



UNIVERSITÀ DI PISA

Corso di Laurea Triennale in Matematica

Tesi di Laurea

METODI DI STIMA
DELL'INDICE DI HURST
DEL MOTO BROWNIANO FRAZIONARIO

Relatore:
Prof. Dario Trevisan

Candidato:
Leonardo Cito

ANNO ACCADEMICO 2021/2022

Indice

| | | |
|----------|---|-----------|
| 1 | Storia del moto Browniano frazionario | 1 |
| 1.1 | Hurst | 1 |
| 1.2 | Definizione dell'indice di Hurst | 2 |
| 1.3 | I risultati di Hurst | 2 |
| 2 | Moto Browniano frazionario | 5 |
| 2.1 | Definizioni e Notazioni | 5 |
| 2.2 | Proprietà del moto Browniano frazionario | 8 |
| 2.3 | Densità spettrale | 11 |
| 3 | Metodi classici di stima dell'indice di Hurst | 15 |
| 3.1 | Stimatori | 15 |
| 3.2 | Metodo delle varianze aggregate | 18 |
| 3.3 | Metodo dei momenti assoluti | 22 |
| 3.4 | Metodo delle variazioni discrete | 23 |
| 3.5 | Metodo del periodogramma | 25 |
| 3.6 | Metodo R/S | 29 |
| 3.7 | Sperimentazione numerica | 35 |
| 3.7.1 | Varianze aggregate | 35 |
| 3.7.2 | Periodogramma | 36 |
| 3.7.3 | R/S | 38 |
| 4 | Reti neurali | 43 |
| 4.1 | Struttura di una rete neurale | 43 |
| 4.1.1 | Algoritmo di discesa stocastica del gradiente | 46 |
| 4.2 | Stima dell'indice di Hurst | 46 |
| 4.2.1 | Modello | 47 |
| 4.2.2 | Sperimentazione numerica | 47 |
| 5 | Conclusioni | 53 |
| A | Codici Python | 55 |
| A.1 | Metodo delle varianze aggregate | 55 |

| | | |
|----------|--|-----------|
| A.2 | Metodo del periodogramma | 56 |
| A.3 | Metodo R/S | 57 |
| A.4 | Rete neurale convuluzionale 1-dimensionale | 59 |
| A.5 | Rete neurale convuluzionale bidimensionale | 60 |
| B | Codici Matlab | 61 |
| B.1 | Algoritmo di Davies e Harte | 61 |
| | Bibliografia | 63 |

Introduzione

Nel 1900 il francese Louis Bachelier (1870-1946) scrisse a Parigi, nell'ambito dei suoi studi di dottorato in economia, una dissertazione dal titolo *Théorie de la Spéculation*. Egli fu il primo a capire che era possibile applicare la teoria della probabilità allo studio dei mercati finanziari, così, cinque anni prima di Einstein e di Wiener, descrisse tutte le caratteristiche essenziali del moto Browniano introducendo il concetto di processo stocastico in tempo continuo. Nella sua tesi il matematico francese trattò la fluttuazione dei prezzi delle obbligazioni francesi rifacendosi ad un concetto prettamente fisico affrontato nel secolo precedente da un botanico scozzese di nome Brown, da cui ha origine il nome moto Browniano.

Nel 1827 il botanico scozzese Robert Brown (1773-1858) osservò al microscopio come le particelle di polline in sospensione nell'acqua si muovessero continuamente in modo casuale e imprevedibile.

Nel 1905 Einstein (1879-1955) fornì un'interpretazione molecolare tramite un modello probabilistico del fenomeno del movimento osservato da Brown, mostrando come il moto delle particelle fosse descrivibile matematicamente ipotizzando che i salti di queste fossero dovuti agli scontri casuali delle particelle di polline con le molecole dell'acqua.

Nei decenni successivi, in ambito finanziario, venne ampiamente discusso da matematici ed economisti se, effettivamente, il moto Browniano fosse il modello corretto per la rappresentazione dell'andamento dei prezzi dei titoli finanziari, che secondo la teoria classica risultava essere casuale. Secondo i risultati conseguiti dal matematico franco-polacco Benoit Mandelbrot (1924-2010), la presunta relazione di normalità dei rendimenti era da eliminare visto che eventi come la crisi del 1929, il crollo della borsa di Wall Street del 1987 (e successivamente la grande recessione del 2007) si sarebbero dovuti verificare ogni migliaia di anni, contrariamente a quanto accaduto nella realtà. Partendo da questo presupposto, Mandelbrot affermò che i movimenti dei prezzi non fossero indipendenti o Browniani, bensì subissero l'influenza degli eventi passati, in grado di alterare le quotazioni future dei titoli. Doveva, quindi, essere in azione un processo la cui evoluzione era dipendente in qualche misura dagli eventi passati. Questi fenomeni seguivano un andamento nel tempo che può essere descritto come un processo stocastico noto come moto Browniano frazionario e in genere implicano la presenza di una dipendenza di lungo periodo nelle osservazioni.

Il moto Browniano frazionario è oggi uno strumento matematico nell'ambito della teoria della probabilità, utilizzato per descrivere un insieme sempre più vasto di fenomeni, studiati

da discipline anche molto diverse tra loro, ad esempio la diffusione del calore, delle popolazioni di animali, dei batteri, di una malattia, del suono o della luce sono tutte modellate con lo stesso strumento: moto Browniano frazionario. L'utilizzo di tale moto si giustifica nei casi in cui una variazione è dipendente dalle variazioni precedenti e quindi il semplice moto Browniano, non è in grado di descriverla.

La tipologia di questa dipendenza è sintetizzata in quello che oggi è noto come indice di Hurst del moto Browniano frazionario, che rappresenta il tema principale di questa tesi. Dopo aver spiegato l'origine e la storia dell'indice di Hurst (**Capitolo 1**) e introdotto le definizioni di processo stocastico e moto browniano frazionario e alcuni risultati teorici ad essi relativi (**Capitolo 2**), nel **Capitolo 3** illustremo e sperimenteremo alcuni dei metodi classici di stima dell'indice di Hurst. Infine, nel **Capitolo 4**, prendendo ispirazione dall'articolo *Long-Range Dependent Traffic Classification with Convolutional Neural Networks Based on Hurst Exponent Analysis*, pubblicato il 15 ottobre 2020 sulla rivista accademica "Entropy", proporremo un metodo di stima dell'indice di Hurst che sfrutta la struttura delle reti neurali.

Capitolo 1

Storia del moto Browniano frazionario

1.1 Hurst

Un apporto significativo allo studio del moto Browniano frazionario fu quello fornito dall'idrologo Harold Edwin Hurst (1880-1978).

Tra la fine del 1800 e l'inizio del 1900 le fabbriche di cotone del nord dell'Inghilterra, in piena espansione industriale, reclamavano crescenti quantità di cotone egiziano, i cui livelli di produzione erano strettamente legati al ciclo delle piene del Nilo. Il problema del Nilo era che non si conosceva di quanto straripasse il fiume in relazione all'anno precedente, in pratica era ignota la sua volatilità. Nel 1906 l'Impero Britannico, di cui l'Egitto faceva parte, incaricò Hurst di realizzare un progetto chiamato "Century Storage", che aveva lo scopo di accantonare una quantità di acqua necessaria per far fronte alla massima siccità possibile.

Lavorando ai progetti di dighe sul Nilo fino all'inizio degli anni '50 ed occupandosi del controllo delle riserve idriche, concluse che una riserva ideale non dovrebbe mai straripare ed una politica di rilascio idrico dovrebbe essere formulata in modo da soddisfare i bisogni idrici a valle senza svuotare completamente la riserva. Nel tentativo di definire un modello matematico in grado di simulare il problema reale, Hurst ipotizzò che la parte incontrollabile del sistema (l'afflusso di acqua nel bacino) si comportasse come una passeggiata aleatoria, per cui l'evento del passato non influisce su quello futuro: essi sono tra loro perfettamente indipendenti. Se l'ipotesi di cammino casuale risultasse verificata, il range di fluttuazione dovrebbe crescere in proporzione alla radice quadrata dell'intervallo temporale di misurazione T .

Le analisi empiriche svolte da Hurst sulle piene del Nilo contraddissero, però, l'ipotesi di casualità e ciò lo spinse a svolgere studi analoghi su ulteriori 690 serie storiche di circa 75 diversi fenomeni geofisici, come piogge, temperature, pressioni atmosferiche, anelli dei tronchi dei pini e delle sequoie. Osservò che gran parte di essi non seguivano un vero e proprio cammino casuale, ma un cammino casuale persistente. Queste considerazioni, riportate nelle esigenze pratiche di Hurst, sostanzialmente implicarono che, dopo un incremento del livello della riserva, fosse più probabile registrare un ulteriore incremento piuttosto che un suo decremento. Ad un certo punto questa tendenza termina, cioè avviene una brusca

inversione del trend e diventa più probabile osservare un decremento del livello seguito da un ulteriore decremento.

1.2 Definizione dell'indice di Hurst

Siano $(x_i)_{i=1}^n$ i valori dell'acqua defluita dalla riserva all' i -esimo anno. Definiamo $X_k = \sum_{i=1}^k x_i$ e consideriamo la deviazione del valore di X_k dalla rispettiva media calcolata su osservazioni raccolte in n anni:

$$Y_k = X_k - \frac{k}{n} X_n.$$

Il range della deviazione (detto anche *range aggiustato*) è:

$$R_n := \max_{1 \leq k \leq n} (Y_k) - \min_{1 \leq k \leq n} (Y_k).$$

Tuttavia, Hurst non operò con i valori R_n , bensì con i valori normalizzati $Q_n := R_n/S_n$ (*range aggiustato normalizzato*), dove

$$S_n = \sqrt{\frac{1}{n} \sum_{k=1}^n x_k^2 - \left(\frac{1}{n} \sum_{k=1}^n x_k\right)^2}$$

rappresenta la deviazione standard, e osservò che, almeno asintoticamente, valesse la seguente relazione:

$$\frac{R_n}{S_n} = \left(\frac{n}{2}\right)^H,$$

dove C è una costante positiva e il parametro H è quello che oggi è noto come *indice di Hurst*. Passando al logaritmo:

$$\log(R_n/S_n) = \log c + H \log(n/2)$$

da cui si ottiene una stima di H , di cui discuteremo successivamente nella sezione dedicata al metodo R/S, della forma:

$$H \approx \frac{\log(R_n/S_n)}{\log(n/2)}.$$

1.3 I risultati di Hurst

Se l'ipotesi iniziale di cammino casuale fosse vera, il valore dell'esponente H risulterebbe 0,5; tuttavia, dalle analisi effettuate sui fenomeni naturali, Hurst ottenne sempre valori di H assai superiori: ad esempio, per le variazioni logaritmiche del livello idrico del Nilo, Hurst calcola empiricamente un valore di H pari a circa 0,9; mentre, in media, rispetto ai fenomeni naturali da lui studiati, il valore di H risulta essere $0,72 \pm 0,006$ con un intervallo di oscillazione tra 0,46 e 0,96 [6]. Il sistema risulta, quindi, caratterizzato da un effetto memoria per il quale ogni osservazione influenzata da quelle passate influenzerà quelle future.

Hurst riuscì, inoltre, ad anticipare ciò che successivamente venne rigorosamente dimostrato, ossia che al variare di H si osservano tre diverse situazioni:

- per $H = 0,5$ si denotano una serie di eventi indipendenti: ogni variazione non è influenzata dalle precedenti e neppure influenzerà quelle future;
- per $0 < H < 0,5$ la serie è antipersistente, ossia, se in un dato periodo il sistema ha subito un incremento (decremento), è più probabile registrare un successivo decremento (incremento) che un ulteriore incremento (decremento). La serie risulta più volatile di una serie casuale (poiché caratterizzata da inversioni più frequenti) tanto più il valore di H si avvicina a zero. Tipici esempi di sistemi con tali valori di H sono dotati di turbolenza;
- per $0,5 < H < 1$ si ha una serie persistente, ossia caratterizzata da una dipendenza positiva tra le variazioni generate dal processo: se nell'ultima osservazione abbiamo registrato un incremento (decremento) è più probabile che l'osservazione successiva registri un ulteriore incremento (decremento). La probabilità di registrare due variazioni di segno concorde risulta tanto più alta quando H si avvicina ad uno. Fenomeni di questo tipo sono riscontrati nel comportamento dei livelli dei fiumi, o dell'attività solare, o nell'analisi dei rendimenti per titoli azionari.

Capitolo 2

Moto Browniano frazionario

Il moto Browniano classico non offre una corretta descrizione dei fenomeni in cui si verificano casi di dipendenza come quelli studiati da Hurst. Questi eventi seguono un andamento che può essere descritto come un particolare processo stocastico: il **moto Browniano frazionario**.

2.1 Definizioni e Notazioni

Nelle definizioni di questa sezione indichiamo con Ω un generico spazio di probabilità $(\Omega, \mathfrak{S}, \mathbb{P})$ e con E un generico spazio misurabile (E, ξ) .

Vettori Gaussiani

Definizione 2.1. Sia $\mathbf{X} = (X_1, \dots, X_d)$ un vettore aleatorio, $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$, definiamo la matrice di covarianza Q associata ad \mathbf{X} come: $(Q)_{i,j} = Cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]$ per $i, j = 1, \dots, d$.

Proposizione 2.2. Q è una matrice simmetrica e semidefinita positiva.

Dimostrazione. La simmetria segue per definizione di Q . Verifichiamo che sia semidefinita positiva, ovvero che $\forall v \in \mathbb{R}^d \langle v, Qv \rangle \geq 0$.

$$\begin{aligned} \langle v, Qv \rangle &= \sum_{i,j=1}^d v_i (Q)_{ij} v_j = \sum_{i,j=1}^d v_i Cov(X_i, X_j) v_j = \sum_{i,j=1}^d Cov(v_i X_i, v_j X_j) = \\ &= \sum_{i=1}^d \sum_{j=1}^d Cov(v_i X_i, \langle v, X \rangle) = Cov(\langle v, X \rangle, \langle v, X \rangle) = Var(\langle v, X \rangle) \geq 0. \end{aligned}$$

□

Un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ è detto *gaussiano* se $\forall t \in \mathbb{R}^d \langle t, X \rangle = \sum_{i=1}^d t_i X_i$ è una variabile aleatoria gaussiana. Per i vettori gaussiani valgono i seguenti risultati:

Proposizione 2.3. Se \mathbf{X} è un vettore gaussiano, allora \mathbf{X} è di quadrato integrabile, ovvero sono ben definite la media $m = \mathbb{E}[\mathbf{X}]$ e la matrice di covarianza $Q = Q_{\mathbf{X}}$ di \mathbf{X} . Inoltre, m e Q identificano le legge di \mathbf{X} , cioè $\mathbf{X} \sim N(m, Q)$.

Dimostrazione. Sia $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$ vettore gaussiano, e sia $e_i \in \mathbb{R}^d$ l' i -esimo vettore della base canonica di \mathbb{R}^d per $i = 1, \dots, d$. Segue che $X_i = \langle e_i, \mathbf{X} \rangle$ è una variabile gaussiana reale $\forall i = 1, \dots, d$, quindi ammette momento secondo finito, cioè $\mathbb{E}[X_i^2] < \infty \forall i = 1, \dots, d$, dunque $\mathbb{E}[\sum_{i=1}^d X_i^2] < \infty$, dunque il vettore delle medie m e la matrice Q sono ben definiti.

Concludiamo mostrando che la funzione caratteristica $\Phi_{\mathbf{X}}$ di \mathbf{X} dipende da m e Q .

Verifichiamo che valga: $\langle t, \mathbf{X} \rangle \sim N(\langle t, m \rangle, \langle t, Qt \rangle)$.

Vale che

$$\langle t, m \rangle = \langle t, \mathbb{E}[\mathbf{X}] \rangle = \sum_{i=1}^d t_i \mathbb{E}[X_i] = \mathbb{E}[\langle t, \mathbf{X} \rangle]$$

e, procedendo analogamente a quanto fatto nella dimostrazione della Proposizione 2.2, si ottiene che $\text{Var}(\langle t, \mathbf{X} \rangle) = \langle t, Qt \rangle$. Quindi $\forall t \in \mathbb{R}^d$:

$$\Phi_{\mathbf{X}}(t) = \Phi_{\langle t, \mathbf{X} \rangle}(1) = \exp\left(i\langle t, m \rangle - \frac{1}{2}\langle t, Qt \rangle\right).$$

□

Da cui segue che se $\mathbf{X} = (X_1, X_2)$ è un vettore gaussiano con $\text{Cov}(X_1, X_2) = 0$, allora X_1 e X_2 sono indipendenti.

