici, ha quindi una struttura diagonale a blocchi dove i blocchi sono costituiti dalle matrici $\Sigma_{j_0,\phi}$ e $\Sigma_{j,\psi}$, $j = j_0, \cdots, j_1$ moltiplicate rispettivamente per i parametri $\lambda_{j_0,\phi}$ e $\lambda_{j,\psi}$, $j = j_0, \ldots, j_1$. Una possibile scelta degli elementi delle matrici di covarianza relative ai coefficienti ai singoli livelli è $\sigma_{i,j} = \rho^{|i-j|}$, $|\rho| < 1$. La struttura di Σ così determinata equivale a supporre che coefficienti a livelli diversi siano indipendenti mentre coefficienti allo stesso livello sono correlati secondo una relazione di dipendenza inversamente proporzionale alla distanza $d(\xi_1, \xi_2) = |\xi_1 - \xi_2|$ tra i coefficienti stessi.

I parametri $\lambda_{j_0,\phi}$ e $\lambda_{j,\psi},\ j = j_0, \ldots, j_1$ determinano la compressione dei coefficienti e devono essere opportunamente scelti. Risultati sperimentali hanno dimostrato che valori "grandi" non comprimono i coefficienti e valori "piccoli" si. Poichè, allo scopo di eliminare il rumore, la compressione deve essere più forte a livelli più fini, è consigliabile scegliere i parametri di smoothing λ in modo esponenzialmente decrescente.

La distribuzione a posteriori è ancora una normale-inversa-gamma

$$[f_{dwt}, \sigma^2 | y_{dwt}] \sim \mathcal{NIG}(\alpha^\star, \delta^\star, m^\star, \Sigma^\star),$$

dove

$$\begin{array}{rcl} \Sigma^\star & = & (I + \Sigma^{-1})^{-1} \\ m^\star & = & \Sigma^\star y_{dwt} \quad (= E(f_{dwt} \mid y_{dwt})) \\ \alpha^\star & = & \alpha + \|y_{dwt}\| + (m^\star)'(\Sigma^\star)^{-1} m^\star \\ \delta^\star & = & \delta + n. \end{array}$$

Lo stimatore a posteriori Bayesiano $m^\star$ di f_{dwt} così determinato è il vettore dei coefficienti compressi che è stato precedentemente indicato con $\eta(y_{dwt})$ ed al quale può quindi essere applicata la IWT.

Allo scopo di verificarne la validità, questa tecnica di shrinkage è stata utilizzata per "lisciare" una serie storica di prezzi azionari giornalieri della IBM relativi al periodo 28/07/1992-06/10/1995 (1024 osservazioni). La figura 3.8 mostra la serie originaria (grafico a)) e quella lisciata (grafico b)). La famiglia di wavelets utilizzata è la *Daub#7*. La scelta dei parametri di smoothing è avvenuta verificando che la struttura dei residui fosse quanto più possibile vicina a quella di un rumore bianco Gaussiano. A tal scopo sono state utilizzate la funzione di autocorrelazione dei residui e quella dei residui al quadrato (per l'individuazione di eventuali dipendenze non lineari). La figura 3.9 mostra tali funzioni nel caso di parametri di smoothing $(100,\ 100,\ 4,\ 1/2,\ 1/4,\ 1/4^2)$ e matrici di covarianza con $\rho = 0.8$. Tale scelta è quella che ha dato i risultati migliori. Un grafico quantile-quantile valutato rispetto ai residui ed a un campione casuale simulato da una distribuzione normale è riportato in figura 3.10.

Le elaborazioni sono state realizzate in Matlab. Per la applicazione delle trasformate DWT e IWT è stato utilizzato il pacchetto *Wavbox 4.2*.

La tecnica di shrinkage Bayesiana qui presentata offre notevoli proprietà di flessibilità. In particolare, la struttura di covarianza proposta può essere modificata mediante l'introduzione di una struttura gerarchica congiuntamente ad un'analisi teorica della matrice di covarianza effettiva.

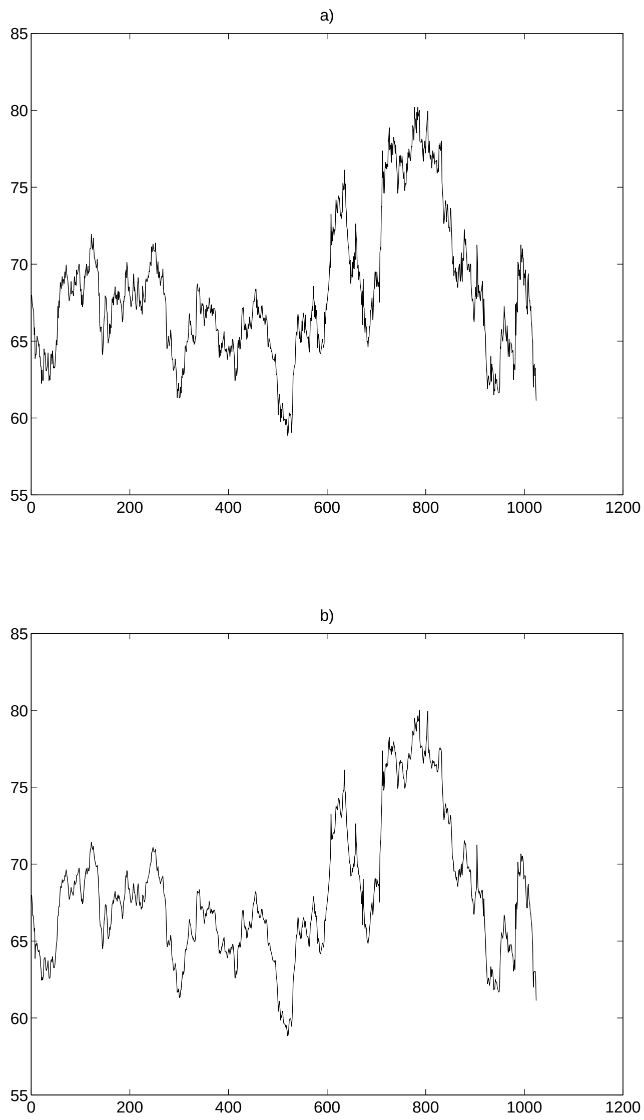

Figura 3.8: Prezzi azionari della IBM. a) Serie osservata. b) Serie lisciata con shrinkage Bayesiano.

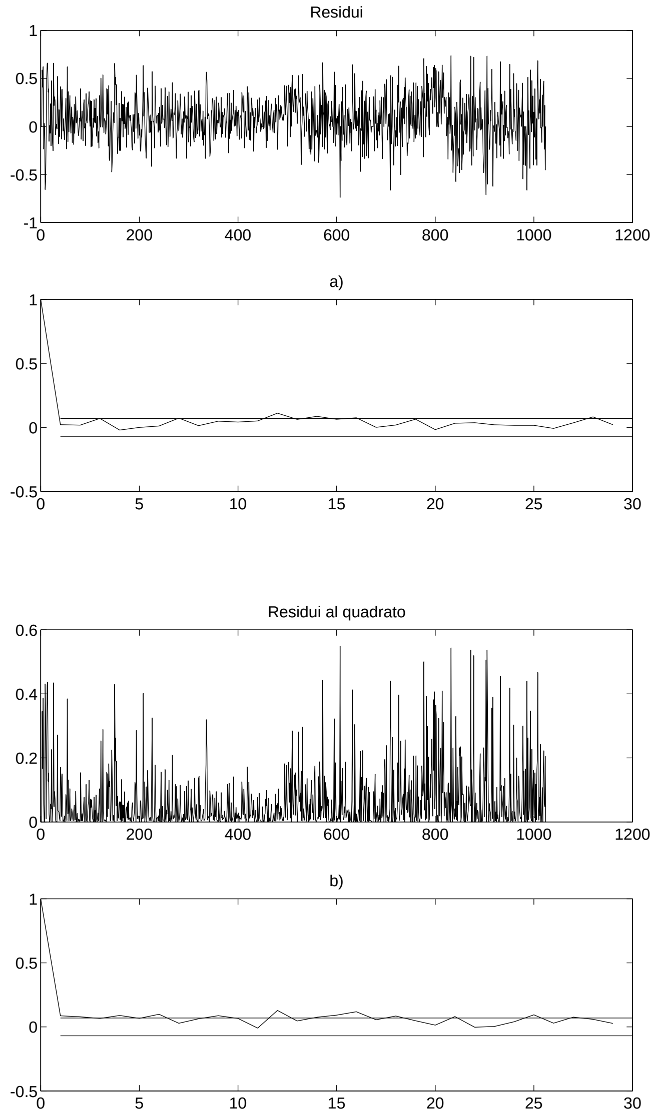

Figura 3.9: Prezzi azionari della IBM. *a*) Autocorrelazione dei residui. *b*) Autocorrelazione dei residui al quadrato.