Proposizione 2.4. *Sia \mathbf{X} vettore gaussiano con matrice di covarianza Q invertibile (ovvero definita positiva), allora $\mathbf{X} \sim N(m, Q)$ con densità:*

$$f_{\mathbf{X}}(x) = \frac{1}{\sqrt{(2\pi)^d \det Q}} \exp\left(-\frac{1}{2}\langle x-m, Q^{-1}(x-m) \rangle\right).$$

Inoltre, è immediato verificare che se $X : \Omega \rightarrow \mathbb{R}^d$ vettore gaussiano, $A \in \mathbb{R}^{k \times d}$ e $b \in \mathbb{R}^k$, allora $AX + b$ è un vettore gaussiano con media $\mathbb{E}[AX + b] = A\mathbb{E}[X] + b$ e matrice di covarianza $Q_{AX+b} = AQA^T$, dove Q è la matrice di covarianza del vettore X .

Processi stocastici

Definizione 2.5. Un *processo stocastico* è una famiglia di variabili aleatorie $(X_t)_{t \in T}$, $T \subseteq \mathbb{R}$, con $X_t : \Omega \rightarrow E$, ossia sono tutte definite sullo stesso spazio di probabilità $(\Omega, \mathfrak{F}, \mathbb{P})$ e a valori in uno spazio misurabile (E, ξ) . T è detto insieme dei tempi, E insieme degli stati.

Fissato $\omega \in \Omega$, si chiama *traiettoria del processo \mathbf{X}* (o realizzazione del processo \mathbf{X}) la funzione: $T \rightarrow E$, con $\omega \in \Omega$ fissato, che ad ogni istante t associa il valore $X_t(\omega)$.

Osservazione 2.6. Dato un processo \mathbf{X} , possiamo rappresentarlo come una famiglia di variabili aleatorie, come suggerito dalla definizione iniziale, oppure, come una variabile aleatoria a valori nelle traiettorie, ovvero $\mathbf{X} : \Omega \rightarrow E^T$. Quest'ultimo caso permette di considerare lo spazio E^T munito della σ -algebra prodotto $\otimes_{t \in T} \xi = \xi^{\otimes T}$ generata dalle proiezioni canoniche e, quindi, di definire la *legge congiunta di un processo* come la misura immagine di \mathbb{P} tramite la mappa $\mathbf{X} : \Omega \rightarrow E^T$ che associa ad $\omega \mapsto (X_t(\omega))_{t \in T}$. Tale legge è una misura di probabilità su $(E^T, \xi^{\otimes T})$.

Dati due processi \mathbf{X}, \mathbf{Y} , diciamo che:

- \mathbf{Y} è una *modificazione di \mathbf{X}* se per ogni istante $t \in T$, $X_t = Y_t$ \mathbb{P} -q.c., cioè $\mathbb{P}(X_t \neq Y_t) = 0 \forall t \in T$;
- sono *indistinguibili* se $\exists N \in \mathfrak{S}$ tale che $\mathbb{P}(N) = 0$ e $\bigcup_{t \in T} \{X_t \neq Y_t\} \subseteq N$, ossia quando \mathbf{X} e \mathbf{Y} coincidono considerati come variabili aleatorie a valori nelle traiettorie.

Osserviamo che se \mathbf{X}, \mathbf{Y} sono una modificazione dell'altro, allora hanno la stessa legge e che se sono indistinguibili, allora sono una modificazione dell'altro, ma, in generale, non vale il viceversa.

Un processo stocastico $(X_t)_{t \in [0, +\infty]}$ a valori in \mathbb{R} si dice:

- ad *incrementi indipendenti* se $\forall 0 \leq t_0 < t_1 < \dots < t_n$ le variabili aleatorie degli incrementi, cioè $(X_{t_1} - X_{t_0}), \dots, (X_{t_n} - X_{t_{n-1}})$, sono indipendenti.
- ad *incrementi stazionari* se la legge dell'incremento $(X_{t_1} - X_{t_0})$ dipende solo da $t_1 - t_0 \forall t_0, t_1$, cioè $(X_{t_1} - X_{t_0}) \sim (X_{t_1 - t_0} - X_0)$, dove \sim indica l'equivalenza in legge tra le due variabili aleatorie.

Un *processo stocastico* è detto *gaussiano* se, dato un qualsiasi numero finito di variabili aleatorie dalla collezione che costituisce il processo stesso, esse hanno una distribuzione di probabilità congiunta gaussiana.

Esempio 2.7. Il processo di Wiener, noto come *moto Browniano*, è un processo stocastico gaussiano univocamente determinato dalle seguenti proprietà:

1. $B_0 = 0$.
2. $(B_t)_{t \geq 0}$ è un processo ad incrementi indipendenti.
3. $\forall s, t \geq 0, s \leq t, B_t - B_s \sim N(0, t - s)$.
4. $\forall \omega \in \Omega$, le traiettorie $t \mapsto B_t(\omega)$ sono continue.

Definizione 2.8. Un processo stocastico (a tempi discreti) $\mathbf{X} = \{X_k : k = 0, 1, 2, \dots\}$ è *stazionario* se i vettori aleatori $(X_{k_1}, \dots, X_{k_d})$ e $(X_{k_1+n}, \dots, X_{k_d+n})$ hanno la stessa distribuzione congiunta per ogni intero $d, n \geq 1$ e $k_1, \dots, k_d > 0$.

Dalla combinazione delle definizioni di processo stocastico gaussiano e di processo stocastico stazionario segue che per i processi Gaussiani è equivalente richiedere che $\gamma(k) := \text{Cov}(X_n, X_{n+k})$ non dipenda da n e che sia costante $\mathbb{E}[X_n] = m$ per ogni n .

Notazione. Da qui in poi γ indicherà la funzione di autocovarianza definita sopra.

Definizione 2.9. Un processo stocastico (discreto) \mathbf{X} è *auto-simile con indice di Hurst* $0 < H < 1$ se \mathbf{X} e $m^{1-H}\mathbf{X}^{(m)}$ hanno la stessa distribuzione congiunta finita per ogni $m \geq 1$, cioè se per ogni $d \geq 1$ e $0 \leq k_1 < \dots < k_d$ i vettori

$$(X_{k_1}, \dots, X_{k_d}) \quad \text{e} \quad (m^{1-H}X_{k_1}^{(m)}, \dots, m^{1-H}X_{k_d}^{(m)})$$

hanno la stessa distribuzione congiunta.

$\mathbf{X}^{(m)}$ è un processo definito come segue: per ogni $m \geq 1$, $\mathbf{X}^{(m)} = \{X_k^{(m)} : k = 0, 1, 2, \dots\}$ dove:

$$X_k^{(m)} = \frac{1}{m} (X_{km} + \dots + X_{(k+1)m-1}).$$

Le definizioni appena date possono essere formulate analagamente per il caso continuo, con la differenza che per l'auto-similarità di un processo stocastico Y continuo si confrontano i processi $\{Y(at) : 0 \leq t < \infty\}$ e $\{a^H Y(t) : 0 \leq t < \infty\}$ per ogni $a > 0$, invece di Y e $m^{1-H} Y^{(m)}$ del caso discreto. Ovviamente, i processi che simuleremo computazionalmente saranno discreti.

Definizione 2.10. Un processo stocastico stazionario (discreto) \mathbf{X} è con dipendenza a lungo raggio (LRD, o a lunga memoria) se $\sum_{k=0}^{\infty} \gamma(k) = \infty$ dove γ è la funzione di autocovarianza. Se, invece, la serie è convergente, il processo viene detto a memoria breve.

Sembrerebbe, quindi, teoricamente impossibile stabilire da un campione finito di un processo se quest'ultimo sia o meno con dipendenza a lungo termine. Ricorreremo, perciò, ad opportuni metodi di stima che illustreremo nei capitoli 3-4.

Moto Browniano frazionario (fBm)

Tratteremo il caso particolare del moto Browniano frazionario normalizzato:

Definizione 2.11. Un processo gaussiano $B_H := \{B_H(t) : 0 \leq t < \infty\}$ con $0 < H < 1$ che rispetta le seguenti proprietà è detto *moto Browniano frazionario normalizzato* :

- B_H ha incrementi stazionari.
- $B_H(0) = 0$ e $\mathbb{E}[B_H(t)] = 0$ per $t \geq 0$.
- $\mathbb{E}[B_H^2(t)] = t^{2H}$ per $t \geq 0$.

Dato B_H moto Browniano frazionario, definiamo il corrispondente processo degli incrementi $\mathbf{X} = \{X_k : k = 0, 1, 2, \dots\}$, noto come *rumore Guassiano frazionario (fGn)*, dove:

$$X_k = B_H(k+1) - B_H(k).$$

È immediato osservare che il rumore Guassiano frazionario è un processo stazionario tale che $X_k \sim N(0, 1) \forall k$ e che, in generale, non c'è indipendenza tra le X_k .

2.2 Proprietà del moto Browniano frazionario

Dalle prime tre proprietà nella definizione di moto Browniano frazionario normalizzato e dall'identità:

$$B_H(t)B_H(s) = \frac{1}{2}(B_H(t)^2 + B_H(s)^2 - (B_H(t) - B_H(s))^2)$$

segue che la funzione di covarianza per il processo B_H è della forma:

$$\rho(t, s) = \mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}) \quad (2.2.1)$$

per $s, t \geq 0$.

Mostriamo, adesso, l'**esistenza** del moto Browniano frazionario, ossia che data una funzione del tipo

$$Q = Q(s, t) = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H})$$

esista un processo con le caratteristiche assegnate al moto Browniano frazionario e di cui Q sia la funzione di covarianza. È sufficiente verificarlo per un campione del processo, cioè che esistano Y_{t_1}, \dots, Y_{t_n} con $0 \leq t_1 < \dots < t_n$ variabili aleatorie tali che il vettore $(Y_{t_1}, \dots, Y_{t_n})$ abbia matrice di covarianza del tipo $(Q(t_i, t_j))_{i,j=1, \dots, n}$.

Dimostrazione. Dobbiamo dimostrare che la matrice $(Q(s, t))_{s,t=1, \dots, n}$ è semidefinita positiva. È sufficiente rappresentare $\forall s, t \geq 0$, $Q(s, t) = \langle \phi_s, \phi_t \rangle_{L^2}$ per $\phi_t, \phi_s \in L^2(\mathbb{R}, \mathcal{L})$, dove \mathcal{L} rappresenta la misura di Lebesgue e $\langle \cdot, \cdot \rangle_{L^2}$ il prodotto scalare su L^2 .

La rappresentazione di Mandelbrot-Van Nees del fBm:

$$B_H(t) = c_H \int_{-\infty}^{\infty} ((t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}}) dB(s),$$

dove c_H è una costante positiva e B un moto Browniano standard, suggerisce di considerare:

$$\phi_t(x) := (t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}} \quad \text{per } x \in \mathbb{R}.$$

Osserviamo che $\phi_0(x) = 0$, quindi basta verificare che $\|\phi_t - \phi_s\|^2 = c_H |t - s|^{2H}$, con $c_H > 0$ per $s \leq t$, per ottenere che $\|\phi_t\|^2 = c_H t^{2H}$,

$$\langle \phi_t, \phi_s \rangle_{L^2} = \frac{1}{2} [\|\phi_t\|^2 + \|\phi_s\|^2 - \|\phi_t - \phi_s\|^2] = \frac{c_H}{2} (t^{2H} + s^{2H} - |t - s|^{2H}) = c_H C(t, s)$$

e dunque la tesi.

Vale che:

$$\begin{aligned} \|\phi_t - \phi_s\|^2 &= \int_{-\infty}^{\infty} |\phi_t(x) - \phi_s(x)|^2 dx = \int_{-\infty}^{\infty} |(t-x)_+^{H-\frac{1}{2}} - (s-x)_+^{H-\frac{1}{2}}|^2 dx = \\ &= \int_{-\infty}^{\infty} |(t-s-y)_+^{H-\frac{1}{2}} - (-y)_+^{H-\frac{1}{2}}|^2 dy = \\ &= \int_{-\infty}^{\infty} |(t-s)^{H-\frac{1}{2}} (1 - \frac{y}{t-s})_+^{H-\frac{1}{2}} - (t-s)^{H-\frac{1}{2}} (-\frac{y}{t-s})_+^{H-\frac{1}{2}}|^2 dy = \\ &= |t-s|^{2H-1} \int_{-\infty}^{\infty} |(1 - \frac{y}{t-s})_+^{H-\frac{1}{2}} - (-\frac{y}{t-s})_+^{H-\frac{1}{2}}|^2 dy = \\ &= |t-s|^{2H} \int_{-\infty}^{\infty} |(1-z)_+^{H-\frac{1}{2}} - (-z)_+^{H-\frac{1}{2}}|^2 dz = \\ &= |t-s|^{2H} \int_{-\infty}^{\infty} |(1+z)_+^{H-\frac{1}{2}} - z_+^{H-\frac{1}{2}}|^2 dz. \end{aligned}$$

Per concludere, basta verificare che

$$c_H := \int_{-\infty}^{\infty} |(1+z)_+^{H-\frac{1}{2}} - z_+^{H-\frac{1}{2}}|^2 dz < \infty \quad \text{per } H \in (0,1),$$

dato che ovviamente $c_H > 0$. Osserviamo che c_H può essere riscritto come:

$$c_H = \int_{-1}^0 (1+z)^{2H-1} dz + \int_0^{\infty} |(1+z)^{H-\frac{1}{2}} - z^{H-\frac{1}{2}}|^2 dz,$$

quindi per il criterio serie-integrali è sufficiente mostrare che

$$\begin{aligned} \sum_{k=0}^{\infty} |(1+k)^{H-\frac{1}{2}} - k^{H-\frac{1}{2}}|^2 &= 1 + \sum_{k=1}^{\infty} |(1+k)^{H-\frac{1}{2}} - k^{H-\frac{1}{2}}|^2 = \\ &= 1 + \sum_{k=1}^{\infty} k^{2H-1} \left| \left(1 + \frac{1}{k}\right)^{H-\frac{1}{2}} - 1 \right|^2 < \infty. \end{aligned}$$

Dallo sviluppo di Taylor della funzione $f(x) = (1+x)^{H-\frac{1}{2}} - 1$ per $x = 0$:
 $f(x) = 0 + (H - \frac{1}{2})x + o(x)$, segue che c_H si comporta come

$$\sum_{k=1}^{\infty} k^{2H-1} \left(\left(H - \frac{1}{2}\right)^2 \frac{1}{k^2} \right) = \sum_{k=1}^{\infty} k^{2H-3} \left(H - \frac{1}{2}\right)^2 = \left(H - \frac{1}{2}\right)^2 \sum_{k=1}^{\infty} \frac{1}{k^{3-2H}} < \infty$$

per ogni $H \in (0,1)$. □

Osservazione 2.12. Il moto Browniano frazionario è un processo stocastico *autosimile* con parametro di Hurst H , infatti, essendo un processo gaussiano, per la Proposizione 2.3 e per l'equazione (2.2.1) possiamo concludere che $\{B_H(at) : 0 \leq t < \infty\}$ e $\{a^H B_H(t) : 0 \leq t < \infty\}$ hanno le stesse distribuzioni congiunte finite $\forall a > 0$.

Da (1.2.1) segue anche che la funzione di autocovarianza $\gamma(\cdot)$ per il rumore gaussiano frazionario è della forma:

$$\gamma(k) = \frac{1}{2} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}). \quad (2.2.2)$$

Osserviamo che per $H = \frac{1}{2}$ si ottengono tutte covarianze nulle (eccetto per $k = 0$), ovvero che gli incrementi di $B_{\frac{1}{2}}$ sono indipendenti. Vale, infatti, che per $H = 1/2$ il moto Browniano frazionario coincide con il moto Browniano standard e il rumore gaussiano frazionario risulta un processo normale identicamente indipendente, cioè un processo puramente casuale, noto come *rumore bianco*.

Proposizione 2.13. *X* rumore gaussiano frazionario è un processo autosimile e con dipendenza a lungo raggio per $\frac{1}{2} < H < 1$.

Dimostrazione. Dallo sviluppo di Taylor della funzione $h(x) = (1-x)^{2H} - 2 + (1+x)^{2H}$ per $x = 0$ e osservando che $\gamma(k) = \frac{1}{2} k^{2H} h(\frac{1}{k})$ per $k \geq 1$, dalla (2.2.2) segue che, per $H \neq \frac{1}{2}$:

$$\gamma(k) \approx H(2H-1)k^{2H-2} \quad \text{per } k \rightarrow \infty. \quad (2.2.3)$$

Ciò implica che per $\frac{1}{2} < H < 1$: $\sum_{k=0}^{\infty} \gamma(k) = \infty$.

Per l'autosimilarità mostriamo che $m\mathbf{X}^{(m)}$ e $m\mathbf{X}^H$ hanno le stesse distribuzioni congiunte finite, quindi è sufficiente verificare che ogni campione finito di variabili abbia la stessa media e covarianza.

Vale che:

$$\begin{aligned} & Cov(X_{km} + \dots + X_{(k+1)m-1}, X_{lm} + \dots + X_{(l+1)m-1}) = \\ & = Cov((B_H((k+1)m) - B_H(km)), B_H((l+1)m) - B_H(lm)) = \\ & =^1 Cov(m^H(B_H(k+1) - B_H(k)), m^H(B_H(l+1) - B_H(l))) = \\ & = Cov(m^H X_k, m^H X_l), \end{aligned}$$

dove $=^1$ segue dal fatto che i vettori $(B_H(mk), B_H(m(k+1)), B_H(ml), B_H(m(l+1)))$ e $(m^H B_H(k), m^H B_H(k+1), m^H B_H(l), m^H B_H(l+1))$ hanno la stessa distribuzione per $k \neq l$ per autosimilarità del moto Browniano frazionario. Dal fatto che $\forall k \geq 0 X_k \sim N(0, 1)$ segue che anche le medie coincidono. \square

Osservazione 2.14. In generale, l'autosimilarità e la dipendenza a lungo raggio non sono equivalenti. Ad esempio, il moto Browniano standard ($H=1/2$) non è chiaramente a dipendenza a lungo raggio (gli incrementi sono indipendenti); mentre il processo ottenuto sommando due rumori gaussiani frazionari X e Y indipendenti con indici di Hurst rispettivamente $H_1 \neq H_2$, per $H_1, H_2 \in (3/4, 1)$ è ancora un processo stazionario e a dipendenza a lungo raggio, ma in generale non autosimile. Infatti, siano γ_X e γ_Y le rispettive funzioni di autocovarianza, $n \in \mathbb{N}$ fissato, per indipendenza dei due processi vale che:

$$\gamma_{X+Y}(k) = \mathbb{E}[X_k Y_k X_{k+n} Y_{k+n}] = \mathbb{E}[X_k X_{k+n}] \mathbb{E}[Y_k Y_{k+n}] = \gamma_X(k) \gamma_Y(k).$$

Dalla (2.2.3) segue che

$$\gamma_X(k) \gamma_Y(k) \approx H_1 H_2 (2H_1 - 1)(2H_2 - 1) k^{2H_2 - 2} k^{2H_1 - 2}$$

per $k \rightarrow \infty$, quindi $\sum_{k=0}^{\infty} \gamma_{X+Y}(k)$ si comporta come

$$\sum_{k=0}^{\infty} k^{2H_1 - 2} k^{2H_2 - 2} \geq \sum_{k=0}^{\infty} \frac{1}{k} = \infty$$

poichè $2H_i - 2 \in (-1/2, 0)$ per $i = 1, 2$.

2.3 Densità spettrale

Abbiamo mostrato due strumenti di analisi dei processi stocastici: valor medio e funzione di autocovarianza, entrambi caratterizzazioni *nel tempo*. È possibile caratterizzare un processo *nelle frequenze*, ovvero studiarlo a partire dal dominio delle frequenze, piuttosto che dal dominio del tempo. Il passaggio da un tipo di dominio all'altro è garantito, senza perdita di

informazioni, dalla Trasformata di Fourier (e la sua inversa).

Fourier provò che una funzione periodica (deterministica) f può essere rappresentata in modo unico come combinazione lineare di funzioni trigonometriche con frequenze differenti tra loro, ovvero f può essere descritta completamente dall'insieme di frequenze di tali funzioni. Una realizzazione di un processo stocastico, tuttavia, non è necessariamente periodica, quindi non è possibile dare una caratterizzazione frequenziale attraverso la trasformata di Fourier (in modo diretto), ma è possibile estendere il concetto anche in questo caso, caratterizzando i processi in termini di spettro di potenza.

Per i processi stocastici stazionari le informazioni sulla "frequenza attesa" sono contenute nella funzione di autocovarianza γ . Si definisce, infatti, la *densità spettrale di un processo stocastico stazionario* come la trasformata di Fourier a tempi discreti (DTFT) della funzione γ :

$$f(\lambda) = \sum_{t=-\infty}^{\infty} \gamma(t) \exp(it\lambda) \quad (2.3.1)$$

per $-\pi \leq \lambda \leq \pi$. La funzione di autocovarianza si può, dunque, ricavare con la formula inversa:

$$\gamma(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) \exp(-it\lambda) d\lambda.$$

Il processo stazionario di nostro interesse è il rumore gaussiano frazionario, la cui densità spettrale del risulta essere [1]:

$$f(\lambda) = 2 \sin(\pi H) \Gamma(2H + 1) (1 - \cos \lambda) (|\lambda|^{-2H-1} + B(\lambda, H))$$

dove

$$B(\lambda, H) = \sum_{t=1}^{\infty} \{(2\pi t + \lambda)^{-2H-1} + (2\pi t - \lambda)^{-2H-1}\}$$

per $-\pi \leq \lambda \leq \pi$.

Non è nota una formula chiusa per la densità spettrale del rumore Gaussiano frazionario, tuttavia, dallo sviluppo di Taylor: $1 - \cos(\lambda) = \frac{1}{2}\lambda^2 + O(\lambda^4)$ per $|\lambda| \rightarrow 0$ e da (2.3.1), segue che la densità spettrale è proporzionale a $|\lambda|^{1-2H}$ vicino a $\lambda = 0$. Osserviamo, inoltre, che per $H > 1/2$ la densità f ha un polo in zero, mentre per $H < 1/2$ non è differenziabile in zero.

Dimostriamo un risultato sulla densità spettrale, noto come *Teorema di Wiener-Khintchine*, che utilizzeremo nel capitolo 3. Il seguente teorema è spesso enunciato senza ipotesi precise. Una delle ragioni è che si può dimostrare a diversi livelli di generalità, con diversi significati dell'operazione di limite (si tratta di un limite di funzioni). Daremo ora un enunciato rigoroso sotto ipotesi molto precise sulla funzione di autocorrelazione.

Teorema 2.15. Sia $((X_n)_{n \in \mathbb{Z}})$ un processo stazionario e sia $R(\cdot)$ la relativa funzione di autocovarianza. Se

$$\sum_{k \in \mathbb{Z}} |R(k)| < \infty \quad (2.3.2)$$

allora:

$$\lim_{N \rightarrow \infty} \frac{1}{2N+1} \mathbb{E}[|\hat{X}_{2N}(\lambda)|^2] = f(\lambda) \quad (2.3.3)$$

dove f è la densità spettrale e

$$\hat{X}_{2N}(\lambda) = \sum_{k=-N}^N X_k \exp(ik\lambda).$$

Il limite è uniforme per $-\pi \leq \lambda \leq \pi$.

Dimostrazione.

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{2N+1} \mathbb{E}[|\hat{X}_{2N}(\lambda)|^2] &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \mathbb{E}\left[\left(\sum_{k=-N}^N X_k \exp(ik\lambda)\right)\left(\sum_{t=-N}^N X_t \exp(-it\lambda)\right)\right] = \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \mathbb{E}\left[\sum_{k=-N}^N \sum_{t=-N}^N X_k X_t \exp(i(k-t)\lambda)\right] = \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N \sum_{t=-N}^N \mathbb{E}[X_k X_t] \exp(i(k-t)\lambda) = \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N \sum_{t=-N}^N R(k-t) \exp(i(k-t)\lambda) = \\ &= \stackrel{1}{=} \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N \sum_{t=-M}^M R(k-t) \exp(i(k-t)\lambda) = \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N \lim_{M \rightarrow \infty} \sum_{t=-M}^M R(k-t) \exp(i(k-t)\lambda) = \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N \left(\sum_{t=-\infty}^{\infty} R(t) \exp(it\lambda)\right) = \\ &= \left(\sum_{t=-\infty}^{\infty} R(t) \exp(it\lambda)\right) \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N 1 = \\ &= \sum_{t=-\infty}^{\infty} R(t) \exp(it\lambda) = f(\lambda). \end{aligned}$$

dove $\stackrel{1}{=}$ segue per convergenza dominata. Infatti,

$$\left| \sum_{t=-N}^N R(k-t) \exp(i(k-t)\lambda) \right| \leq \sum_{t=-N}^N |R(k-t) \exp(i(k-t)\lambda)| = \sum_{t=-N}^N |R(k-t)|$$

e

$$\sum_{k \in \mathbb{Z}} \sum_{t=-N}^N |R(k-t)| \leq 2 \sum_{s \in \mathbb{Z}} |R(s)| < \infty.$$

□

Osserviamo che, per $H \leq 1/2$, possiamo applicare il Teorema di Wiener-Khintchine al rumore gaussiano frazionario.

Capitolo 3

Metodi classici di stima dell'indice di Hurst

In questo capitolo analizziamo alcuni metodi classici per stimare l'indice di Hurst H . Generalmente, stimare H quando è maggiore di $1/2$ risulta piuttosto complicato, dal momento che sperimentalmente possiamo studiare soltanto campioni di taglia finita. Tuttavia, alcune caratteristiche di dipendenza a lungo raggio (LRD) e autosimilarità possono emergere anche da campioni finiti.

Nel Capitolo 2 abbiamo denotato con $(B_H(0), \dots, B_H(N-1))$ un campione del moto Browniano frazionario su $\{0, \dots, N-1\}$. Adesso useremo la stessa notazione per il campione $(B_H(0), \dots, B_H(N-1))$ su $\{0, 1/N, \dots, N-1/N\}$ e, analogamente, per il corrispondente campione del rumore Gaussiano frazionario; dunque d'ora in poi le variabili $B_H(t)$ e X_k avranno media nulla e varianza rispettivamente uguale a $N^{-2H}t^{2H}$ e N^{-2H} .

Prima di illustrare i metodi di stima, richiamiamo alcuni concetti di statistica.

3.1 Stimatori

Gli ambienti in cui sono definiti gli oggetti statistici che useremo per stimare l'indice di Hurst sono i *modelli statistici (parametrici)*, ossia una terna $(\Omega, \mathfrak{S}, \{\mathbb{P}^\theta, \theta \in \Theta\})$ dove:

- Ω è lo spazio campione;
- \mathfrak{S} è una σ -algebra di Ω ;
- $\{\mathbb{P}^\theta, \theta \in \Theta\}$ è una famiglia di leggi di probabilità su (Ω, \mathfrak{S}) dipendente dal parametro $\theta \in \Theta$.

Un modello statistico si dice *dominato* se esiste una misura σ -finita μ su (Ω, \mathfrak{S}) tale che, per ogni $\theta \in \Theta$, risulti $\mathbb{P}^\theta \ll \mu$, cioè \mathbb{P}^θ è assolutamente continua rispetto a μ .

Dato un modello statistico $(\Omega, \mathfrak{S}, \{\mathbb{P}^\theta, \theta \in \Theta\})$ dominato da una misura μ , chiamiamo

verosimiglianza del modello ogni funzione $L : \Theta \times \Omega \rightarrow \mathbb{R}$ tale che, per ogni θ fissato, la funzione $L^\theta := L(\theta, \cdot) : \omega \mapsto L(\theta, \omega)$ sia una versione della densità di \mathbb{P}^θ rispetto a μ , che possiamo indicare con la seguente scrittura: $L^\theta = \frac{d\mathbb{P}^\theta}{d\mu}$.

Definizione 3.1. *Il modello del campione.*

- Sia μ una misura di probabilità su \mathbb{R} , chiamiamo *campione di taglia n e legge μ* una famiglia (X_1, \dots, X_n) di variabili aleatorie indipendenti, aventi tutte legge μ .
- Assegnata una famiglia $\{\mu^\theta, \theta \in \Theta\}$ di leggi di probabilità su \mathbb{R} , chiamiamo *campione di taglia n e legge μ^θ* una famiglia (X_1, \dots, X_n) di variabili aleatorie indipendenti, aventi tutte legge μ^θ . Con leggero abuso di linguaggio indicheremo allo stesso modo il modello statistico $(\Omega, \mathfrak{S}, \{\mathbb{P}^\theta, \theta \in \Theta\})$ dove:

$$\Omega = \mathbb{R}^n, \quad \mathfrak{S} = \mathbb{B}(\mathbb{R}^n), \quad \mathbb{P}^\theta = \mu^\theta \otimes \mu^\theta \otimes \dots \otimes \mu^\theta = (\mu^\theta)^{\otimes n}.$$

Lo scopo della statistica sarebbe individuare la legge \mathbb{P}^θ su tutta la σ -algebra \mathfrak{S} conoscendo solo i valori che essa assume su una sotto σ -algebra ξ . Definiamo quindi il concetto di *statistica e riassunto esaustivo* (o *statistica esaustiva*).

Dato un modello statistico $(\Omega, \mathfrak{S}, \{\mathbb{P}^\theta, \theta \in \Theta\})$ si chiama *statistica* ogni applicazione misurabile $T : (\Omega, \mathfrak{S}) \rightarrow (E, \xi)$ che non dipenda da θ , dove (E, ξ) è uno spazio misurabile qualsiasi. Se il modello statistico è dominato da una misura μ , $T : (\Omega, \mathfrak{S}) \rightarrow (E, \xi)$ è detta *statistica esaustiva* se, per ogni v.a. Y definita su (Ω, \mathfrak{S}) , a valori reali e limitata, esiste, definita μ -q.o., una versione della speranza condizionale $\mathbb{E}^\theta[Y|T]$ che non dipenda da θ . Ricordiamo che:

Definizione 3.2. Fissati uno spazio di probabilità $(\Omega, \mathfrak{S}, \mathbb{P})$, $\xi \subseteq \mathfrak{S}$ sotto σ -algebra e Y v.a. reale con $\mathbb{E}[Y] < \infty$, è detta *speranza condizionale di Y dato ξ* ogni v.a. Z tale che:

- Z sia ξ -misurabile;
- $\forall A \in \xi, \int_A Z d\mathbb{P} = \int_A Y d\mathbb{P}$.

Tale classe di funzioni viene denotata: $Z = \mathbb{E}[Y|\xi]$.

Esempio 3.3. Dato un campione di taglia n e legge $N(m, \sigma^2)$ con entrambi m, σ^2 sconosciuti, si verifica che il vettore

$$\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

è un riassunto esaustivo del parametro $\theta = (m, \sigma^2)$.

Gli strumenti statistici che utilizzeremo nei metodi illustrati di seguito sono gli stimatori. Dato un modello statistico $(\Omega, \mathfrak{S}, \{\mathbb{P}^\theta, \theta \in \Theta\})$ e una funzione $g : \Theta \rightarrow D$ (con D aperto di \mathbb{R}), si chiama *stimatore* di $g(\theta)$ una v.a. $U : \Omega \rightarrow D$ che non dipende da θ . Assegnato $\omega \in \Omega$, $U(\omega)$ si chiama *stima* di $g(\theta)$. Il costo medio rispetto a \mathbb{P}^θ dello stimatore U è detto *rischio*; considereremo come rischio, il rischio quadratico, ossia la variabile $R_U(\theta) = \mathbb{E}[|g(\theta) - U|^2]$,

tramite cui viene definito un ordine (parziale) tra gli stimatori. Infatti, uno stimatore U si dice *preferibile* rispetto ad uno stimatore V se si ha

$$R_U(\theta) \leq R_V(\theta), \quad \forall \theta \in \Theta.$$

Diciamo che U è uno stimatore corretto (o non distorto) di $g(\theta)$ se è integrabile per ogni \mathbb{P}^θ e

$$\mathbb{E}^\theta[U] = g(\theta), \quad \forall \theta \in \Theta.$$

Si verifica che se U è corretto, allora $R_U(\theta) = \text{Var}^\theta(U)$.