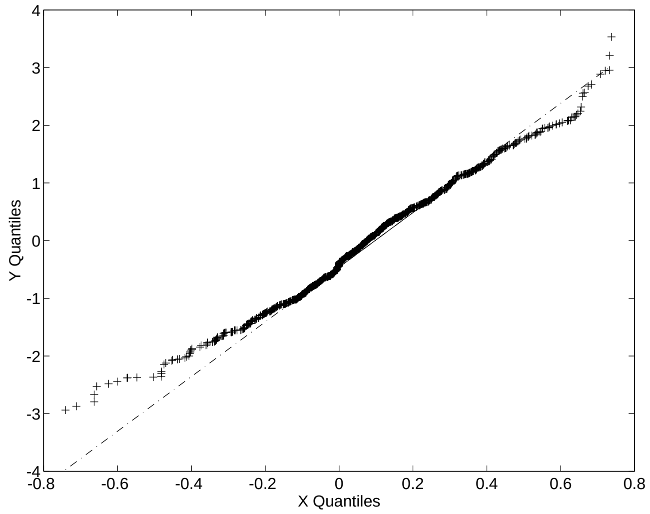

Figura 3.10: Prezzi azionari della IBM. Grafico quantile-quantile dei residui verso una Normale simulata.

3.4 Altre Applicazioni

La seconda parte di questo capitolo è dedicata alla descrizione di altre applicazioni delle wavelets ritenute di particolare interesse statistico. In particolare, la tecnica di wavelet shrinkage, precedentemente descritta, verrà utilizzata per la stima di log-spettri ed una versione modificata di tale tecnica sarà invece adattata al caso di outliers in una serie storica. Infine verranno introdotte le definizioni di spettro wavelet e di scalogramma come funzioni che descrivono, al variare dei parametri di scala e di traslazione, la distribuzione dell'energia di una funzione.

Le elaborazioni relative agli esempi riportati sono state realizzate con il pacchetto S-PLUS *S+Wavelets* [2].

3.4.1 Stima di Log-Spettri

Gao, [51] [52], riconduce il problema della stima dello spettro di una serie storica a quello di eliminazione di un rumore da una serie di dati. Questo consente la applicazione della tecnica di wavelet shrinkage di Donoho e Johnstone prima presentata.

In particolare, data una serie storica $X_t, t = 0, \cdots, 2n-1,\ n = 2^m$, realizza-

[2] Copyright (c) 1994, MathSoft. Inc.

zione finita di un processo stazionario Gaussiano del secondo ordine a media zero, si è interessati alla stima della funzione di densità spettrale

$$h(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s) cos(s\omega)$$

dove $R(s) = E(X_t X_{t+s})$ è la funzione di autocovarianza.

La funzione $h(\omega)$ viene comunemente stimata ricorrendo alla definizione di *periodogramma*

$$I_n(\omega) = \frac{1}{4n\pi} | \sum_{t=0}^{2n-1} X_t e^{-i\omega t} |^2$$

che è una stima asintoticamente corretta ma non consistente di $h(\omega)$. Stime consistenti possono essere ottenute, ad esempio, utilizzando funzioni finestra (funzioni peso nella terminologia statistica) dipendenti da un parametro che ne determina l'ampiezza (vedi [96] per una lista di possibili finestre).

Gao propone una nuova tecnica basata sulla applicazione al logaritmo del periodogramma della tecnica di shrinkage proposta da Donoho e Johnstone. Questa può essere descritta come una procedura che consiste di quattro passi:

i) Calcolo del log-periodogramma

$$z_l = log I_n(\omega_l), \quad l = 0, \cdots, n-1$$

dove $\omega_l = l/2n$.

ii) DWT di z_l

iii) Shrinkage dei coefficienti wavelet $d_{j,k}$ mediante la tecnica di soft threshold

$$\eta(d_{j,k}) = sign(d_{j,k})(|d_{j,k}| - \lambda)_+$$

con parametro di soglia dipendente dal livello j, $\lambda = \lambda_{j,n}$, scelto come

$$\lambda_{j,n} = max(\frac{\pi}{\sqrt{3}}\sqrt{logn}, \alpha_j logn). \qquad (3.9)$$

I valori delle costanti α_j, per diversi valori di j e nel caso si utilizzino le wavelets di Daubechies, sono tabulati in [51].

iv) IWT dei coefficienti. Il vettore risultante è una stima del log-spettro alle frequenze ω_l.

La tecnica di wavelet shrinkage viene dunque applicata non al periodogramma stesso ma al logaritmo di questo; la motivazione è dovuta al fatto che il passaggio al logaritmo permette di ottenere stime positive e, soprattutto, al fatto che il logaritmo è la trasformazione, proposta da Wahba, [124], che stabilizza la varianza del periodogramma. Si consideri il seguente modello con componente di rumore additiva non Gaussiana

$$z_l = g_l + \epsilon_l, \quad l = 1, \cdots, n-1 \tag{3.10}$$

dove

$$\epsilon_l = log(\xi_l/2) + \gamma$$

con ξ_l, $l = 1, \cdots, n-1$ indipendenti e $\chi_2^2 = exp(1/2)$ distribuiti e dove $\gamma \approx 0.57721...$ è la costante di Eulero-Mascheroni. E' possibile dimostrare che $E(\epsilon_l) = 0$ e $var(\epsilon_l) = \pi^2/6$. Wahba propose la (3.10) come modello per il logaritmo del periodogramma, cioè $z_l \approx logI(\omega_l)$ e $g_l = log(h(\omega_l)) - \gamma$. La trasformazione logaritmica permette quindi di ricondurre il problema della stima dello spettro a quello di smoothing di una serie di dati consentendo l'applicazione della tecnica di wavelet shrinkage. Poichè nella (3.10) il rumore è, in generale, non Gaussiano, il valore di soglia universale proposto da Donoho e Johnstone non dà buoni risultati.

La soglia (3.9) proposta da Gao è basata su un attento esame del carattere non Gaussiano del rumore e verifica la seguente

Proposizione 3.4.1 *Data la approssimazione di Wahba per il logaritmo del periodogramma, si ha*

$$P(\cup_j [Sup_k |d_{j,k} - E(d_{j,k})| > \lambda_{j,n}]) \to 0, \quad n \to \infty.$$

La figura 3.11 mostra il grafico di una serie storica AR simulata il cui spettro presenta alcuni picchi. I coefficienti AR utilizzati per la simulazione corrispondono ad un processo AR con radici complesse di modulo e fase

$$mod = (1.03, 1.03, 1.005, 1.03, 1.03, 1.03, 1.03, 1.03, 1.03, 1.03, 1.03, 1.03)$$

$$fas = (0.2, 0.4, 0.6, 0.8, 1.0, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6).$$

La figura 3.12 mostra invece lo spettro della serie AR simulata, il logaritmo del periodogramma, il logaritmo del periodogramma "lisciato" con una funzione finestra di tipo triangolare ed infine il logaritmo del periodogramma lisciato con la tecnica wavelet prima introdotta. Per una esposizione più completa di questo esempio vedi [15].

Questo stesso approccio al problema della stima di spettri è stato indipendentemente proposto da Moulin, [85] [86], che ha suggerito altre scelte del parametro di soglia ed ha generalizzato la tecnica al caso di immagini radar.

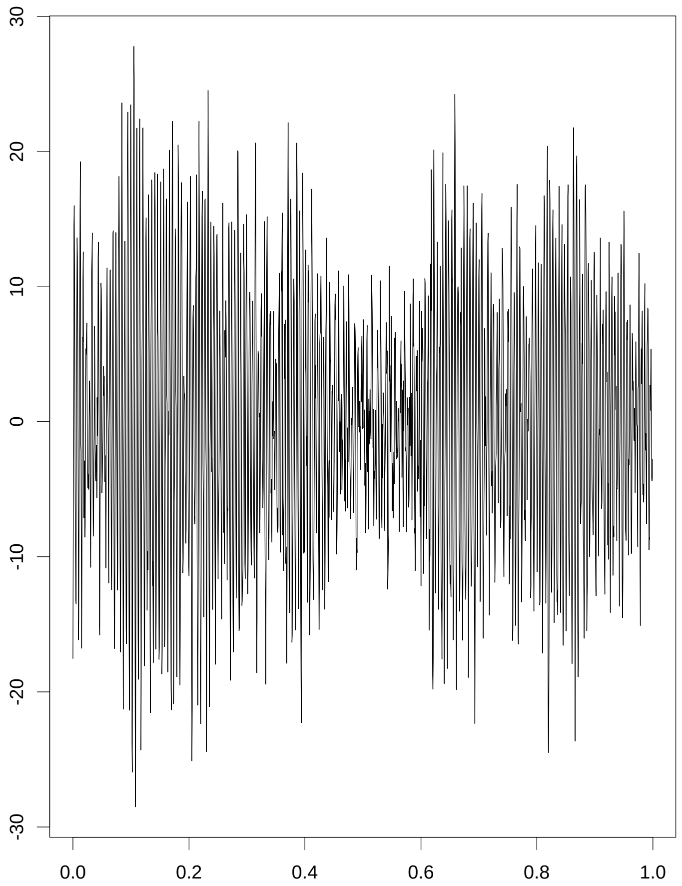

Figura 3.11: Serie storica AR simulata.