Esempio 3.4. Dato un campione di legge su \mathbb{R} $\{\mu^\theta, \theta \in \Theta\}$ con momento secondo finito, si verifica che

$$s^2 := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

è uno stimatore corretto della varianza $\sigma^2(\theta)$, dove \bar{X} è la media campionaria del campione.

Tra la classe degli stimatori, ricordiamo, infine, il concetto di stimatori di massima verosimiglianza.

Sia $U : \Omega \rightarrow \Theta$ uno stimatore di θ , U è detto *stimatore di massima verosimiglianza* se, per ogni $\omega \in \Omega$ si ha:

$$L(U(\omega), \omega) = \sup_{\theta \in \Theta} L(\theta, \omega).$$

In generale, non è detto che un tale stimatore esista, o che sia univocamente determinato; tuttavia, quando esiste, viene denotato solitamente con $\hat{\theta}$. Nel caso in cui, per ω fissato, la funzione $\theta \mapsto L(\theta, \omega)$ sia differenziabile, $\hat{\theta}$ verifica l'equazione

$$\left. \frac{d}{d\theta} L(\theta, \omega) \right|_{\theta=\hat{\theta}(\omega)} = 0. \quad (3.1.1)$$

Tuttavia, per calcolare $\hat{\theta}$ si considera la seguente equazione:

$$\left. \frac{d}{d\theta} \log L(\theta, \omega) \right|_{\theta=\hat{\theta}(\omega)} = 0,$$

che equivale alla (3.1.1) sotto opportune assunzioni ed è nota come *equazione di massima verosimiglianza*.

Esempio 3.5. Sia (X_1, \dots, X_n) un campione di legge $N(m, \sigma^2)$ con verosimiglianza

$$L(m, \sigma^2; x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}\right).$$

Si verifica che lo stimatore di massima verosimiglianza per (m, σ^2) è

$$\left(\frac{\sum_{i=1}^n X_i}{n}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\right) = \left(\bar{X}, \frac{n-1}{n} s^2\right).$$

Osservazione 3.6. In tutti gli esempi proposti in questa sezione le variabili X_1, \dots, X_n erano indipendenti tra loro; in generale, nel caso del rumore gaussiano frazionario questa condizione non è soddisfatta.

3.2 Metodo delle varianze aggregate

Il metodo delle varianze aggregate sfrutta l'autosimilarità del rumore gaussiano frazionario \mathbf{X} , ossia il fatto che $\forall m \geq 1, \mathbf{X} = \{X_k : k = 0, 1, 2, \dots\}$ e $m^{1-H}\mathbf{X}^{(m)}$ hanno le stesse distribuzioni congiunte finite, dove $\mathbf{X}^{(m)} = \{X_k^{(m)} : k = 0, 1, 2, \dots\}$ con

$$X_k^{(m)} = \frac{1}{m} (X_{km} + \dots + X_{(k+1)m-1}) \quad \text{per } k = 0, 1, 2, \dots$$

Segue che $\text{Var}(X_k^{(m)}) = m^{2H-2}\text{Var}(X_k) = m^{2H-2}N^{-2H}$ per ogni k .

Per stimare l'indice di Hurst H , consideriamo il seguente stimatore della varianza:

$$\widehat{\text{Var}}(X_k^{(m)}) = \frac{1}{M} \sum_{i=0}^{M-1} (X_i^{(m)} - \overline{X^{(m)}})^2, \quad (3.2.1)$$

dove $\overline{X^{(m)}}$ rappresenta la media campionaria:

$$\overline{X^{(m)}} = \frac{1}{M} \sum_{i=0}^{M-1} X_i^{(m)}$$

e M la parte intera di N/m in entrambe le espressioni, e riportiamo i valori di $\widehat{\text{Var}}(X_k^{(m)})$ rispetto a m in un grafico bi-logaritmico. Se i valori calcolati fossero esatti, giacerebbero tutti su una retta con coefficiente angolare $2H - 2$, quindi, per stimare H tracciamo la retta di regressione lineare tra i punti del grafico che abbiamo ottenuto e ne calcoliamo il coefficiente angolare.

Tuttavia, questo metodo risulta poco affidabile, a partire dallo stimatore scelto per la varianza, la cui efficienza è legata al valore di M . Infatti, se le variabili $X_i^{(m)}$ fossero indipendenti, in (3.2.2) dovremmo sostituire M con $M - 1$ (per quanto visto nella sezione 3.1). Beran [8] dimostra che per valori di M molto grandi (ad esempio, per N grande e m piccolo) questo bias scompare poiché vale il seguente risultato:

$$\mathbb{E} \left[\widehat{\text{Var}}(X_k^{(m)}) \right] \approx \text{Var}(X_k^{(m)}) (1 - CM^{2H-2}), \quad (3.2.2)$$

per una costante C e per $M \rightarrow \infty$.

Dimostrazione. Fissiamo m e denotiamo con $\sigma^2 := \text{Var}(X_k^{(m)})$, $\hat{\sigma}^2 := \widehat{\text{Var}}(X_k^{(m)})$, $Y_i := X_i^{(m)}$, $\bar{Y} := \overline{X^{(m)}}$. Segue che, $\forall i = 0, \dots, M - 1$ le variabili Y_i hanno media 0, varianza σ^2 e, in generale, non sono indipendenti. Vale, inoltre:

$$\begin{aligned} \mathbb{E}[\hat{\sigma}^2] &= \frac{1}{M} \sum_{i=0}^{M-1} \mathbb{E}[(Y_i - \bar{Y})^2] = \frac{1}{M} \sum_{i=0}^{M-1} \mathbb{E}[Y_i^2] - 2\mathbb{E}[Y_i\bar{Y}] + \mathbb{E}[\bar{Y}^2] = \\ &= \frac{1}{M} \sum_{i=0}^{M-1} \left[\sigma^2 - \frac{2}{M} \sum_{j=0}^{M-1} \text{Cov}(Y_i, Y_j) \right] + \sum_{k,j=0}^{M-1} \frac{1}{M^2} \text{Cov}(Y_k, Y_j) = \end{aligned}$$

$$\begin{aligned}
&= \sigma^2 - \frac{1}{M^2} \sum_{i,j=0}^{M-1} \text{Cov}(Y_i, Y_j) = \\
&= \sigma^2 - \frac{1}{M^2} \sum_{i,j=0}^{M-1} \text{Cov}\left(\frac{(B_{(i+1)m} - B_{im})}{m}, \frac{(B_{(j+1)m} - B_{jm})}{m}\right) =
\end{aligned}$$

Dal calcolo eseguito nella dimostrazione della Proposizione 2.13 segue che

$$\begin{aligned}
&= \sigma^2 - \frac{m^{-2}}{M^2} \sum_{i,j=0}^{M-1} \text{Cov}(m^H X_i, m^H X_j) = \sigma^2 - \frac{N^{-2H} m^{2H-2}}{M^2} \sum_{i,j=0}^{M-1} \gamma(|i-j|) = \\
&= \sigma^2 - \frac{\sigma^2}{M^2} \sum_{k=0}^{M-1} \left(\gamma(k) \sum_{i,j=0}^{M-1} \mathbb{1}_{\{|i-j|=k\}} \right) = \\
&= \sigma^2 - \frac{\sigma^2}{M^2} \left(\gamma(0)M + \sum_{k=1}^{M-1} \gamma(k)2(M-k) \right) = \\
&= \sigma^2 \left(1 - \frac{1}{M} - \frac{1}{M^2} \sum_{k=1}^{M-1} \gamma(k)2(M-k) \right) = \\
&= \sigma^2 \left(1 - \frac{1}{M} - \frac{\gamma(1)2(M-1)}{M^2} - \frac{1}{M^2} \sum_{k=2}^{M-1} \gamma(k)2(M-k) \right) = \\
&= \sigma^2 \left(1 - \frac{1}{M} - \frac{\gamma(1)2(M-1)}{M^2} - \frac{1}{M} \sum_{k=2}^{M-1} \gamma(k)2 \left(1 - \frac{k}{M} \right) \right).
\end{aligned}$$

Dallo sviluppo di $f(x) = (1+x)^{2H}$ in $x=0$, per $k \geq 2$ vale che:

$$\begin{aligned}
\gamma(k) &= \frac{1}{2} \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right) = \\
&= \frac{1}{2} \left(k^{2H} \left(1 + \frac{1}{k} \right)^{2H} - 2k^{2H} + k^{2H} \left(1 - \frac{1}{k} \right)^{2H} \right) = \\
&= \frac{1}{2} \left(k^{2H} \left(1 + \frac{2H}{k} + o\left(\frac{1}{k^2}\right) \right) - 2k^{2H} + k^{2H} \left(1 - \frac{2H}{k} + o\left(\frac{1}{k^2}\right) \right) \right) = \\
&= \frac{k^{2H}}{2} o\left(\frac{1}{k^2}\right) = o(k^{2H-2}) \quad \text{per } k \rightarrow \infty.
\end{aligned}$$

Quindi $\mathbb{E}[\sigma^2]$ si comporta come

$$\sigma^2 \left(1 - \frac{1}{M} \sum_{k=2}^{M-1} k^{2H-2} \left(1 - \frac{k}{M} \right) \right) = \sigma^2 \left(1 - CM^{2H-3} \sum_{k=2}^{M-1} 1 - \frac{k}{M} \right),$$

per un'opportuna costante C , e da

$$\sum_{k=2}^{M-1} 1 - \frac{k}{M} = M - 2 - \frac{(M-1)M}{2M} + 1 = \frac{M(M+1)}{2M} = \frac{M+1}{2}$$

si ottiene la tesi per $M \rightarrow \infty$. □

Beran [8] dimostra, inoltre, che per $H > 1/2$ la costante moltiplicativa C è positiva. Tuttavia, per valori di m troppo piccoli il metodo della varianze aggregate si rivela inefficiente,

poiché il grafico che otterremmo risulterebbe troppo sensibile a valori locali dello stimatore e, quindi, scarsamente significativo.

Un ulteriore punto di debolezza si trova nella scelta arbitraria della regione dei punti del grafico rispetto a cui calcolare la retta di approssimazione. Ciò può condurre a stime per H molto diverse fra loro, come mostrato nei grafici delle Figure 3.1, 3.2, 3.3, 3.4, in cui sono state tracciate 3 rette di approssimazione di diverso colore: in nero e in rosso le rette ottenute per regressione lineare sui punti rispettivamente della regione iniziale e finale del grafico; in verde la retta ottenuta per regressione lineare sui valori di m secondo i quali, per quanto visto sopra, ci aspettiamo che il metodo restituisca una stima più accurata dell'indice H .

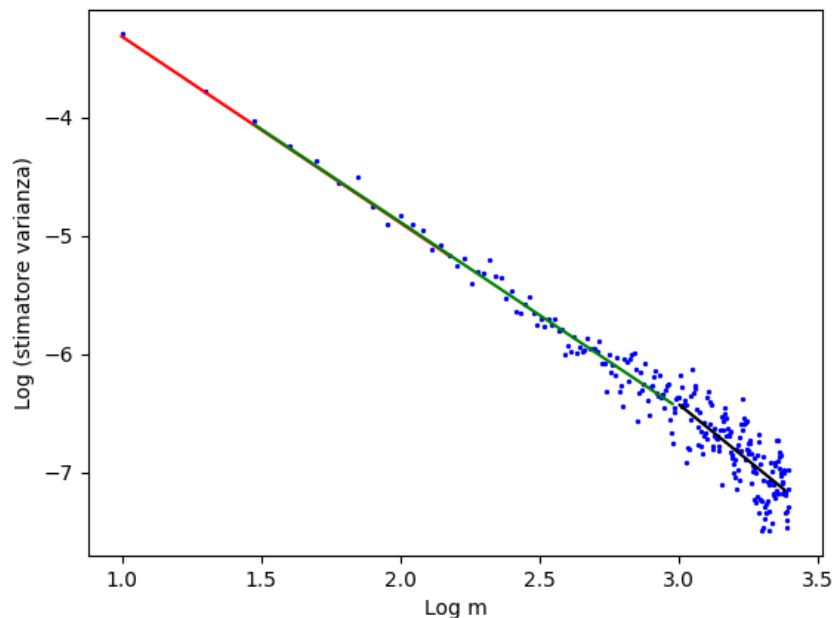


Figura 3.1: $H = 0,2$. Stime di H ottenute:
verde = 0,212; rosso = 0,214; nero = 0,048.

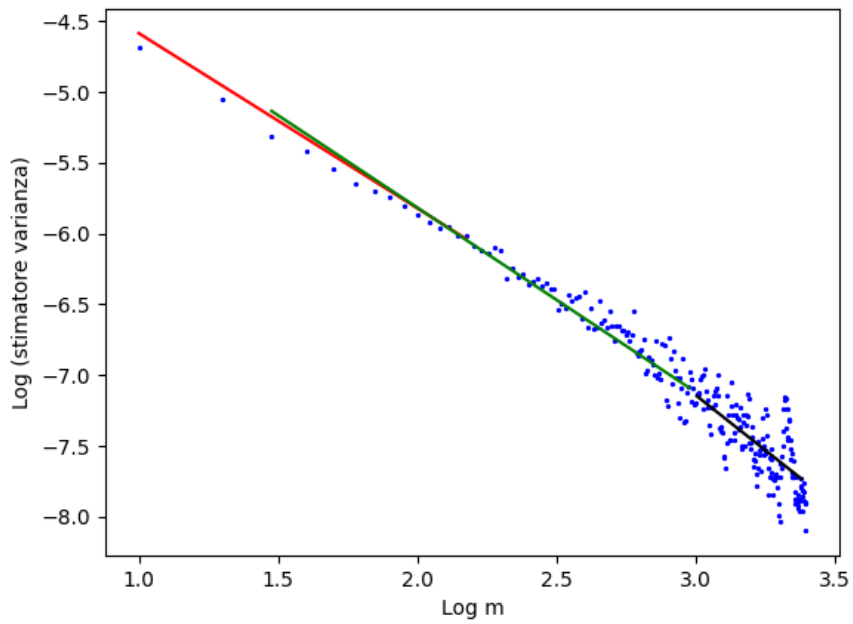


Figura 3.2: $H = 0,42$. Stime di H ottenute:
verde = 0,382; rosso = 0,348; nero = 0,222.

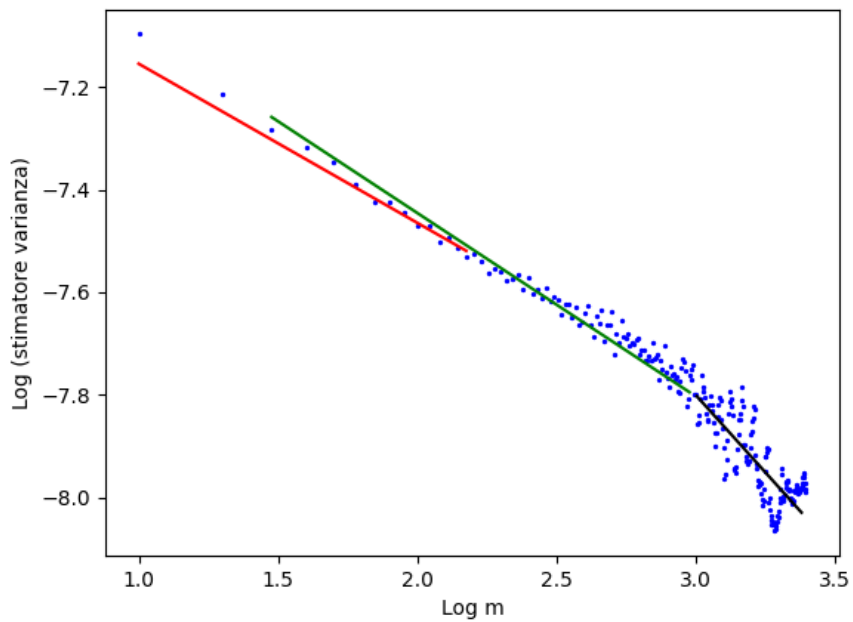


Figura 3.3: $H = 0,8$. Stime di H ottenute:
verde = 0,845; rosso = 0,822; nero = 0,699.