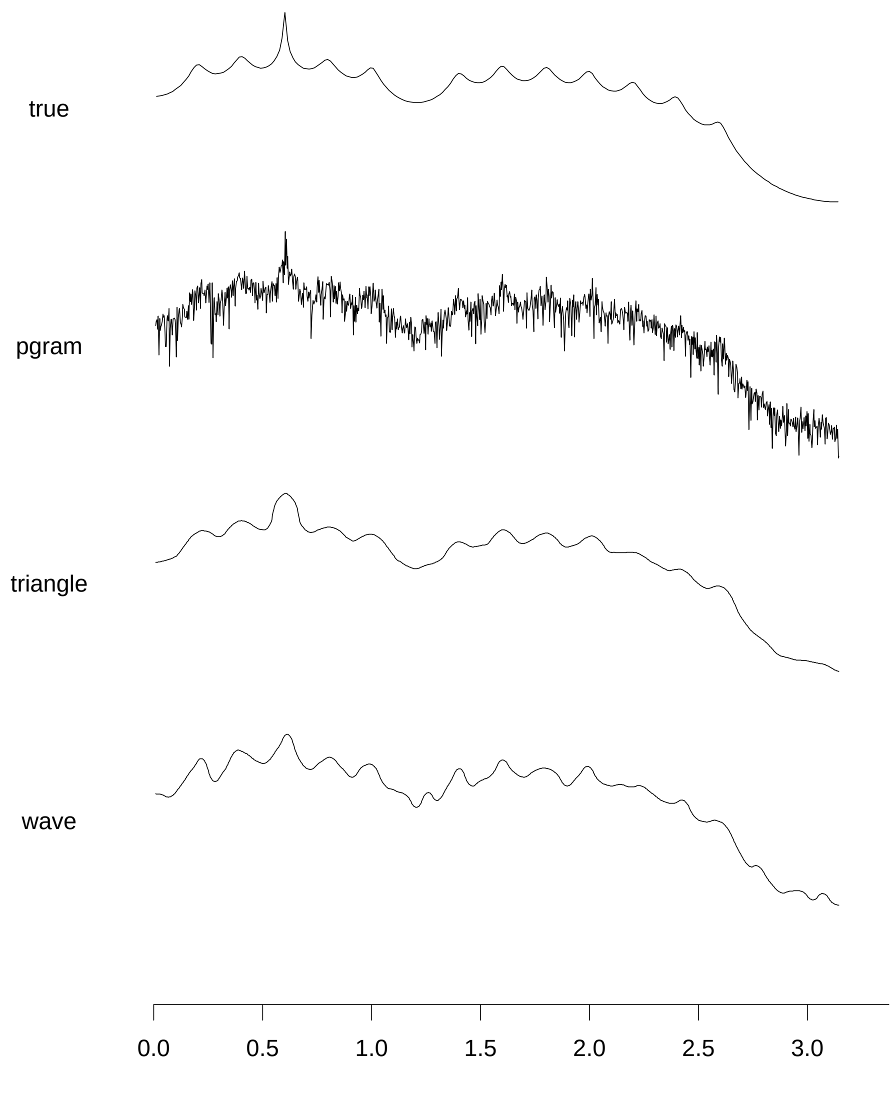

Figura 3.12: Serie storica AR simulata. Spettro (true), log-periodogramma (pgram), log-periodogramma "lisciato" con funzione finestra triangolare (triangle) e log-periodogramma lisciato con la tecnica wavelet (wave).

3.4.2 Outliers

Data una serie storica, si definisce outlier un valore localizzato ad una alta frequenza che appare, nel grafico dei dati, come una osservazione "fuori posto" e che può essere dovuto, per esempio, ad errori di misurazione. La tecnica di shrinkage proposta da Donoho e Johnstone per l'eliminazione di un rumore da un insieme di dati osservati non riesce ad eliminare gli outliers. Tale tecnica è infatti particolarmente indicata per la rimozione di rumori "di fondo", quale quello Gaussiano, che hanno quindi, rispetto al segnale, carattere globale. Un outlier è invece una caratteristica locale che è catturata, in generale, dai coefficienti più grandi ottenuti dalla DWT e che quindi non viene rimossa.

Bruce, Donoho, Gao e Martin, [15] [16], propongono un *algoritmo di smoothing robusto* (robust smoother algorithm) per la rimozione degli outliers: a partire dall'insieme delle osservazioni (viste come un vettore di coefficienti di scala ad un determinato livello), l'algoritmo produce una scomposizione wavelet del segnale introducendo, rispetto alla usuale DWT, una terza componente, detta vettore dei residui robusti, che cattura gli outliers. Ad ogni livello di risoluzione il vettore dei coefficienti di scala viene "ripulito" usando un filtro ripulitore robusto; al vettore ottenuto vengono poi applicati i due filtri H (passa basso) e G (passa banda) della DWT. In particolare, ad ogni passo l'algoritmo produce:

1) Un vettore di residui robusti dato da

$$r^{(j)} = \delta_{t_1,t_2}(c^{(j)} - \hat{c}^{(j)})$$

dove δ_{t_1,t_2} è la funzione di shrinkage

$$\delta_{t_1,t_2} = \begin{cases} 0 & |x| \le t_1 \\ sign(x)\frac{t_2(|x|-t_1)}{t_2-t_1} & t_1 < |x| \le t_2 \\ x & |x| > t_2 \end{cases}$$

e $\hat{c}^{(j)}$ è uno smoothing robusto di $c^{(j)}$ (ottenuto per esempio con mediane di 5 elementi). Le costanti t_1 e t_2 sono scelte in modo tale che la maggior parte dei residui siano zero. Soft e hard threshold possono essere ottenuti scegliendo rispettivamente $t_1 = t, t_2 = \infty$ e $t_1 = t_2 = t$.

I residui così ottenuti contengono i coefficienti che catturano gli outliers.

2) Un vettore di coefficienti di scala $c^{(j-1)}$ ottenuto applicando il filtro H al vettore dei coefficienti di scala ripuliti $u^{(j)} = c^{(j)} - r^{(j)}$.

3) Un vettore di coefficienti wavelet $d^{(j-1)}$ ottenuto applicando il filtro G al vettore $u^{(j)}$.

Gli autori suggeriscono inoltre di applicare le tecniche shrinkage di Donoho e Johnstone ai coefficienti wavelet ottenuti dopo la rimozione dei residui. Questo permette di ottenere risultati migliori.

Un esempio è riportato in figura 3.13. Il grafico in alto è relativo a misurazioni radar i cui valori sono gli angoli di un oggetto. Si indichi con il nome *glint* tale segnale. Il segnale radar "vero" è dato da piccole oscillazioni intorno allo zero mentre le osservazioni presentano outliers (picchi) dovuti ad errori di misurazione. Il secondo grafico mostra il segnale "lisciato" con la tecnica shrinkage proposta da Donoho e Johnstone. Molti degli outliers sono ancora presenti. Il grafico in basso è invece la stima ottenuta con l'algoritmo di smoothing robusto sopra descritto. Gli outliers sono stati adesso rimossi. Vedi [15] per questo esempio.

3.4.3 Spettro Wavelet e Scalogramma

Questa sezione inizia con una breve descrizione dell'utilizzo della trasformata di Fourier nell'ambito dell'analisi spettrale di serie storiche. In seguito, utilizzando la trasformata continua wavelet, saranno definite le funzioni spettro wavelet e scalogramma evidenziandone l'analogia con la funzione di densità spettrale. Tali funzioni descrivono infatti la distribuzione dell'energia di una funzione al variare dei parametri di scala e di traslazione. L'applicazione di questi strumenti nell'ambito dell'analisi di serie storiche è una delle più recenti applicazioni statistiche delle wavelets e promette interessanti risultati.

Sia $X(t)$ una funzione di t che descrive un fenomeno fisico che varia nel tempo. La trasformata di Fourier

$$\mathcal{F}_X(\omega) = \int_{\mathbb{R}} e^{-it\omega} X(t) dt$$

è utilizzata per studiare le caratteristiche della funzione nel dominio frequenziale.

In particolare, la formula di Plancherel

$$\|X\|^2 = \frac{1}{2\pi} \|\mathcal{F}_X\|^2 \qquad X \in L_2(\mathbb{R}),$$

nota in questo contesto come *conservazione dell'energia*, permette di esprimere l'energia totale $E = \|X\|^2$ come

$$E = \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{F}_X(\omega)|^2 d\omega. \tag{3.11}$$

Il termine $|\mathcal{F}_X(\omega)|^2 d\omega$ rappresenta quindi il contributo all'energia totale di quelle componenti di $X(t)$ le cui frequenze sono comprese tra ω e $\omega + d\omega$. Al variare di ω la funzione

$$h(\omega) = |\mathcal{F}_X(\omega)|^2$$

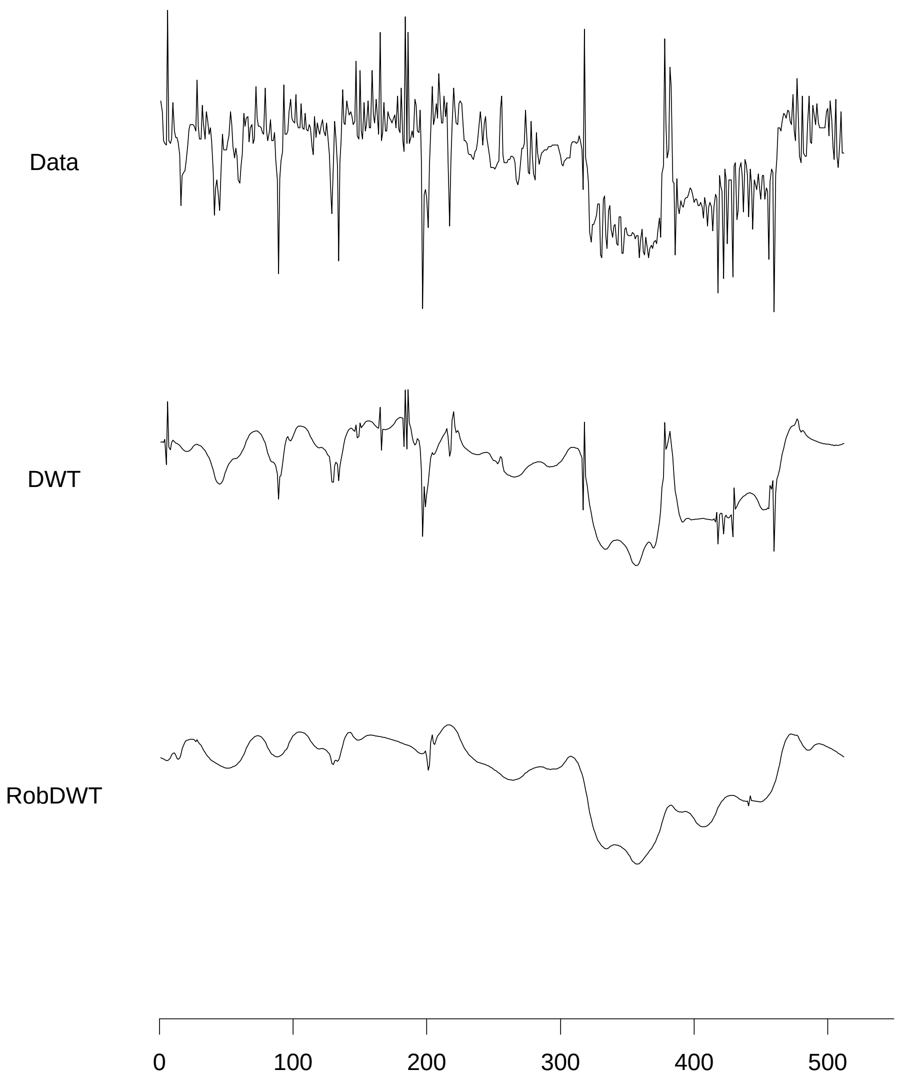

Figura 3.13: Segnale *glint* (Data), stima ottenuta con la tecnica di wavelet shrinkage (DWT) e con l'algoritmo robusto (RobDWT).

misura la densità dell'energia relativa alle componenti con frequenze in un intorno di ω ed è detta *funzione di densità spettrale* o *spettro di potenza* [3].

Per il teorema di Wiener-Khintchine si ha

$$|\mathcal{F}_X(\omega)|^2 = \mathcal{F}_R(\tau)$$

dove $R(\tau)$ è la funzione di autocorrelazione di $X(t)$ definita come

$$R(\tau) = \int \overline{X(t)} X(t+\tau) dt.$$

La funzione di densità spettrale è quindi la trasformata di Fourier della funzione di autocorrelazione

$$h(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} R(\tau) e^{-i\omega\tau} d\tau.$$

Se $X(t)$ è a valori reali, allora $|\mathcal{F}_X(\omega)|^2 = |\mathcal{F}_X(-\omega)|^2$ e

$$h(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} R(\tau) \cos(\omega\tau) d\tau.$$

Questo tipo di analisi è stato largamente utilizzato in statistica per la analisi spettrale di serie storiche stazionarie, [96]. In tale contesto, $X(t)$ rappresenta la realizzazione di un processo stocastico a media zero e la funzione $h(\omega)$ viene comunemente stimata ricorrendo alla definizione di periodogramma

$$I_N(\omega) = \frac{1}{N} |\sum_{t=1}^{N} X_t e^{-i\omega t}|^2$$

dove la funzione $A(\omega) = \sum_{t=1}^{N} X_t e^{-i\omega t}$ è la *trasformata discreta di Fourier* dei valori osservati $X_1, \cdots, X_N$.

Si ha inoltre

$$\sigma_X^2 = R(0) = \int_{\mathbb{R}} h(\omega) d\omega \tag{3.12}$$

cioè la varianza del processo è una misura dell'energia totale [4]. La varianza σ_X^2 può essere stimata con la varianza campionaria; l'espressione (3.12) diventa quindi

$$\frac{1}{N} \sum_{t=1}^{N} X_t^2 = \sum_k I_N(\omega_k) \quad \omega_k = \frac{2\pi}{N} k, \quad k = 0, \cdots, N-1.$$

[3] La potenza è l'energia per unità di tempo.

[4] In statistica la funzione $R(\tau)$ è detta autocovarianza. Il nome autocorrelazione è invece assegnato alla funzione $R(\tau)$ normalizzata $\rho(\tau) = R(\tau)/R(0)$.

In conclusione, la trasformata di Fourier determina una scomposizione della varianza della serie in un insieme di componenti, ognuna delle quali può essere associata ad una particolare frequenza; dal grafico del periodogramma è possibile stabilire quali sono le componenti, e quindi le frequenze, che contribuiscono in maggior misura.

Si consideri adesso la trasformata continua wavelet di $X(t)$

$$W_X(a,b) = \langle X, \psi_{a,b} \rangle = \int_{\mathbb{R}} X(t)\overline{\psi_{a,b}(t)}dt \tag{3.13}$$

dove

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a}).$$

La formula di Plancherel diventa in questo caso

$$E = \frac{1}{C_\psi}\int\int |W_X(a,b)|^2 \frac{dadb}{a^2}$$

dove C_ψ è la costante di ammissibilità definita in (1.5).

L'equivalente wavelet dello spettro di potenza $h(\omega)$ deve essere una funzione il cui integrale è proporzionale all'energia totale E. La funzione

$$h_w(a) = \int |W_X(a,b)|^2 db$$

soddisfa questo requisito ed è quindi detta *spettro wavelet di potenza*, [62] [63] [94]. Lo spettro wavelet descrive la distribuzione dell'energia al variare del parametro di scala a.

Al variare di a ed b la funzione $|W_X(a,b)|^2/C_\psi a^2$ descrive invece la distribuzione dell'energia nel piano scala-tempo. Flandrin, [47], propone per la funzione

$$S_w(a,b) = |W_X(a,b)|^2$$

il nome di *scalogramma* [5], in analogia con lo *spettrogramma*, definito come il modulo al quadrato della trasformata con finestra di Fourier, STFT, di $X(t)$

$$STFT_X(t_0,\omega) = \int_{\mathbb{R}} X(t)g(t-t_0)e^{-i\omega t}dt,$$

che descrive, al variare di t_0 e ω, la distribuzione dell'energia nel piano tempo-frequenza (vedi anche [98]).

[5]In letteratura la terminologia è purtroppo non chiara. Alcuni autori usano il nome di *varianza wavelet* invece di spettro wavelet. Altri addirittura usano la terminologia inversa a quella qui presentata chiamando scalogramma la funzione $h_w(a)$ e spettro wavelet quella $S_w(a,b)$.

Un vantaggio nell'utilizzo dello spettro wavelet è costituito dal fatto che piccoli cambiamenti sull'asse temporale alterano lo spettro wavelet solo localmente a differenza dello spettro di Fourier che risulta invece globalmente alterato. Una tale caratteristica è dovuta alla proprietà di localizzazione nel tempo delle funzioni wavelets che non è invece soddisfatta dalle funzioni trigonometriche.

Nel caso discreto, dati i valori osservati $X_1, \cdots, X_N$, $N = 2^m$, e posti $a = (1/2)^j$ e $b = k(1/2)^j$, $j, k \in \mathbb{Z}$, la trasformata wavelet discreta, DWT, permette di determinare una scomposizione della varianza della serie in un insieme di componenti ognuna delle quali può essere associata a particolari valori dei parametri di scala e di tempo data da

$$\frac{1}{N}\sum_{t=1}^{N} X_t^2 = \|c^{(0)}\|^2 + \sum_{j=0}^{m-1} \|d^{(j)}\|^2$$

dove il vettore $(d^{(m-1)}, d^{(m-2)}, \cdots, d^{(0)}, c^{(0)})$ è quello dei coefficienti ottenuti dalla DWT. Il grafico dello scalogramma $S_w(j,k)$ permette di determinare quali coefficienti contribuiscono in misura maggiore. Lo spettro wavelet (discreto), versione wavelet del periodogramma,

$$h_w(j) = \sum_k |d_{j,k}|^2, \quad j = m-1, \cdots, 0$$

determina invece il contributo alla varianza della serie relativo ai diversi valori del parametro di scala j.

In [48] spettro wavelet e scalogramma sono utilizzati in geofisica da vari autori per lo studio dei moti di turbolenza. In particolare, Collineau e Brunet osservano che lo spettro wavelet presenta dei "picchi" in corrispondenza di valori del parametro di scala proporzionali alla durata media di componenti periodiche del segnale. Tali valori sono utilizzati per separare queste componenti secondo il seguente criterio: si supponga che lo spettro wavelet presenti due picchi e sia a_c il valore del parametro di scala che li separa. I coefficienti wavelet relativi alle scale successive ad a_c vengono posti uguali a zero; la trasformata wavelet inversa applicata all'insieme dei coefficienti così ottenuti determina la ricostruzione della componente corrispondente al primo picco. In maniera analoga si procede per il secondo.

Arino e Vidakovic, [6], applicano questa tecnica nell'ambito dell'analisi di componenti di serie storiche economiche. Arino infine, [7], applica alle singole componenti estratte da h_w tecniche standard di previsione statistica: la somma di tali previsioni determina una previsione sull'intera serie.