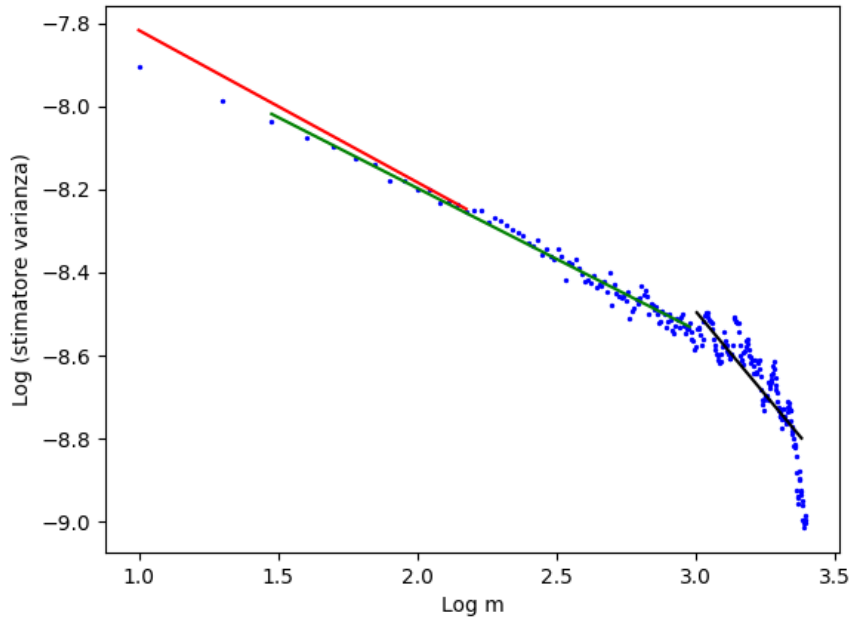


Figura 3.4: $H = 0,9$. Stime di H ottenute:
verde = 0,817; rosso = 0,830; nero = 0,598.

Dalle figure sopra possiamo apprezzare come il metodo fornisca stime dell'indice H molto distanti dal valore corretto per valori di m grandi.

3.3 Metodo dei momenti assoluti

Il metodo dei momenti assoluti è una generalizzazione del precedente e, come esso, sfrutta la proprietà di autosimilarità del rumore gaussiano frazionario e considera come stimatore una generalizzazione di quello appena studiato:

$$AM_m = \frac{1}{M} \sum_{i=0}^{M-1} |X_i^{(m)} - \overline{X^{(m)}}|^n$$

per qualche $n > 0$ e M come nel metodo precedente. In modo analogo a (3.2.1), si può mostrare che $\mathbb{E}[AM_m]$ è proporzionale a $m^{n(H-1)}$, e, quindi, stimare l'indice di Hurst H per regressione lineare rispetto ai punti del grafico bi-logartmico in cui sono rappresentati i valori di $\mathbb{E}[AM_m]$ rispetto a m .

Questo metodo viene tipicamente usato per $n = 1$, ossia il caso dei *momenti assoluti*. Per $n = 2$, invece, si ottiene il metodo delle varianze aggregate, quindi, come quest'ultimo, anche il metodo dei momenti assoluti è caratterizzato dalle stesse condizioni di inaffidabilità.

3.4 Metodo delle variazioni discrete

Tra i metodi basati sulla regressione lineare, quello delle variazioni discrete è uno dei più recenti. Proposto da Coeurjolly [3], si differenzia da quelli appena illustrati per l'uso delle tecniche di filtraggio.

Sia a un filtro di lunghezza $l + 1$ e ordine $p \geq 1$, cioè a è una sequenza tale che

$$\sum_{q=0}^l a_q q^r = 0 \quad \text{per } r = 0, 1, \dots, p - 1,$$

ma non per $r = p$.

Denotiamo con $Y = \{Y_i = \sum_{k=0}^i X_k \mid i = 0, 1, \dots, N - 1\}$ e definiamo il processo V^a come il filtrato del processo Y , ossia:

$$V_k^a = \sum_{q=0}^l a_q Y_{k-q} \quad \text{per } k = l, \dots, N - 1.$$

Dalla definizione segue che $\mathbb{E}[V_i^a] = 0$ perché $\mathbb{E}[Y_i] = 0$ per ogni i . Osserviamo, inoltre, che $V_i^a \sim V_j^a$ (equivalenza in legge), infatti:

$p \geq 1$ implica che $\sum_{q=0}^l a_q = 0$ quindi possiamo riscrivere V_i^a come:

$$V_i^a = \sum_{q=0}^l a_q (Y_{i-q} - Y_i)$$

e per la proprietà degli incrementi stazionari del moto Browniano frazionario segue che $V_i^a \sim V_j^a$.

Dunque V^a è un processo gaussiano stazionario e per l'autosimilarità del moto Browniano frazionario segue che la funzione di autocovarianza $\pi_H^a(\cdot)$ per V^a è del tipo (Coeurjolly [3]):

$$\pi_H^a(j) = \mathbb{E}[V_i^a V_{i+j}^a] = -\frac{1}{2N^{2H}} \sum_{q,r=0}^l a_q a_r |q - r + j|^{2H},$$

dove il termine N^{2H} al denominatore segue dal fatto che le variabili Y_i sono definite su $\{0, \dots, N - 1\}$.

Il metodo delle varianze discrete si basa sul momento k -esimo delle variazioni discrete, così definito:

$$S_N(k, a) = \frac{1}{N-l} \sum_{i=l}^{N-1} |V_i^a|^k \quad \text{per } k > 0.$$

Indichiamo con E_k il momento k -esimo di una variabile gaussiana standard:

$$E_k = 2^{\frac{k}{2}} \Gamma\left(k + \frac{1}{2}\right) / \Gamma\left(\frac{1}{2}\right).$$

Segue che:

$$\mathbb{E}[S_N(k, a)] = E_k N^{-kH} (\text{Var}(N^H V_l^a))^{\frac{k}{2}} = E_k N^{-kH} \left(\sum_{q,r=0}^l a_q a_r |q-r|^{2H} \right)^{\frac{k}{2}}. \quad (3.4.1)$$

Dal risultato appena ottenuto per $\mathbb{E}[S_N(k, a)]$ possiamo ricavare una stima per l'indice H : stimiamo $\mathbb{E}[S_N(k, a)]$ con $S_N(k, a)$, passiamo al logaritmo in entrambi i membri e nel membro di destra si ottiene un'espressione del tipo:

$$-kH \log N + \left[\log E_k + \frac{k}{2} \log \left(\sum_{q,r=0}^l a_q a_r |q-r|^{2H} \right) \right],$$

dove la dipendenza da N compare solo nel termine $\log N$. Quindi dal grafico bi-logaritmico dei valori dello stimatore per $\mathbb{E}[S(k, a)]$ rispetto a N , otteniamo una stima di H tramite il coefficiente angolare della retta di approssimazione lineare.

Tuttavia, questo metodo risulta poco affidabile perché basato solo su l stime di varianze a breve termine, cioè non è in grado di stimare correttamente la presenza, o meno, di dipendenza a lungo raggio. (Dieker [1])

Coeurjolly [3] propone di stimare l'indice H a partire dalla (3.4.1), definendo una sequenza di stimatori del tipo:

$$\hat{H}_N(k, a) := g_{k,a,N}^{-1}(S_N(k, a))$$

dove N è un intero ≥ 1 , k è un reale strettamente positivo, a un filtro e

$$g_{k,a,N}(t) = N^{-tk} (\pi_t^a(0))^{\frac{k}{2}} E_k.$$

Gli stimatori $\hat{H}_N(k, a)$ sono ben definiti poiché dalla derivata della funzione $g_{k,a,N}(t)$ si osserva (Coeurjolly [3]) che $g_{k,a,N}(t)$ è monotona decrescente non appena

$$N > \max_{t \in (0,1)} \exp \left\{ \frac{\sum_{|q-r|>2} a_q a_r \log(|q-r|) |q-r|^{2t}}{\sum_{q,r} a_q a_r |q-r|^{2t}} \right\}.$$

Per il caso $p = 1$, in cui viene scelto il filtro $a = (-1, 1)$ e quindi il processo V^a coincide con il rumore gaussiano frazionario, Peltier and Lévy-Véhel (1994) [4] forniscono una formula esplicita per la sequenza di stimatori:

$$\hat{H}_N(k) := \hat{H}_N(k, a) = -\frac{\log[S_N(k)/E_k]}{k \log N},$$

dove

$$S_N(k) := S_N(k, a) = \frac{1}{N-1} \sum_{i=1}^{N-1} |Y_{i+1} - Y_i|^k \quad \text{con} \quad \mathbb{E}[S_N(k)] = N^{-Hk} E_k,$$

dimostrando che

$$\lim_{N \rightarrow \infty} \hat{H}_N(k) = H.$$

3.5 Metodo del periodogramma

Tra quelli illustrati in questo capitolo il metodo del periodogramma è l'unico che propone una stima dell'indice di Hurst sfruttando la teoria della densità spettrale di potenza.

In generale, il periodogramma è la stima della densità spettrale di un segnale. Nel nostro caso, per stimare la densità spettrale definita come in (2.3.1):

$$f(\lambda) = \sum_{t=-\infty}^{\infty} \gamma(t) \exp(it\lambda) \quad \text{per } -\pi \leq \lambda \leq \pi,$$

consideriamo il periodogramma definito come segue:

$$I(\lambda) = \sum_{k=-(N-1)}^{N-1} \hat{\gamma}(k) \exp(ik\lambda) \quad (3.5.1)$$

dove $\hat{\gamma}(\cdot)$ è la funzione di autocovarianza del campione:

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{j=0}^{N-|k|-1} (X_j - \bar{X})(X_{j+|k|} - \bar{X})$$

e \bar{X} è la media campionaria.

Verifichiamo che per $H \leq 1/2$, il periodogramma è uno stimatore asintoticamente corretto della densità spettrale f del rumore gaussiano frazionario, ossia:

$$\lim_{N \rightarrow \infty} \mathbb{E}[I(\lambda)] = f(\lambda) \quad \text{per } -\pi \leq \lambda \leq \pi.$$

Dimostrazione. Vale che:

$$\mathbb{E}[I(\lambda)] = \mathbb{E} \left[\sum_{k=-(N-1)}^{N-1} \hat{\gamma}(k) \exp(ik\lambda) \right] = \sum_{k=-(N-1)}^{N-1} \mathbb{E}[\hat{\gamma}(k)] \exp(ik\lambda)$$

dove

$$\begin{aligned} \mathbb{E}[\hat{\gamma}(k)] &= \frac{1}{N} \mathbb{E} \left[\sum_{j=0}^{N-|k|-1} (X_j - \bar{X})(X_{j+|k|} - \bar{X}) \right] = \frac{1}{N} \sum_{j=0}^{N-|k|-1} \mathbb{E}[(X_j - \bar{X})(X_{j+|k|} - \bar{X})] = \\ &= \frac{1}{N} \sum_{j=0}^{N-|k|-1} \mathbb{E}[X_j X_{j+|k|}] + \mathbb{E}[\bar{X}^2] - \mathbb{E}[X_{j+|k|} \bar{X}] - \mathbb{E}[X_j \bar{X}] = \\ &= \frac{N-|k|}{N} \gamma(|k|) + \frac{N-|k|}{N} \mathbb{E}[\bar{X}^2] - \frac{1}{N^2} \sum_{j=0}^{N-|k|-1} \sum_{i=0}^{N-1} \left[(\gamma(|i-j|) + \gamma(|i-j-|k|)) \right] = \\ &= \frac{N-|k|}{N} \gamma(k) + \frac{N-|k|}{N^3} \sum_{i,j=0}^{N-1} \gamma(|i-j|) - \frac{1}{N^2} \sum_{i=0}^{N-1} \left(\sum_{j=0}^{N-|k|-1} \gamma(|i-j|) + \right. \\ &\quad \left. + \sum_{j=|k|}^{N-1} \gamma(|i-j|) \right) = \end{aligned}$$

$$= \frac{N - |k|}{N} \gamma(k) - \frac{|k|}{N^3} \sum_{i,j=0}^{N-1} \gamma(|i-j|) - \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=m}^M \gamma(|i-j|)$$

dove $M = \max(|k|, N - |k| - 1)$ e $m = \min(|k|, N - |k| - 1)$.

Dato che la funzione di autocovarianza γ che stiamo considerando è quella relativa al rumore gaussiano frazionario, dalla Proposizione 2.13 segue che per $H \leq 1/2$

$$\sum_{k=0}^{\infty} \gamma(k) < \infty,$$

quindi

$$\mathbb{E}[\hat{\gamma}(k)] \rightarrow \gamma(k) \quad \text{per } N \rightarrow \infty.$$

Inoltre, sappiamo che $H \leq 1/2$ implica $\gamma(k) \leq 0$, da cui segue $\mathbb{E}[\hat{\gamma}(k)] \geq \gamma(k)$ per ogni k ; quindi definitivamente $|\mathbb{E}[\hat{\gamma}(k)]| \leq |\gamma(k)|$. Allora per il teorema di convergenza dominata si ottiene la tesi. \square

Per stimare l'indice di Hurst riportiamo su un grafico bi-logaritmico i valori di $I(\lambda_k)$ per $k = 1, \dots, N$, dove $\lambda_k = \pi k/N$ e tracciamo la retta di approssimazione lineare che, per quanto visto nel capitolo 2, ha coefficiente angolare $1 - 2H$. Dato che la densità spettrale è proporzionale a λ^{1-2H} vicino a $\lambda = 0$, per tracciare la retta di approssimazione prendiamo come riferimento il 10% delle frequenze più basse. Ovviamente, la scelta di tale numero è arbitraria e può condurre a stime differenti dell'indice H , come mostrato nei grafici delle Figure 3.5, 3.6, 3.7, 3.8, in cui sono state tracciate 3 diverse rette di approssimazione: in verde, rosso e nero sono rappresentate le rette ottenute per regressione lineare rispettivamente sul 10%, il 15% e il 20% delle frequenze più basse.

Da un punto di vista computazionale risulta più efficiente implementare questo metodo considerando una formulazione equivalente del periodogramma. Priestley [10] dimostra che la definizione fornita per il periodogramma nella (3.5.1) è equivalente a:

$$I(\lambda) = \frac{1}{N} \left| \sum_{k=0}^{N-1} (X_k - \bar{X}) \exp(ik\lambda) \right|^2, \quad (3.5.2)$$

che può essere facilmente ottenuta sfruttando la trasformata di Fourier a tempi discreti (DTFT).

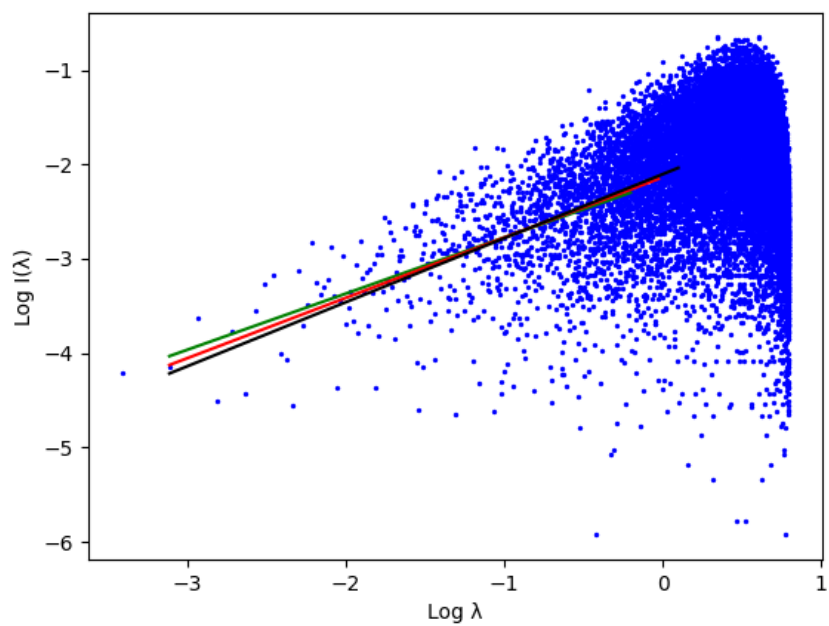


Figura 3.5: $H = 0,2$. Stime di H ottenute:
verde = 0,202; rosso = 0,181; nero = 0,161.