Appendice A

Alcuni Spazi Funzionali

Spazi di Hilbert

Indicato con X uno spazio vettoriale, si definisce *prodotto scalare* una applicazione bilineare tale che per ogni $x, y \in X$

$$\langle x, y\rangle = \overline{\langle y, x\rangle}, \quad \langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$$

$$\langle \alpha x, y\rangle = \alpha\langle x, y\rangle, \quad \langle x, x\rangle \geq 0, \quad \langle x, x\rangle = 0 \Rightarrow x = 0.$$

Uno spazio vettoriale in cui sia definito un prodotto scalare è uno spazio normato con *norma*

$$\|x\| = \sqrt{\langle x, x\rangle}, \quad \forall x \in X.$$

Data una successione $\{x_n\} \in X$, questa è detta *di Cauchy* se $\|x_m - x_n\| \to 0$, per $m, n \to \infty$. Uno *spazio di Hilbert* è uno spazio vettoriale munito di prodotto scalare e *completo*, cioè tale che ogni successione di Cauchy è convergente in X. Un semplice esempio di spazio di Hilbert è lo *spazio Euclideo* $E^n = \{x = (x_1, \cdots, x_n)\}$ dove

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i, \quad \|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Due sottospazi A e B di uno spazio vettoriale X costituiscono una scomposizione in *somma diretta* di X, $X = A \oplus B$, se ogni elemento di X può essere unicamente scritto come somma di un elemento di A e un elemento di B, cioè se $\forall x \in X$ esistono $a \in A$ e $b \in B$ tali che $x = a + b$. Dato uno spazio vettoriale X ed un suo sottospazio $U \subset X$, si definisce *complemento ortogonale* di U in X il sottospazio W di X tale che $W \perp U$ e $X = U \oplus W$.

Spazi $L_2(\mathbb{R})$ e $l_p(\mathbb{Z})$

Si indichi con $L_p(\mathbb{R})$, $1 \leq p < \infty$ lo spazio delle funzioni $f : \mathbb{R} \to \mathbb{C}$ misurabili e tali che

$$\int_{\mathbb{R}} |f(x)|^p dx < +\infty.$$

Per $p = +\infty$ lo spazio $L_\infty(\mathbb{R})$ è invece quello delle funzioni misurabili e tali che

$$ess\ sup_{x \in \mathbb{R}} |f(x)| < +\infty.$$

Di particolare interesse sono gli spazi ottenuti ponendo $p = 1$ e $p = 2$. Si definisce quindi *spazio delle funzioni integrabili* l'insieme

$$L_1(\mathbb{R}) = \{f \ : \ \int_{\mathbb{R}} |f(x)| dx < +\infty\}$$

e *spazio delle funzioni di quadrato integrabile* l'insieme

$$L_2(\mathbb{R}) = \{f \ : \ \int_{\mathbb{R}} |f(x)|^2 dx < +\infty\}.$$

$L_2(\mathbb{R})$ è uno spazio di Hilbert con prodotto scalare

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x) \overline{g(x)} dx$$

e norma

$$\|f\|_2^2 = \int_{\mathbb{R}} |f(x)|^2 dx$$

che è detta *energia* di f. Una funzione f appartenente ad $L_2(\mathbb{R})$ ha quindi energia finita.

Se f è una funzione periodica di periodo 2π la cui restrizione a $(-\pi, \pi)$ è in $L_1(-\pi, \pi)$, f è detta soddisfare la *condizione di Lipschitz* di ordine $\alpha > 0$ in x_0 se esiste una costante C tale che

$$|f(x) - f(x_0)| \leq C|x - x_0|^\alpha$$

in un intorno di x_0. Se tale condizione si verifica per ogni x_0, allora f verifica la *condizione di Lipschitz uniforme.*

Si indica poi con $l_p(\mathbb{Z})$ lo spazio delle successioni $\{a_i\}_{i \in \mathbb{Z}}$ a valori complessi tali che

$$\sum\nolimits_{i \in \mathbb{Z}} |a_i|^p < +\infty$$

e si definisce *spazio delle successioni di quadrato sommabile* l'insieme

$$l_2(\mathbb{Z}) = \{\{a_i\}_{i\in\mathbb{Z}} \ : \ \sum_{i\in\mathbb{Z}} |a_i|^2 < +\infty\}.$$

Lo spazio $l_2(\mathbb{Z})$ è uno spazio di Hilbert con prodotto scalare

$$\langle a, b\rangle = \sum_{i\in\mathbb{Z}} a_i \overline{b_i}$$

e norma

$$\|a\|^2 = \sum_{i\in\mathbb{Z}} |a_i|^2.$$

Data una successione di funzioni $\{f_n(x)\}_{n\in\mathbb{N}}$, si definisce *spazio generato* da tale successione, $Span\{f_n\}$, l'insieme di funzioni che possono essere espresse come combinazione lineare delle f_n,

$$Span\{f_n\} = \{f \in L_2(\mathbb{R}) : f(x) = \sum_{n\in\mathbb{N}} a_n f_n(x)\}.$$

Se i coefficienti a_n sono unici per ogni f, allora $\{f_n\}_{n\in\mathbb{N}}$ costituisce una *base* in $Span\{f_n\}$. Una successione di funzioni $\{f_n\}_{n\in\mathbb{N}}$ in $L_2(\mathbb{R})$ è detta *sistema ortogonale* se

$$\langle f_n, f_m\rangle = \int_{\mathbb{R}} f_n(x)\overline{f_m(x)}dx = 0 \qquad n \neq m, \ n, m \in \mathbb{N}$$

e *sistema ortonormale* se si verifica anche

$$\langle f_n, f_n\rangle = 1, \ n \in \mathbb{N}.$$

In particolare, data una base ortogonale in $L_2(\mathbb{R})$, ogni funzione in tale spazio può essere espressa come

$$f(x) = \sum_{n\in\mathbb{N}} \langle f(x), f_n(x)\rangle f_n(x) \tag{A.1}$$

dove i coefficienti $a_n = \langle f(x), f_n(x)\rangle$ rappresentano la proiezione ortogonale di f nei sottospazi V_n generati da $\{f_0, \cdots, f_n\}$.

Dato lo sviluppo in serie $f(x) = \sum_{n\in\mathbb{N}} a_n f_n(x)$ di una funzione $f \in L_2(\mathbb{R})$ ed indicata con f_N la serie troncata ad N, di particolare interesse nella teoria dell'approssimazione di funzioni è la determinazione del *tasso di convergenza* di una serie, ovvero la velocità con la quale l'errore $\|f_N - f\|$ decade a zero. Per

$a \to 0$ (oppure ad ∞), si indica con $O(a)$ una quantità che, divisa per a, è limitata da una costante e con $o(a)$ una quantità che, divisa per a, tende a zero.

Un sistema ortonormale $\{f_n\}_{n\in\mathbb{N}}$ è detto *completo* in $L_2(\mathbb{R})$ se, data $f \in L_2(\mathbb{R})$, $\langle f, f_n\rangle = 0, \quad n = 0, 1, \cdots,$ implica $f = 0$.

Spazi di Sobolev e spazi di Besov

Sia S lo spazio delle funzioni C^∞ in $\mathbb{R}$ che decadono rapidamente, cioè tali che

$$|g^{(k)}(x)| \le C_{k,p}(1+|x|)^{-p}, \quad p, k \in \mathbb{N} \tag{A.2}$$

con $C_{k,p}$ costante. Sia poi S_r lo spazio delle funzioni C^r che soddisfano la condizione (A.2) per $k \le r$ e per ogni p. S_r è detto *spazio di Schwartz* di ordine r. Si definisce *spazio delle distribuzioni temperate* lo spazio duale di S, cioè l'insieme S' dei funzionali $f : S \to \mathbb{C}$ lineari e continui su S. Indicheremo con (f, g) il valore del funzionale f in $g \in S$. Gli elementi di S' sono *funzioni generalizzate.* Ogni funzione localmente integrabile f a crescita esponenziale appartiene ad S' e il funzionale corrispondente è

$$(f, g) = \int_{\mathbb{R}} f(x)g(x)dx.$$

La definizione di trasformata di Fourier può essere estesa allo spazio S' nel seguente modo: data $f \in S'$ si definisce $\mathcal{F}_f$ come

$$(\mathcal{F}_f, g) = (f, \mathcal{F}_g).$$

Si definisce poi *spazio di Sobolev* H^α, per $\alpha \in \mathbb{R}$, il sottospazio di S' dato da

$$H^\alpha = \{f \in S' \ : \ \int_{\mathbb{R}} |\mathcal{F}_f(\omega)|^2(\omega^2+1)^\alpha d\omega < \infty\}.$$

Per $\alpha = 0$ si ha che H^α è lo spazio $L_2(\mathbb{R})$; per $\alpha = 1, 2, \cdots$, invece, H^α è lo spazio delle funzioni in $L_2(\mathbb{R})$ che sono $(\alpha-1)$-volte derivabili e la cui α-esima derivata è in $L_2(\mathbb{R})$; infine, per $\alpha = -1, -2, \cdots$ H^α è lo spazio delle distribuzioni con supporto costituito da un punto e di ordine $\le \alpha$ (per esempio, δ, δ', $\cdots$, $\delta^{(\alpha-1)}$ dove δ è la delta di Dirac). Gli spazi di Sobolev sono spazi di Hilbert con prodotto scalare

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{F}_f(\omega)\overline{\mathcal{F}_g(\omega)}(\omega^2+1)^\alpha d\omega.$$

Più in generale si definisce *spazio di Sobolev* H_p^α, per $\alpha \in \mathbb{R}$ e $1 \le p \le \infty$, il sottospazio di S' dato da

$$H_p^\alpha = \{f \in S' \ : \ \|\mathcal{F}_f(\omega)(\omega^2+1)^{\alpha/2}\|_p < \infty\}.$$

Infine, si definisce *spazio di Besov* B^s_{pq}, per $s \in \mathbb{R}$ e $1 \leq p, q \leq \infty$, il sottospazio di S' dato da

$$B^s_{pq} = \{f \in S' \ : \ \left(\|\psi * f\|_p + (\sum_{k=1}^{\infty} (2^{sk} \|\phi_k * f\|_p)^q))^{1/q} \right) < \infty\}$$

con ψ e ϕ_k appartenenti ad S e definite da

$$\mathcal{F}_{\phi_k}(\omega) = \phi(2^{-k}\omega)$$

$$\mathcal{F}_{\psi}(\omega) = 1 - \sum_{k=1}^{\infty} \phi(2^{-k}\omega)$$

$(k = 0, \pm 1, \pm 2, \cdots)$ dove ϕ è scelta in S tale che

$$Supp(\phi) \subset [-1, 1] \quad \text{e} \quad \int \phi(x) dx = 1.$$

Anche gli spazi di Besov sono spazi di Hilbert.

Per una trattazione più completa di questi spazi si rimanda a [11].

Appendice B

Proprietà di Minimax per Stimatori Wavelet di Densità in Spazi di Besov

Il tasso minimax di convergenza associato ad uno spazio funzionale F, ad un insieme di stimatori V e ad una funzione di perdita $d(f^*, f)$ è una costante a tale che

$$C_1 n^{-a} \leq Inf_{f^* \in V} Sup_{f \in F} E(d(f^*, f)) \leq C_2 n^{-a}.$$

Prima di enunciare le proprietà di minimax degli stimatori wavelet di funzioni di densità descritti nel secondo capitolo è opportuno dimostrare come gli spazi di Besov (vedi Appendice A) possono essere caratterizzati utilizzando le basi wavelet. Si indichi con $P_0 f$ la proiezione di f in V_0 e con $D_j f$ quella in W_j

$$P_0 f = \sum_k c_{0,k} \phi_{0,k}, \quad D_j f = \sum_k d_{j,k} \psi_{j,k}.$$

Proposizione B.1 *Data una MRA r-regolare con $r > s$, allora*

$$f \in B^s_{p,q} \text{ se e solo se } J^s_{p,q}(f) = \|P_0 f\|_p + \Big(\sum_{j \geq 0} (\|D_j f\|_p 2^{js})^q\Big)^{1/q} < +\infty$$

per $s > 0$ e $1 \leq p,\ q \leq +\infty$.

La norma $J^s_{p,q}$ è equivalente alla usuale norma definita per gli spazi di Besov. Utilizzando i coefficienti delle proiezioni di f in V_0 e W_j si può determinare un'altra norma, ancora equivalente,

$$J^{s'}_{p,q}(f) = \|c_{0,\cdot}\|_p + \Big(\sum_{j \geq 0} (\|d_{j,\cdot}\|_p 2^{j(s+1/2-1/p)})^q\Big)^{1/q}$$

dove la norma $\| \cdot \|_p$ è quella definita negli spazi l_p.

Stimatori Wavelet Lineari

Si indichi con $F^s_{p,q}$ lo spazio

$$F^s_{p,q} = \{f \ : \ f \geq 0, \ \int f = 1, \ J^s_{p,q}(f) \leq M\}$$

e si considerino le distanze

$$d(f^*, f) = (J^{s'}_{p,q'}(f^* - f))^q \quad \text{per } 0 < s' < s, \ q' \geq q$$

$$d(f^*, f) = \|f^{*(s')} - f^{(s')}\|^p_p \quad \text{per } 0 \leq s' < s, \ s' \text{ intero.}$$

Si ha la seguente, [70] [71]

Proposizione B.2 *Sia $d(\cdot,\cdot)$ una delle due norme sopra definite e sia V un insieme di stimatori contenente quelli lineari. Data una MRA r-regolare con $r > s$ e $\phi \in C^r$ a supporto compatto (per esempio le wavelets Daubechies), si ha*

(i) Per $p \geq 2$ esiste C_1 : per ogni $f \in F^s_{p,q}$

$$E(d(\hat{f}, f)) \leq C_1 2^{\bar{\jmath}s'p}\Big\{2^{-\bar{\jmath}sp} + \Big(\frac{2^{\bar{\jmath}}}{n}\Big)^{p/2}\Big\} \quad per\ 2^{\bar{\jmath}}/n \to 0, \ n \to +\infty.$$

(ii) Per M sufficientemente grande esiste C_2 :

$$Inf_{f^* \in V} Sup_{f \in F^s_{p,q}} E(d(f^*, f)) \geq C_2 n^{(s'-s)p/(1+2s)}$$

Si osservi che nella disuguaglianza (i) il limite inferiore è ottenuto per $2^{\bar{\jmath}} = n^{1/(1+2s)}$; per tale valore si ha quindi

$$Sup_{f \in F^s_{p,q}} E(d(\hat{f}, f)) \leq C_1 n^{(s'-s)p/(1+2s)}.$$

Da questa disuguaglianza e dalla (ii) segue che lo stimatore wavelet lineare (2.19) raggiunge il tasso minimax di convergenza per $\bar{\jmath}$ tale che $2^{\bar{\jmath}} = n^{1/(1+2s)}$.

Stimatori Wavelet a Soglia

Nel caso dello stimatore a soglia proposto da Donoho, Johnstone, Kerkyacharian e Picard si ha la seguente

Proposizione B.3 *Se f è a supporto compatto e norma limitata in uno spazio di Besov $B_{s,p,q}$ e la misura globale di errore (il rischio) adottata è $E\|\hat{f} - f\|^2_{s,\pi}$ con $\|\cdot\|_{s,\pi}$ la norma nello spazio di Sobolev H^s_π, lo stimatore a soglia raggiunge (a meno di un fattore logaritmico) il tasso minimax ottimo, risultando migliore dello stimatore lineare (2.19) nel caso $\pi > p$.*

Bibliografia

[1] Special issue on Wavelet Transform and Multiresolution Signal Analysis. *IEEE Transactions on Information Theory*, **38(2)**, March 1992.

[2] Special issue on Wavelets and Signal Processing. *IEEE Transactions on Signal Processing*, **41(12)**, Dec. 1993.

[3] Antoniadis, A. e Carmona, R. (1991). Multiresolution Analyses and Wavelets for Density Estimation. *Technical Report*, University of California at Irvine.

[4] Antoniadis, A., Gregoire, G. e McKeague, I.W. (1994). Wavelet Methods for Curve Estimation. *JASA*, **89(428)**, Dec..

[5] Antoniadis, A. e Oppenheim, G. editori (1995). *Wavelets and Statistics.* Proceedings of the Villard de Lans Conference November 1994. Lecture Notes in Statistics. Springer Verlag.

[6] Arino, M.A. e Vidakovic, B. (1995). On Wavelet Scalograms and their Applications in Economic Time Series. *Discussion Paper*, **95-21**, ISDS, Duke University.

[7] Arino, M.A. (1995). Time Series Forecast via Wavelets: an Application to Car Sales in the Spanish Market. *Discussion Paper*, **95-30**, ISDS, Duke University.

[8] Aronszajn, N. (1950). Theory of Reproducing Kernels. *Math. Soc.*, **68**, 337-404.

[9] Azzalini, A. e Bowman, A.W. (1990). A Look at Some Data on the Old Faithful Geyser. *Applied Statistics*, **39**, 357-365.

[10] Benedetto, J.J. e Frazier, M.W., editori, (1994). *Wavelets: Mathematics and Applications.* CRC Press.

[11] BERGH, J. e LÖFSTRÖM, J. (1976). *Interpolation Spaces, an Introduction.* Springer Verlag.

[12] BEYLKIN, G. (1992). On the Representation of Operators in Bases of Compactly Supported Wavelets. *SIAM J. Numer. Anal.*, **29(6)**, 1716-1740.

[13] BEYLKIN, G., COIFMAN, R. e ROKHLIN, V. (1991). Fast Wavelet Transforms and Numerical Algorithms I. *Comunications on Pure and Applied Mathematics*, **XLIV**, 141-183.

[14] BOCHNER, S. e CHANDRASHEKHARAN. K. (1949). *Fourier transforms.* Princeton University, Princeton.

[15] BRUCE, A.G. e GAO, H.Y. (1994). *S+Wavelets, user manual, version 1.0.* Statsci Division, Mathsoft Inc., Seattle, Washington.

[16] BRUCE, A.G., DONOHO, D.L., GAO, H.Y. e MARTIN, R.D. (1994). Denoising and Robust Non-Linear Wavelet Analysis. In *Proceedings SPIE, Wavelet Applications.* SZU, H.H., editore, **2242**, 5-8 April, Orlando, Florida.