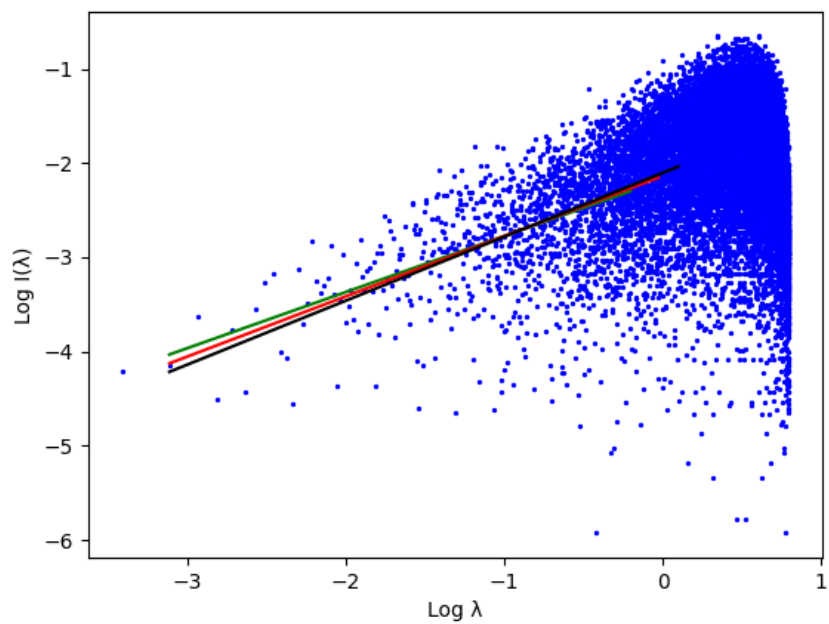


Figura 3.6: $H = 0,42$. Stime di H ottenute:
verde = 0,420; rosso = 0,416; nero = 0,418.

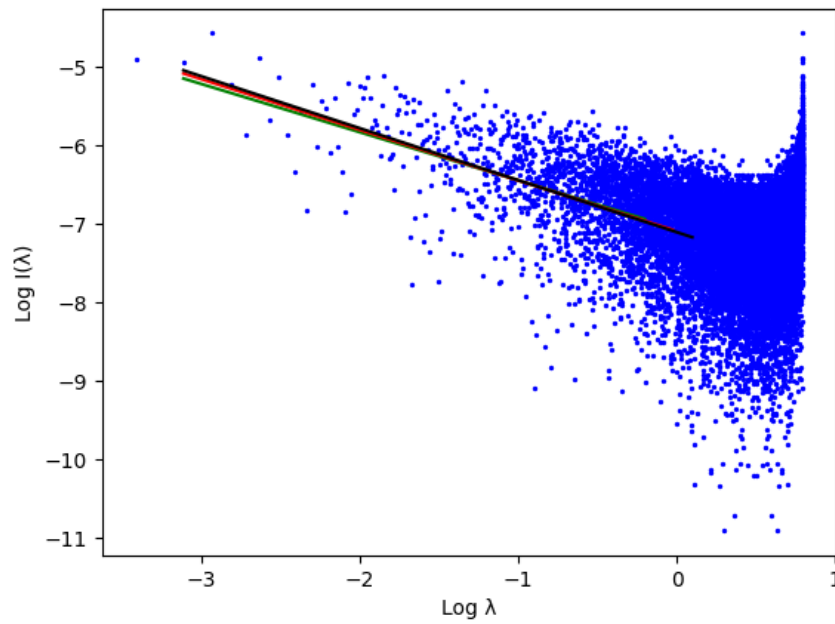


Figura 3.7: $H = 0,8$. Stime di H ottenute:
verde = 0,807; rosso = 0,823; nero = 0,831.

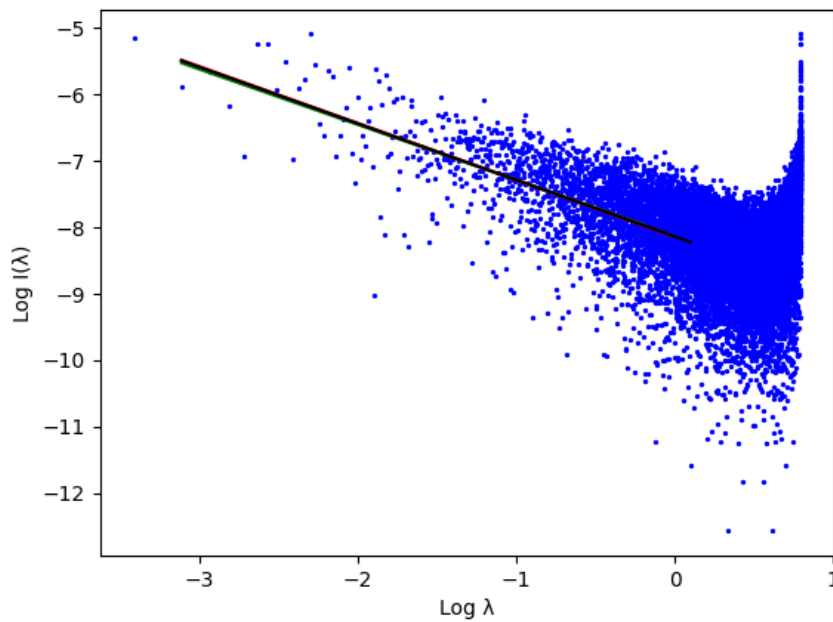


Figura 3.8: $H = 0,9$. Stime di H ottenute:
verde = 0,917; rosso = 0,928; nero = 0,925.

3.6 Metodo R/S

Il metodo R/S, adottato da Hurst nel 1951, prende il nome dalla statistica R/S (in italiano *range aggiustato normalizzato*) utilizzata per stimare l'indice di Hurst H e viene spesso utilizzato per distinguere serie storiche correlate dalle passeggiate aleatorie.

La serie, o il campione, di cui vogliamo stimare l'indice di Hurst, viene divisa in K blocchi disgiunti tutti di lunghezza M , dove M rappresenta la parte intera di N/K , e per ogni blocco viene calcolato il range aggiustato normalizzato $R(t_i, r)/S(t_i, r)$ per ogni valore r del range, dove $t_i = M(i - 1)$ rappresenta la prima posizione dell' i -esimo blocco, per $i = 1, \dots, K$. Analogamente a quanto illustrato nel capitolo 1, $R(t_i, r)$ è calcolato come segue:

$$R(t_i, r) = \max\{W(t_i, 1), \dots, W(t_i, r)\} - \min\{W(t_i, 1), \dots, W(t_i, r)\}$$

dove

$$W(t_i, k) := \sum_{j=0}^{k-1} X_{t_i+j} - k \left(\frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j} \right), \quad k = 1, \dots, r;$$

mentre $S(t_i, r)$ è la radice quadrata della varianza del campione $X_{t_i}, \dots, X_{t_i+r-1}$:

$$S^2(t_i, r) = \frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j}^2 - \left(\frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j} \right)^2.$$

Notiamo che, dato un range r , $R(t_i, r)$ può essere calcolato solo quando $t_i + r \leq N$, e, in tal caso, $R(t_i, r) \geq 0$ poiché $W(t_i, r) = 0$ per definizione.

Le variabili $W(t_i, k)$ possono essere rappresentate in modo equivalente come

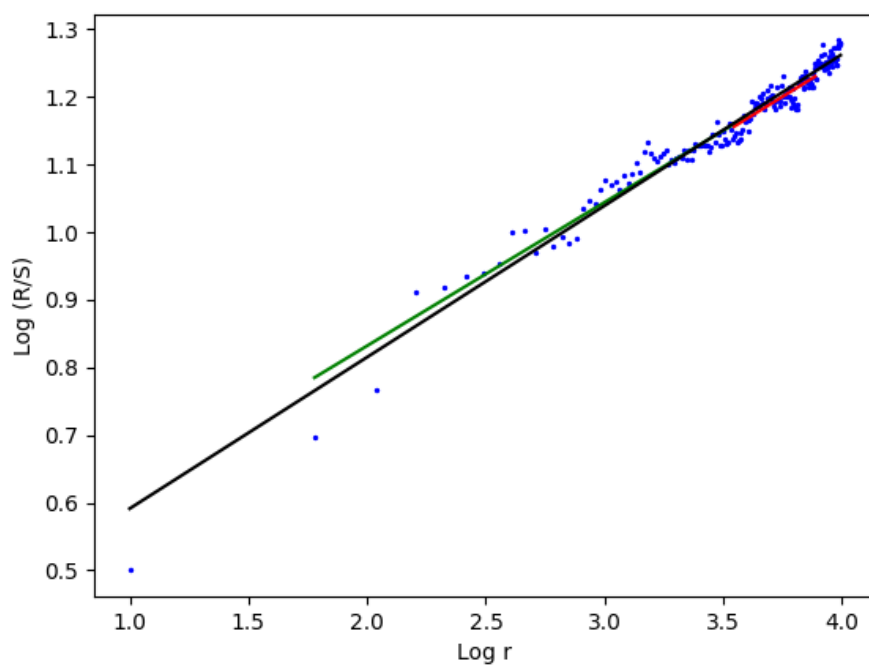
$$W(t_i, k) = \sum_{j=0}^{k-1} (X_{t_i+j} - m)$$

dove m è la media delle variabili contenute nel range r , ossia

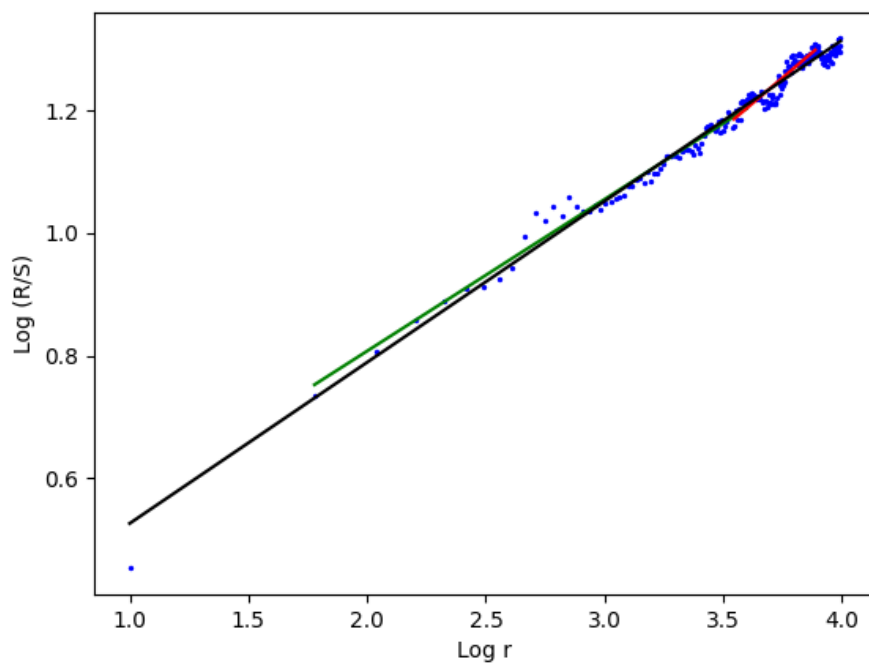
$$m = \frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j}.$$

Dato che le X_j sono le variabili aleatorie del rumore gaussiano frazionario, la media m è un valore vicino a zero. Per valori di r molto grandi possiamo assumere che $m = 0$, da cui segue che l'insieme $\{W(t_i, k) : k = 1, 2, \dots, r\}$ è un campione di moto Browniano frazionario di taglia r e il range $R(t_i, r)$ è la differenza tra il massimo e il minimo di tale campione, ossia una misura di dispersione, come la varianza. Poiché l'ultimo punto di tale campione ha varianza r^{2H} , si tende ad assegnare questo valore anche al range $R(t_i, r)$, quindi, dividendo per la radice della varianza, si ottiene che la statistica R/S è proporzionale a r^H per $r \rightarrow \infty$.

Fissato K , per ogni valore di r otteniamo campioni della statistica R/S di diversa lunghezza; in particolare, per valori piccoli di r otteniamo campioni di lunghezza K , in quanto viene soddisfatta la condizione $t_i + r \leq N$ per ogni $i = 1, \dots, K$, mentre per valori di r vicini ad N il numero di statistiche R/S ottenute sarà sempre più vicino ad 1. Perciò, per stimare l'indice H , riportiamo su un grafico bi-logaritmico la media del campione di R/S rispetto al valore del range r a cui si riferisce il campione e il coefficiente angolare della retta di regressione lineare sarà esso stesso la stima dell'indice di Hurst. Anche in questo caso, la regione del grafico in cui calcolare la retta di approssimazione è scelta arbitrariamente. Dalle figure 3.9, 3.10, 3.11, 3.12 possiamo apprezzare come la scelta della regione influisca sulla precisione della stima: in verde e in rosso sono rappresentate le rette ottenute per regressione lineare sui punti rispettivamente della regione centrale e finale del grafico; in nero la retta ottenuta per regressione lineare su tutti i punti del grafico.

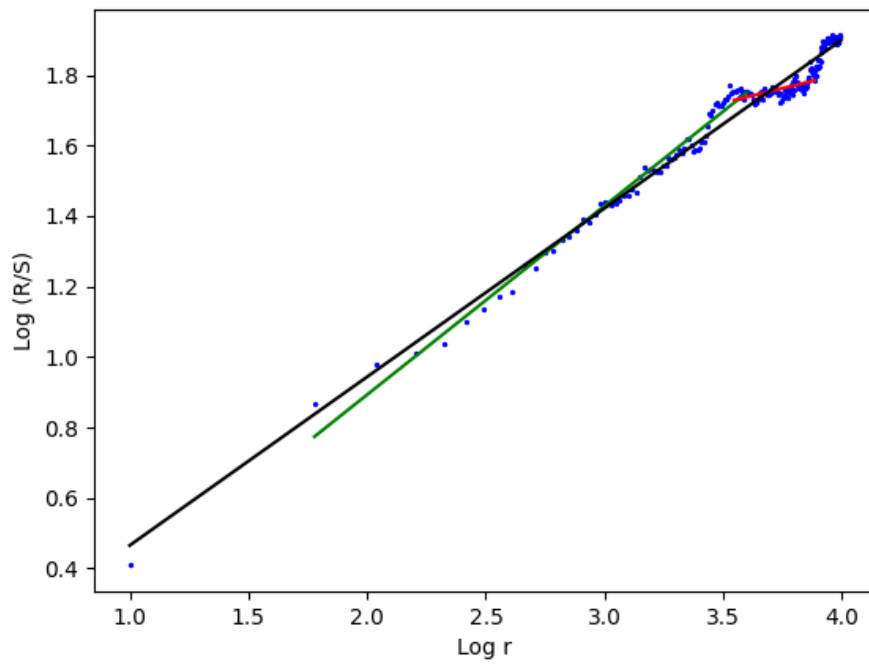


(a) $K = 3$. Stime di H ottenute:
nero = 0,224 ; verde = 0,212; rosso = 0,216.

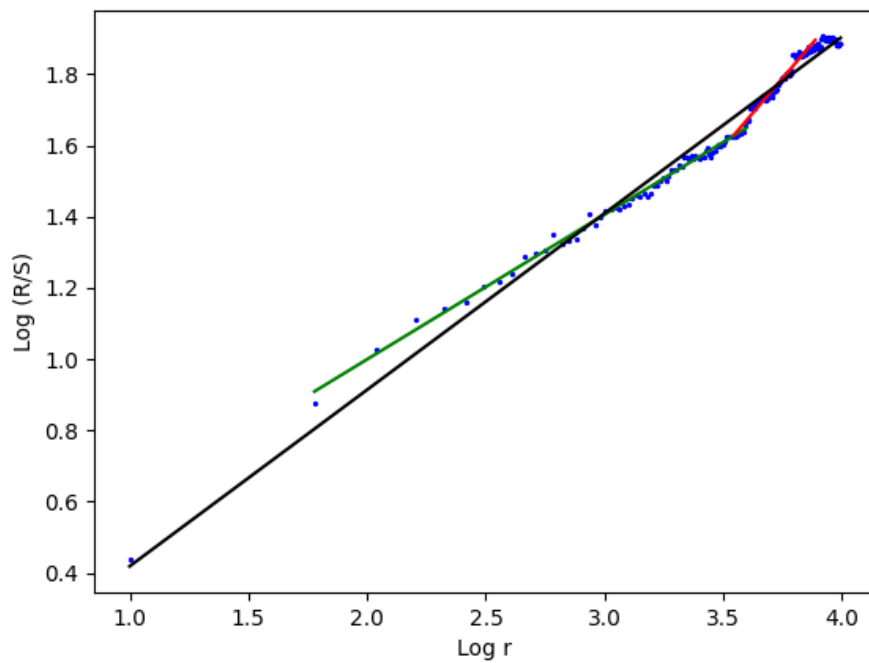


(b) $K = 8$. Stime di H ottenute:
nero = 0,263 ; verde = 0,248; rosso = 0,330.