[17] BURRUS, C.S. e GOPINATH, R.A. (1995). Introduction to Wavelets and Wavelet Transforms. *Lecture Notes*, Rice University, Houston, Texas.

[18] BYRNES, J.S., BYRNES, J.L., HARGREAVES, K.A. e BERRY, K. editori (1994). *Wavelets and their Applications : Proceedings of the NATO Advanced Study Institute, Il Ciocco, Italy, 16-29 August 1992* Kluwer Academic Publishers.

[19] ČENCOV, N. N. (1962). Evaluation of an Unknown Distribution Density from Observations. *Doklady*, **3**, 1559-1562.

[20] CHAN, Y.T. (1995). *Wavelet Basics.* Kluver Academic Publihsers, Boston.

[21] CHUI, C.K. (1992). *An Introduction to Wavelets.* Academic Press, New York.

[22] CHUI, C.K., editore, (1992). *Wavelets: A Tutorial in Theory and Applications.* Academic Press, New York.

[23] CHUI, C.K., MONTEFUSCO, L. e PUCCIO, L., editori, (1994). *Wavelets: Theory, Algorithms and Applications.* Academic Press.

[24] COHEN, L. (1989). Time-Frequency Distribution-A Review. *Proc. IEEE* **77(7)***, 941-981.*

[25] COHEN, A. (1992). Biorthogonal Wavelets. In *Wavelets: A Tutorial in Theory and Applications*. CHUI, C.K., editore. Academic Press, New York.

[26] COHEN, A., DAUBECHIES, I. e VIAL, P. (1993). Wavelets on the Interval and Fast Wavelet Transforms. *ACHA* **1**, 54-81.

[27] DAUBECHIES, I. (1988). Orthonormal Bases of Compactly Supported Wavelets. *Comunications on Pure and Applied Mathematics*, **XLI**, 909-996.

[28] DAUBECHIES, I. (1990). The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, **36(5)**, 961-1005.

[29] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets.* SIAM, CBMS-NSF Conference Series **61**.

[30] DAUBECHIES, I., editore, (1993). *Different Perspectives on Wavelets.* Proceedings of Symposia in Applied Mathematics, **47**. American Mathematical Society, Short Course, January 11-12, San Antonio, Texas.

[31] DAUBECHIES, I. (1993). Orthonormal Bases of Compactly Supported Wavelets II. Variations on the theme. *SIAM, J. Math. Anal.*, **24**, 499-519.

[32] DAUBECHIES, I., GROSSMANN, A. e MEYER, Y. (1986). Painless Nonorthogonal Expansions. *J. Math. Phys.* **27**, 1271-1283.

[33] DAUBECHIES, I. e LAGARIAS, J. (1991). Two-scale Difference Equations I. Existence and Global Regularity of Solutions. *SIAM J. Math. Anal.*, **22(5)**, 1388-1410.

[34] DAUBECHIES, I. e LAGARIAS, J. (1992). Two-scale Difference Equations II. Local Regularity, Infinite Products of Matrices and Fractals. *SIAM J. Math. Anal.*, **23(4)**, 1031-1079.

[35] DELYON, B. e JUDITSKY, A. (1993). Wavelet Estimators, Global Error Measures: Revisited. *Publication interne n.782*, IRISA-INRIA.

[36] DEVROYE, L. e GYORFI, L. (1985). *Nonparametric Density Estimation, the L_1 view.* John Wiley & sons.

[37] DONOHO, D.L. (1992). De-Noising via Soft Thresholding. *Technical Report 409*, Statistical Department, Stanford University.

[38] DONOHO, D.L. (1993). Unconditional Bases are Optimal Bases for Data Compression and for Statistical Estimation. *Applied and Computational Harmonic Analysis*, **1**, 100-115.

[39] DONOHO, D.L. (1993). Nonlinear Wavelet Methods for Recovery of Signals, Densities and Spectra from Indirect and Noisy Data. In *Different Perspectives on Wavelets*, I. Daubechies editore.

[40] DONOHO, D.L. e JOHNSTONE, I.M. (1990). Minimax Risk over l_p-balls. *Technical Report*, Statistics, University of California, Berkeley.

[41] DONOHO, D.L. e JOHNSTONE, I.M. (1992). Minimax Estimation via Wavelet Shrinkage. *Technical Report 402*, Statistics, Stanford University.

[42] DONOHO, D.L. e JOHNSTONE, I.M. (1994). Ideal Spatial Adaption via Wavelet Shrinkage. *Biometrika*, **81**, 425-455.

[43] DONOHO, D.L. e JOHNSTONE, I.M. (1995). Adapting to Unknown Smoothness by Wavelet Shrinkage. *Journal of the American Statistical Association*, **90**, 1200-1224.

[44] DONOHO, D.L., JOHNSTONE, I.M., KERKYACHARIAN, G. e PICARD, D. (1993). Density estimation by wavelet thresholding. *Technical Report 426*, Statistics, Stanford University.

[45] DONOHO, D.L., JOHNSTONE, I.M., KERKYACHARIAN, G. e PICARD, D. (1995). Wavelet shrinkage: Asymptopia?. *J. R. Statis. Soc. B*, **57 (2)**, 301-369.

[46] EUBANK, R.L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker inc.

[47] FLANDRIN, P. (1988). Time-Frequency and Time-Scale. *IEEE Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling.* Minneapolis, Minnesota, 77-80.

[48] FOUFOULA-GEORGIOU, E. e KUMAR, P., editori, (1994). *Wavelets in Geophisics.* Academic Press.

[49] FOX, C. (1987). *An Introduction to Calculus of Variations.* Dover

[50] GABOR, D. (1946). Theory of Communication. *J. Inst. Elec. Eng.* **93**, 429-441.

[51] GAO, H. (1993). Choice of Thresholding for Wavelet Shrinkage Estimate of the Spectrum. *Technical Report*, Statistics, Stanford University.

[52] GAO, H. (1993). Wavelet Estimation of Spectral Densities in Time Series Analysis. *PhD Thesis*, University of California, Berkeley.

[53] GOEL, P. e VIDAKOVIC, B. (1995). Wavelet Transformations as Diversity Enhancers. *Discussion Paper*, **95-04**, ISDS, Duke University.

[54] GOOD, I.J. e GASKIN, R.A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, **58(2)**, 255-77.

[55] GROSSMANN, A. e MORLET, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.* **15**, 723-736.

[56] HAAR, A. (1910). Zur Theorie der Orthogonalen Funktionen-Systeme. *Math. Ann.* **69**, 331-371.

[57] HALL, P. e PATIL, P. (1995). On Wavelet Methods for Estimating Smooth Functions. *Bernoulli*, **1(1/2)**, 41-58.

[58] HÄRDLE, W. (1991). *Smoothing Techniques with Implementation in S.* New York, Springer Verlag.

[59] HÄRDLE, W. (1991). *Applied Nonparametric Regression.* Cambridge University Press.

[60] HEIL, C.E. e WALNUT, D.F. (1989). Continous and Discrete Wavelet Transforms. *SIAM Review*, **31(4)**, 628-666.

[61] HUCK, F.O. e JUDAY, R.D., editori, (1993). *Visual Information Processing II.* Proceedings SPIE, 14-16 April , Orlando, Florida.

[62] HUDGINS, L.H., MAYER, M.E. e FRIEHE, C.A. (1993). Fourier and Wavelet Analysis of Atmospheric Turbulence. In *Progress in Wavelet Analysis and Applications*, Y. Meyer e S. Roques editori.

[63] HUDGINS, L.H. (1993). Wavelet Transform and Spectral Estimation. *Phisical Review Letters*, Novembre.

[64] IZENMAN, A. (1991). Recent Developments in Nonparametric Density Estimation. *JASA*, **86(413)**, 205-224.

[65] JAWERTH, B. e SWELDENS, W. (1994). An Overview of Wavelet Based Multiresolution Analyses. *SIAM Review*, **36(3)**, 377-412.

[66] JANSSEN (1988). The Zak Transform: a signal transform for sampled time-continuous signals. *Phillips J. Res.*, **43**, 23-69.

[67] JOHNSTONE, I.M. e SILVERMAN, B.W. (1994). Wavelet Threshold Estimators for Data with Correlated Noise. *Technical report*. Dept. of Mathematics. University of Bristol.

[68] KAISER, G. (1994). *A Friendly Guide to Wavelets*. Birkhauser.

[69] KAY, J. (1994). Wavelets. *Advances in Applied Statistics*, capitolo 10.

[70] KERKYACHARIAN, G. e PICARD, D. (1992). Density Estimation in Besov Spaces. *Statistics & Probability Letters*, **13**, 15-24.

[71] KERKYACHARIAN, G. e PICARD, D. (1993). Density Estimation by Kernel and Wavelets Methods: Optimality of Besov Spaces. *Statistics & Probability Letters*, **18**, 327-336.

[72] KOORNWINDER, T.H., editore, (1993). *Wavelets: an Elementary Treatment of Theory and Applications*. Number 1 in Series in Approximations and Decompositions. World Scientific, Singapore.

[73] KRONMAL, R. e TARTER, M. (1968). The Estimation of Probability Densities and Cumulatives by Fourier Series Methods. *JASA*, **63**, 925-952.

[74] LAINE, A.F. e UNSER, M.A., editori, (1995). *Wavelet Applications in Signal and Image Processing III*. Proceedings SPIE, 12-14 July , San Diego, California.