Figura 3.9: $H = 0,2$.

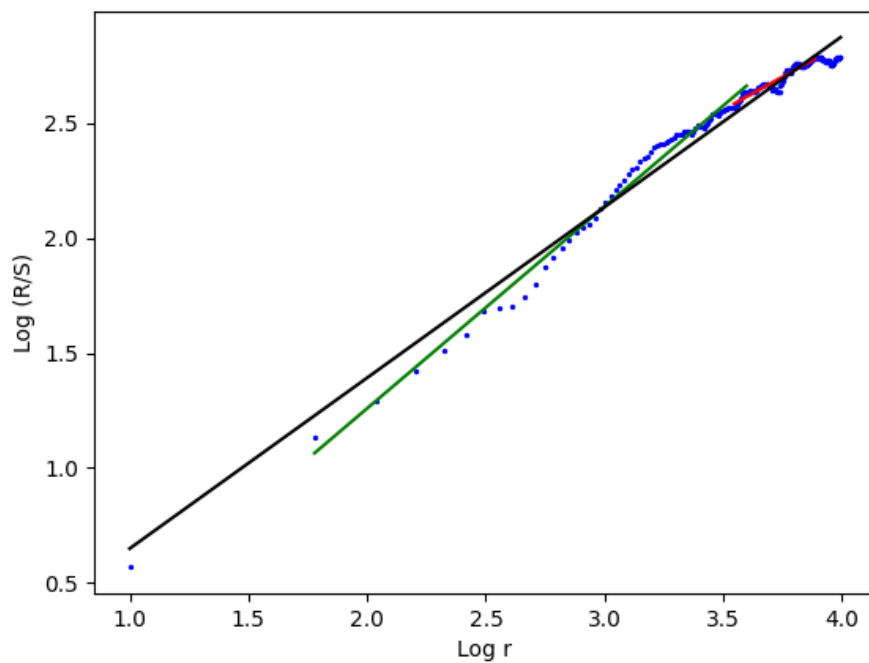


(a) $K = 3$. Stime di H ottenute:
nero = 0,478; verde = 0,536; rosso = 0,165.

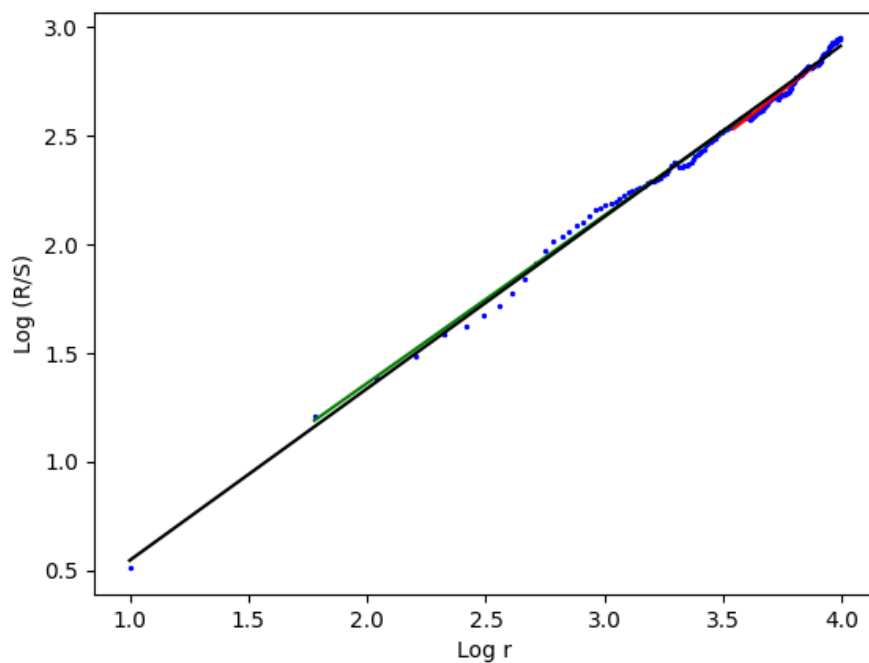


(b) $K = 8$. Stime di H ottenute:
nero = 0,496; verde = 0,407; rosso = 0,779.

Figura 3.10: $H = 0,42$.

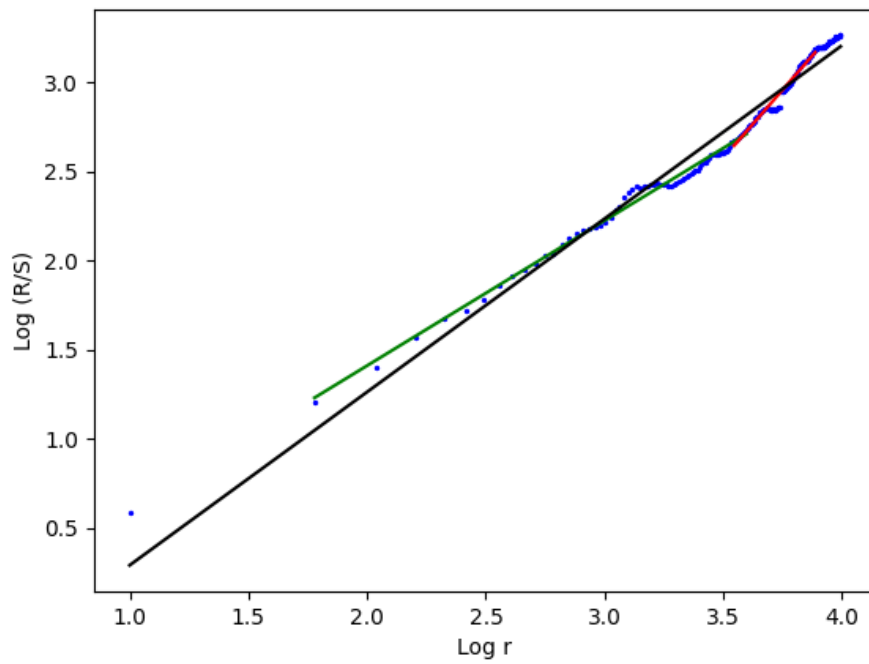


(a) $K = 3$. Stime di H ottenute:
nero = 0,744; verde = 0,879; rosso = 0,573.

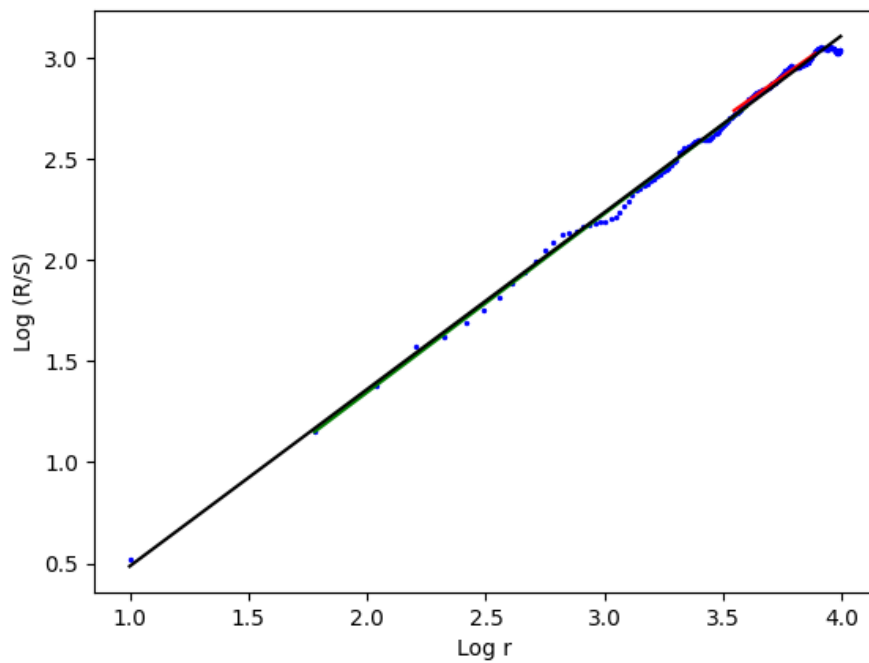


(b) $K = 8$. Stime di H ottenute:
nero = 0,790; verde = 0,772; rosso = 0,853.

Figura 3.11: $H = 0,8$.



(a) $K = 3$. Stime di H ottenute:
nero = 0,973; verde = 0,815; rosso = 1,534.



(b) $K = 8$. Stime di H ottenute:
nero = 0,781; verde = 0,801; rosso = 0,723.

Figura 3.12: $H = 0,9$.

3.7 Sperimentazione numerica

Come i grafici delle figure di questo capitolo, i dati riportati in questa sezione sono stati ottenuti testando i metodi su campioni di rumore gaussiano di taglia $N = 2^{14}$ generati usando il metodo di Davies e Harte (vedi Appendice B). Per effettuare i test abbiamo usato PYTHON in ambiente PYCHARM CE 2022, versione 3.1 su Mac.

3.7.1 Varianze aggregate

Nelle Tabelle 3.1, 3.2, 3.3, 3.4 abbiamo riportato le approssimazioni dell'indice di Hurst H ottenute dal coefficiente della retta di approssimazione lineare, ricavata prendendo come riferimento i punti nelle tre regioni del grafico indicate nella sezione 3.2: `reg_iniz` per la regione iniziale, `reg_fin` per quella finale e `reg_teo` per la regione consigliata dalla teoria e gli errori relativi per ogni stima ottenuta.

| <code>reg_iniz</code> | <code>reg_teo</code> | <code>reg_fin</code> | <code>err_iniz</code> | <code>err_teo</code> | <code>err_fin</code> |
|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|
| 0.199 | 0.189 | 0.017 | 0.006 | 0.055 | 0.915 |
| 0.170 | 0.185 | 0.212 | 0.150 | 0.075 | 0.059 |
| 0.117 | 0.171 | 0.103 | 0.417 | 0.147 | 0.486 |
| 0.144 | 0.225 | -0.161 | 0.279 | 0.124 | 1.808 |
| 0.159 | 0.144 | 0.295 | 0.204 | 0.282 | 0.473 |
| 0.187 | 0.203 | 0.381 | 0.065 | 0.014 | 0.906 |

Tabella 3.1. $H = 0,2$.

| <code>reg_iniz</code> | <code>reg_teo</code> | <code>reg_fin</code> | <code>err_iniz</code> | <code>err_teo</code> | <code>err_fin</code> |
|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|
| 0.456 | 0.445 | 0.656 | 0.085 | 0.059 | 0.562 |
| 0.368 | 0.381 | 0.231 | 0.125 | 0.093 | 0.451 |
| 0.403 | 0.451 | 0.371 | 0.040 | 0.073 | 0.116 |
| 0.453 | 0.455 | 0.460 | 0.078 | 0.084 | 0.095 |
| 0.441 | 0.416 | 0.517 | 0.050 | 0.008 | 0.232 |
| 0.458 | 0.428 | 0.257 | 0.090 | 0.019 | 0.388 |

Tabella 3.2. $H = 0,49$.

| reg_iniz | reg_teo | reg_fin | err_iniz | err_teo | err_fin |
|----------|---------|---------|----------|---------|---------|
| 0.689 | 0.689 | 0.692 | 0.129 | 0.138 | 0.134 |
| 0.751 | 0.756 | 0.674 | 0.061 | 0.055 | 0.157 |
| 0.845 | 0.855 | 0.793 | 0.057 | 0.068 | 0.009 |
| 0.668 | 0.681 | 0.591 | 0.166 | 0.148 | 0.261 |
| 0.741 | 0.730 | 0.741 | 0.074 | 0.087 | 0.074 |
| 0.758 | 0.751 | 0.680 | 0.053 | 0.061 | 0.150 |

Tabella 3.3. $H = 0,8$.

| reg_iniz | reg_teo | reg_fin | err_iniz | err_teo | err_fin |
|----------|---------|---------|----------|---------|---------|
| 0.824 | 0.854 | 0.852 | 0.084 | 0.051 | 0.053 |
| 0.794 | 0.781 | 0.720 | 0.118 | 0.132 | 0.200 |
| 0.742 | 0.744 | 0.369 | 0.176 | 0.173 | 0.590 |
| 0.855 | 0.848 | 0.898 | 0.050 | 0.058 | 0.003 |
| 0.815 | 0.819 | 0.529 | 0.094 | 0.090 | 0.413 |
| 0.839 | 0.826 | 0.873 | 0.067 | 0.083 | 0.0304 |

Tabella 3.4. $H = 0,9$.

Dalle tabelle appena riportate riusciamo a quantificare ed apprezzare quanto la scelta della regione del grafico su cui eseguire la regressione lineare impatti sull'efficienza del metodo e, quindi, sull'efficienza della stima dell'indice di Hurst. Dai dati, inoltre, risulta difficile individuare una regione "di affidabilità" del grafico; in letteratura solitamente si consiglia di scegliere quella centrale.

3.7.2 Periodogramma

Nelle Tabelle 3.5, 3.6, 3.7, 3.8 abbiamo riportato le approssimazioni dell'indice di Hurst H ottenute dal coefficiente della retta di approssimazione lineare, ricavata prendendo come riferimento il 10% (freq_10), il 15% (freq_15) e il 20% (freq_20) delle frequenze più basse, e gli errori relativi per ogni stima. Come i grafici delle figure 3.5, 3.6, 3.7 e 3.8, i seguenti dati sono stati calcolati sfruttando la formulazione equivalente per il periodogramma (3.5.2).

| freq_10 | freq_15 | freq_20 | err_10 | err_15 | err_20 |
|---------|---------|---------|--------|--------|--------|
| 0.172 | 0.162 | 0.156 | 0.141 | 0.191 | 0.221 |
| 0.169 | 0.146 | 0.134 | 0.157 | 0.268 | 0.330 |
| 0.195 | 0.182 | 0.165 | 0.026 | 0.089 | 0.173 |
| 0.188 | 0.177 | 0.162 | 0.061 | 0.116 | 0.191 |
| 0.183 | 0.157 | 0.159 | 0.083 | 0.213 | 0.204 |
| 0.177 | 0.164 | 0.144 | 0.116 | 0.181 | 0.278 |

Tabella 3.5. $H = 0,2$.

| freq_10 | freq_15 | freq_20 | err_10 | err_15 | err_20 |
|---------|---------|---------|--------|--------|--------|
| 0.426 | 0.429 | 0.424 | 0.015 | 0.022 | 0.009 |
| 0.428 | 0.435 | 0.429 | 0.020 | 0.036 | 0.022 |
| 0.428 | 0.426 | 0.414 | 0.019 | 0.015 | 0.015 |
| 0.413 | 0.412 | 0.407 | 0.017 | 0.018 | 0.030 |
| 0.440 | 0.432 | 0.424 | 0.049 | 0.028 | 0.010 |
| 0.429 | 0.421 | 0.411 | 0.022 | 0.001 | 0.021 |

Tabella 3.6. $H = 0,42$.

| freq_10 | freq_15 | freq_20 | err_10 | err_15 | err_20 |
|---------|---------|---------|--------|--------|--------|
| 0.820 | 0.820 | 0.819 | 0.026 | 0.025 | 0.024 |
| 0.818 | 0.819 | 0.830 | 0.023 | 0.024 | 0.038 |
| 0.788 | 0.786 | 0.798 | 0.015 | 0.017 | 0.003 |
| 0.784 | 0.795 | 0.803 | 0.020 | 0.006 | 0.004 |
| 0.828 | 0.814 | 0.827 | 0.034 | 0.018 | 0.034 |
| 0.804 | 0.796 | 0.805 | 0.005 | 0.006 | 0.006 |

Tabella 3.7. $H = 0,8$.

| freq_10 | freq_15 | freq_20 | err_10 | err_15 | err_20 |
|---------|---------|---------|--------|--------|--------|
| 0.911 | 0.911 | 0.915 | 0.012 | 0.012 | 0.016 |
| 0.944 | 0.920 | 0.919 | 0.049 | 0.022 | 0.021 |
| 0.892 | 0.910 | 0.914 | 0.009 | 0.011 | 0.016 |
| 0.909 | 0.915 | 0.913 | 0.010 | 0.016 | 0.015 |
| 0.896 | 0.910 | 0.914 | 0.004 | 0.010 | 0.015 |
| 0.918 | 0.919 | 0.923 | 0.020 | 0.021 | 0.025 |

Tabella 3.8. $H = 0,9$.