[75] MALLAT, S.G. (1989). Multiresolution Approximations and Wavelet Orthonormal Bases of $L_2(\mathbb{R})$. *Transactions of the American Mathematical Society* **315 (1)**, 69-87.

[76] MALLAT, S.G. (1989). A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11 (7)**, 674-693.

[77] MEYER, Y., editore, (1992). *Wavelets and Applications: Proceedings of the International Conference on Wavelets, May 1989, Marseille, France*. Springer Verlag.

[78] MEYER, Y. (1992). *Wavelets and Operators.* Cambridge University Press, translated by D.H. Salinger.

[79] MEYER, Y. (1993). *Wavelets: Algorithms and Applications.* SIAM, Philadelphia, translated by R.D. Ryan.

[80] MEYER, Y. (1993). Review of Books *An Introduction to Wavelets* by C.K. Chui and *Ten Lectures on Wavelets* by I. Daubechies. *Bulletin (New Series) of the American Mathematical Society*, **28(2)**, 350-360.

[81] MEYER, Y. e ROQUES, S.(1993). *Progress in Wavelet Analysis and Applications.* Editions Frontieres, France.

[82] MIZRACH, B. (1993). Multivariate Nearest-Neighbour Forecasts of EMS Exchange Rates. In *Nonlinear Dynamics, Chaos and Econometrics*, Pesaran and Potter editori. J.Wiley & Sons Ltd.

[83] MORLET, J., ARENS, G., FOURGEAU, I. e GIARD, D. (1982). Wave Propagation and Sampling Theory, part II. *Geophysics*, **47**, 203-236.

[84] MOTARD, R.L. e JOSEPH, B., editori, (1994). *Wavelet Applications in Chemical Engineering.* Kluver Academic Publishers.

[85] MOULIN, P. (1994). Wavelet Thresholding Techniques for Power Spectrum Estimation. *IEEE Transactions on Signal Processing*, November.

[86] MOULIN, P. (1993). A Wavelet Regularization Method for Diffuse Radar-Target Imaging and Speckle-Noise Reduction. *Journal of Mathematical Imaging and Vision*, Special Issue on Wavelets, **3(1)**, 123-134.

[87] NADARAYA, E. (1964). On Regression Estimators. *Theory Prob. Appl.*, **9**,157-159.

[88] NASON, G.P. (1994). Wavelet Regression by Cross-Validation. *Technical report 447*, Statistics, Stanford University.

[89] NASON, G.P. e SILVERMAN, B.W. (1994). The Discrete Wavelet Transform in S. *Journal of Computational and Graphical Statistics*, **3**, 163-191.

[90] O'HAGAN, A. (1994). *Kendall's Advanced Theory of Statistics, vol. 2b: Bayesian Inference* University Press Cambridge.

[91] PACINI, B. (1994). Metodi Non Parametrici e Semiparametrici per l'Analisi di Modelli Economico-Finanziari. *Tesi di Dottorato*, Dipartimento Statistico G.Parenti, Università di Firenze.

[92] Pacini, B. e Vannucci, M.(1996). Stima non parametrica di funzioni di densità e di regressione con particolare riferimento ad applicazioni in ambito econometrico. Dipartimento di Statistica "G. Parenti", Università di Firenze.

[93] Penev, S. e Dechevsky, L. (1995). On Non-Negative Wavelet-Based Density Estimators. *Report* **S95-11**. School of Mathematics. Dept. of Statistics. University of New South Wales. Sidney. Australia.

[94] Percival, D.B. (1993). An Introduction to Spectral Analysis and Wavelets. *Proceedings of the Workshop Advanced Mathemetical Tools in Metrology* Torino, Italy, Ottobre.

[95] Pinheiro, A. e Vidakovic, B. (1995). Estimating the Square Root of a Density via Compactly Supported Wavelets. *Discussion Paper 95-14*, ISDS, Duke University.

[96] Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Vol.1,2. Academic Press.

[97] Rinehart, J.S. (1969). Thermal and Seismic Indications of Old Faithful Geyser's Inner Working. *J. Geophys. res.*, **74**, 566-573.

[98] Rioul, O. e Vetterli, M. (1991). Wavelet and Signal Processing. *IEEE Signal Processing Magazine*, October, 14-38.

[99] Rudin, W. (1970). *Real and Complex Analysis.* McGraw-Hill Series in Higher Mathematicsf, New York.

[100] Ruskay, M.B, Beylkin, G., Coifman, R., Daubechies, I.. Mallat, S., Meyer, Y. e Raphael, L., editori, (1992). *Wavelets and their Applications.* Jones and Bartlett Publishers.

[101] Samarov, A. (1994). Nonparametric Curve and Functional Estimation. Lecture Notes. Università di Padova.

[102] Sansone, G. (1955). *Orthogonal Functions.* Interscience, New York.

[103] Schumaker, L.L. e Webb, G. (1994). *Recent Advances in Wavelet Analysis.* Academic Press.

[104] Scott, D.W. (1992). *Multivariate Density Estimation.* Wiley, New York.

[105] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

[106] STEIN, C. (1981). Estimation of the Mean of a Multivariate Normal Distribution. *Annals of Statistics*, **9(6)**, 1135-1151.

[107] STONE, C.J. (1977). Consistent Nonparametric Regression. *Annals of Statistics*, **5**, 595-645.

[108] STRANG, G. (1989). Wavelets and Dilation Equations: a Brief Introduction. *SIAM Review*, **31(4)**, 614-627.

[109] STRANG, G. (1993). Wavelet Transforms versus Fourier Transforms. *Bulletin (New Series) of the American Mathematical Society*, **28(2)**.

[110] SZU, H.H., editore, (1994). Wavelet Applications. *Proceedings SPIE*, **2242**, 5-8 April, Orlando, Florida,

[111] TAPIA, R.A. e THOMPSON, J.R. (1978). *Nonparametric Probability Density Estimation.* The John Hopkins University Press, Baltimore and London.

[112] TASWELL, C. (1995). WavBox 4: A Software Toolbox for Wavelet Transforms and Adaptive Wavelet Packet Decompositions. In Proceedings of the Villard de Lans Conference November 1994, *Wavelets and Statistics*, Lecture Notes in Statistics, A. Antoniadis and G. Oppenheim editori. Springer Verlag.

[113] TASWELL, C. (1995). Algorithms for Wavelet Transforms and Adaptive Wavelet Packet Decompositions. *PhD Thesis*, Scientific Computing and Computational Mathematics, Stanford University.

[114] TASWELL, C. e MCGILL, K.C. (1993). Wavelet Transform Algorithms for Finite-Duration Discrete-Time Signals. *Technical Report*, Stanford University.

[115] THOMPSON, J.R. e TAPIA, R.A. (1990). *Nonparametric Function Estimation, Modeling and Simulation.* Siam, Philadelphia.

[116] TRIBOULEY, K. (1995). Practical Estimation of Multivariate Densities using Wavelt Methods. *Statistica Neerlandica*, **49(1)**, 41-62.

[117] VANNUCCI, M. (1995). Nonparametric Density Estimation using Wavelets. *Discussion Paper 95-26*, ISDS, Duke University.

[118] VANNUCCI, M. e VIDAKOVIC, B. (1995). Preventing the Dirac Disaster: Wavelet Based Density Estimation. *Discussion Paper 95-27*, ISDS, Duke University.

[119] Venables, W.N. e Ripley, B.D. (1992). *Modern Applied Statistics with S-Plus*. New York, Springer Verlag.

[120] Vetterli, M. e Herley, C. (1992). Wavelets and Filter Banks: Theory and Design. *IEEE Transactions on Signal Processing*, **40**, 2207-2232.

[121] Vidakovic, B. (1994). Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factor. *Discussion Paper 94-24*, ISDS, Duke University.

[122] Vidakovic, B. e Müller, P. (1994). Wavelets for kids. *Discussion Paper 94-13*, ISDS, Duke University.

[123] Vidakovic, B. e Müller, P. (1995). Affine Bayes Rules in Wavelet Shrinkage and Applications. Manoscritto.

[124] Wahba, G. (1980). Automatic Smoothing of the Log Periodogram. *JASA* **75**, 122-132.

[125] Walter, G.G. (1992). Approximation of Delta Function by Wavelets. *Journal of Approximation Theory*, **71**, 329-343.

[126] Walter, G.G. (1994). *Wavelets and Others Orthogonal Systems with Applications*. CRC Press, Boca Raton, FL.

[127] Walter, G.G. (1995). Estimation with Wavelets and the Curse of Dimensionality. *Manoscritto*, Department of Mathematical Sciences, University of Wisconsin-Milwaukee.

[128] Walter, G.G. e Blum, J. (1979). Probability Density Estimation using Delta Sequences. *The Annals of Statistics*, **7(2)**, 328-340.

[129] Wu, D. (1994). Probability Density Estimation with Wavelets. *PhD Thesis*, University of Wisconsin-Milwaukee.

[130] Watson, G.S. (1964). Smooth Regression Analysis. *Sankya*, serie A, **26**, 359-372.

[131] Young, R.K. (1993). *Wavelet Theory and its Applications*. Kluwer Academic Publishers.

[132] Zayed, A. (1993). *Advances in Sampling Theory and Applications*. CRC Press, Boca Raton.

[133] Zygmund, A. (1957). *Trigonometric Series*. Cambridge University Press, New York.