Per il periodogramma osserviamo che, ad eccezione del caso vicino a 0, cioè $H = 0,2$, scegliere di eseguire la regressione lineare su una porzione più ampia del 10% non comporta differenze significative in termini di precisione della stima dell'indice di Hurst.

3.7.3 R/S

Nelle tabelle di seguito abbiamo riportato le stime dell'indice di Hurst H ottenute dal coefficiente della retta di regressione lineare, ricavata prendendo come riferimento i punti della regione centrale (reg_cent) e finale (reg_fin) e tutti i punti del grafico (reg_tot) al variare del numero di blocchi K in cui abbiamo diviso il campione di rumore gaussiano; per ogni stima è stato riportato anche l'errore assoluto corrispondente.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.288 | -0.020 | 0.247 | 0.439 | 1.102 | 0.237 |
| 0.252 | 0.276 | 0.260 | 0.261 | 0.379 | 0.300 |
| 0.244 | 0.070 | 0.220 | 0.221 | 0.649 | 0.101 |
| 0.302 | 0.249 | 0.272 | 0.509 | 0.245 | 0.358 |
| 0.219 | 0.311 | 0.220 | 0.096 | 0.557 | 0.102 |
| 0.292 | 0.403 | 0.315 | 0.459 | 1.014 | 0.574 |

Tabella 3.9. $H = 0,2$, $K = 3$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.243 | 0.235 | 0.239 | 0.217 | 0.173 | 0.193 |
| 0.256 | 0.325 | 0.282 | 0.281 | 0.623 | 0.412 |
| 0.253 | 0.296 | 0.268 | 0.264 | 0.478 | 0.340 |
| 0.266 | 0.248 | 0.257 | 0.331 | 0.240 | 0.286 |
| 0.281 | 0.317 | 0.309 | 0.404 | 0.583 | 0.543 |
| 0.258 | 0.140 | 0.236 | 0.290 | 0.300 | 0.178 |

Tabella 3.10. $H = 0,2$, $K = 8$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.393 | 0.590 | 0.420 | 0.065 | 0.404 | 0.000 |
| 0.478 | 0.352 | 0.472 | 0.138 | 0.162 | 0.123 |
| 0.415 | 0.474 | 0.444 | 0.012 | 0.129 | 0.057 |
| 0.496 | 0.573 | 0.479 | 0.182 | 0.365 | 0.141 |
| 0.356 | 0.994 | 0.459 | 0.152 | 1.367 | 0.094 |
| 0.526 | 0.135 | 0.467 | 0.251 | 0.678 | 0.112 |

Tabella 3.11. $H = 0,42$, $K = 3$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.463 | 0.413 | 0.445 | 0.103 | 0.016 | 0.058 |
| 0.399 | 0.442 | 0.403 | 0.051 | 0.052 | 0.042 |
| 0.459 | 0.354 | 0.445 | 0.092 | 0.156 | 0.060 |
| 0.390 | 0.368 | 0.382 | 0.071 | 0.125 | 0.091 |
| 0.443 | 0.374 | 0.433 | 0.054 | 0.111 | 0.032 |
| 0.499 | 0.483 | 0.511 | 0.187 | 0.150 | 0.217 |

Tabella 3.12. $H = 0,42$, $K = 8$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.828 | 0.474 | 0.827 | 0.035 | 0.408 | 0.034 |
| 0.808 | 0.296 | 0.761 | 0.010 | 0.630 | 0.049 |
| 0.775 | 0.680 | 0.782 | 0.031 | 0.150 | 0.023 |
| 0.784 | 0.508 | 0.689 | 0.020 | 0.365 | 0.139 |
| 0.773 | 0.830 | 0.833 | 0.034 | 0.038 | 0.042 |
| 0.874 | 0.923 | 0.893 | 0.092 | 0.154 | 0.116 |

Tabella 3.13. $H = 0,8$, $K = 3$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.856 | 0.705 | 0.795 | 0.070 | 0.118 | 0.007 |
| 0.759 | 0.596 | 0.719 | 0.051 | 0.255 | 0.102 |
| 0.709 | 0.909 | 0.741 | 0.114 | 0.136 | 0.074 |
| 0.758 | 0.873 | 0.764 | 0.053 | 0.091 | 0.045 |
| 0.783 | 0.674 | 0.749 | 0.022 | 0.157 | 0.064 |
| 0.830 | 0.475 | 0.730 | 0.037 | 0.406 | 0.088 |

Tabella 3.14. $H = 0,8$, $K = 8$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 1.001 | 0.756 | 0.893 | 0.112 | 0.160 | 0.008 |
| 0.928 | 0.230 | 0.753 | 0.031 | 0.744 | 0.163 |
| 0.910 | 0.735 | 0.852 | 0.012 | 0.183 | 0.053 |
| 0.984 | 1.041 | 0.932 | 0.093 | 0.158 | 0.035 |
| 0.939 | 0.604 | 0.874 | 0.043 | 0.329 | 0.029 |
| 0.865 | 1.177 | 0.932 | 0.039 | 0.308 | 0.035 |

Tabella 3.15. $H = 0,9$, $K = 3$.

| reg_cent | reg_fin | reg_tot | err_cent | err_fin | err_tot |
|----------|---------|---------|----------|---------|---------|
| 0.834 | 1.182 | 0.914 | 0.074 | 0.314 | 0.015 |
| 0.878 | 0.945 | 0.888 | 0.025 | 0.050 | 0.013 |
| 0.930 | 0.867 | 0.940 | 0.034 | 0.037 | 0.045 |
| 0.784 | 0.648 | 0.795 | 0.129 | 0.280 | 0.117 |
| 0.809 | 0.538 | 0.736 | 0.101 | 0.403 | 0.183 |
| 0.899 | 0.716 | 0.816 | 0.001 | 0.204 | 0.093 |

Tabella 3.16. $H = 0,9$, $K = 8$.

Sperimentalmente osserviamo che la regione finale del grafico, ossia quella che racchiude i valori della statistica R/S corrispondenti ai valori più grandi del range r , risulta poco affidabile per stimare H tramite regressione lineare. Tra le altre due regioni scelte, invece, nessuna risulta sostanzialmente più affidabile dell'altra, coerentemente a quanto previsto dalla teoria.

Concludiamo osservando che il metodo del periodogramma risulta il più efficiente e stabile tra i metodi implementati, sia per $H < 1/2$, sia per $H > 1/2$.

Capitolo 4

Reti neurali

Le reti neurali sono modelli parametrici molto versatili in grado di approssimare complessi pattern di dati tramite moderni ed ottimizzati metodi di addestramento. Sebbene tali modelli abbiano avuto molto successo, la teoria sviluppata attorno a questi è ancora debole e, soprattutto, molto distante dalle applicazioni pratiche che ne vengono fatte. In questo capitolo illustreremo il concetto di reti neurali e alcuni dei principali risultati teorici relativi ad esse; infine, sfrutteremo la loro struttura per ottenere, dato in input un campione di rumore gaussiano frazionario, una stima dell'indice di Hurst corrispondente.

4.1 Struttura di una rete neurale

Le reti neurali sono famiglie di funzioni non lineari parametrizzate da $(\theta_i = (W_i, b_i))_i$, dove W_i sono detti pesi e b_i bias; sono costituite da una serie di livelli, detti *layers*, in cui si alternano funzioni di attivazione non lineari e funzioni affini del tipo $x \mapsto W_i x + b_i$. Formalmente:

Definizione 4.1. Una rete neurale completamente connessa feed-forward (NN) è definita tramite un'architettura $\alpha = (n, \psi)$ con $l \in \mathbb{N}_{\geq 2}$ numero di layers (o livelli), $n \in \mathbb{N}^{l+1}$ e $\psi : \mathbb{R} \rightarrow \mathbb{R}$ funzione di attivazione. Si indicano con n_0, n_l e n_i per $i = 1, \dots, l-1$ il numero di neuroni (o unità) rispettivamente in input, output e nell' i -esimo hidden layer. Denotando con $p(n) = \sum_{i=1}^l n_i n_{i-1} + n_l$ si definisce una funzione di realizzazione

$$\Phi_\alpha : \mathbb{R}^{n_0} \times \mathbb{R}^{p(n)} \rightarrow \mathbb{R}^{n_l}$$

che, fissato un parametro θ ,

$$\theta = (\theta_i)_{i=1}^l = (W_i, b_i)_{i=1}^l \in \times_{i=1}^l (\mathbb{R}^{n_i \times n_{i-1}} \times \mathbb{R}^{n_i}) \cong \mathbb{R}^{p(n)},$$

per ogni input $x \in \mathbb{R}^{n_0}$, calcola $\Phi_\alpha(\cdot; \theta) : \mathbb{R}^{n_0} \rightarrow \mathbb{R}^{n_l}$ come $\Phi_\alpha(x; \theta) = \Phi_\alpha^{(l)}(x; \theta)$ dove:

$$\begin{aligned}\Phi_\alpha^{(1)}(x; \theta) &= W_1 x + b_1, \\ \bar{\Phi}_\alpha^{(1)}(x; \theta) &= \psi(W_1 x + b_1),\end{aligned}$$

$$\begin{aligned} & \vdots \\ \Phi_{\alpha}^{(l-1)}(x; \theta) &= W_{(l-1)} \bar{\Phi}_{\alpha}^{(l-2)}(x; \theta) + b_{l-1}, \\ \bar{\Phi}_{\alpha}^{(l-1)}(x; \theta) &= \psi(\Phi_{\alpha}^{(l-1)}(x; \theta)), \\ \Phi_{\alpha}^{(l-1)}(x; \theta) &= W_{(l)} \bar{\Phi}_{\alpha}^{(l-1)}(x; \theta) + b_l. \end{aligned}$$

Le funzioni $\Phi_{\alpha}^{(i)}$ e $\bar{\Phi}_{\alpha}^{(i)}$ sono note, rispettivamente, come pre-attivazione e attivazione dei neuroni nell' i -esimo layer, mentre si definiscono ampiezza e profondità di una rete neurale rispettivamente $\|n\|_{\infty}$ e l ($l - 1$ è il numero di hidden layers).

Quando il numero di livelli è maggiore o uguale a 2, la rete viene detta *profonda*.

Il grafo aciclico in Figura 4.1 è un esempio di NN ottenuto intervallando le mappe affini $x \mapsto W_i x + b_i \forall i = 1, \dots, l$ con la funzione di attivazione ψ .

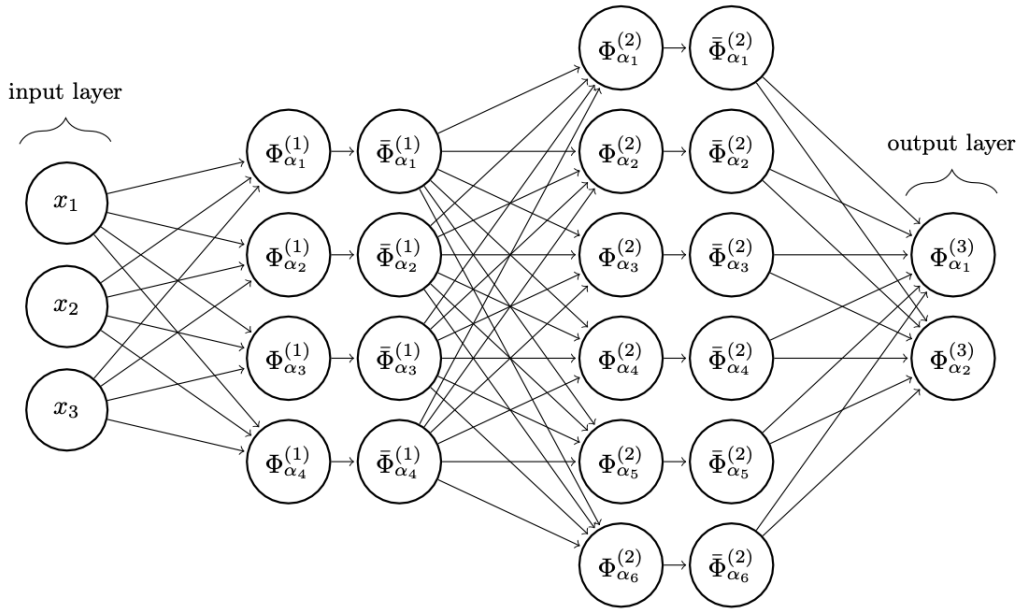


Figura 4.1: Grafo associato ad una rete neurale con 2 hidden layers; architettura $\alpha = ((3, 4, 6, 2), \psi)$ e parametri $\theta = (W_i, b_i)_{i=1}^3$ con $\Phi_{\alpha} : \mathbb{R}^3 \times \mathbb{R}^{60} \rightarrow \mathbb{R}^2$.

Analogamente a quanto accade in statistica parametrica, in cui l'obiettivo, date delle osservazioni $(X_i)_{i=1}^n$ (solitamente i.i.d.) tutte con legge \mathbb{P}^{θ} , è trovare una stima del parametro θ , lo scopo dell'utilizzo della struttura di rete neurale è addestrare il computer a trovare proprio il "miglior" $\hat{\theta}$. Nel nostro caso, ad esempio, vorremmo definire una funzione che, dato in input un campione di rumore gaussiano frazionario, stabilisce il corrispondente valore dell'indice di Hurst H . Fissando le funzioni di attivazione, il numero di livelli e di neuroni per ogni livello (in questo caso n_0 è la taglia del campione), per ogni $\theta = ((W_i, b_i)_{i=1}^l)$ abbiamo appena definito una rete neurale, ossia un modello, che verrà istruito a trovare la "miglior" sequenza di parametri $(\hat{\theta})_{i=1}^l$ affinché la funzione $f(\cdot, \theta) : \mathbb{R}^{n_0} \rightarrow \mathbb{R}$, dato in input un campione di rumore gaussiano frazionario o una serie storica, restituisca una stima, più precisa e affidabile

possibile, del corrispondente valore dell'indice di Hurst.

Una volta definita l'architettura di una rete neurale, per poter determinare $\hat{\theta}$ è, quindi, necessario stabilirne la fase di apprendimento.

Per determinare $\hat{\theta}$ si introduce un criterio di ottimalità, detto funzione di perdita, o *loss function*. Le più note sono la funzione che calcola l'errore assoluto medio (MAE) o l'errore quadratico medio (MSE); quest'ultima sarà quella che utilizzeremo.

Dato un insieme, detto *data set*, di elementi del tipo (x_i, y_i) , dove, $\forall i$, x_i è un valore di input (ad esempio, il campione di fGn) e y_i l'output corrispondente (ad esempio, l'esatto valore dell'indice di Hurst del campione in input), la fase di apprendimento della rete consiste nel minimizzare la funzione:

$$\theta \mapsto \frac{1}{M} \sum_{i=1}^M ((f(x_i; \theta) - y_i)^2 =: \mathfrak{L}((x_i, y_i)_{i=1}^M; \theta), \quad (4.1.1)$$

ossia la nostra funzione di perdita.

La minimizzazione avviene generalmente tramite algoritmi iterativi, il più utilizzato è quello della discesa stocastica del gradiente, e termina appena viene raggiunto il criterio d'arresto stabilito.

L'apprendimento viene eseguito su un sottoinsieme del data set, noto come *training set* e, una volta terminato, la rete viene testata sul *test set*, sottoinsieme del data set costituito da elementi che la rete non ha mai incontrato durante la fase di addestramento. A questo punto, se i risultati del test sono soddisfacenti, abbiamo ottenuto la funzione che cercavamo, la quale verrà eseguita ogni volta con i pesi e i bias presenti nel $\hat{\theta}$ ottenuto nella fase di apprendimento.

Presentiamo uno dei principali risultati teorici relativi alle reti neurali: il teorema di approssimazione universale, dimostrato nel 1989 da George Cybenko.

Data una NN con un solo hidden-layer composto da m neuroni, k neuroni in input, uno in output e con funzione di attivazione ψ , l'insieme delle funzioni implementabili al variare di $\theta = (W, b, c) \in \mathbb{R}^{m \times k} \times \mathbb{R}^m \times \mathbb{R}^m$ è

$$\mathcal{N}_{k,m}(\psi) = \left\{ h : \mathbb{R}^k \rightarrow \mathbb{R} \mid h(x) = \sum_{i=1}^m c_i \psi(W_i x - b_i) \right\}.$$

Al variare di $m \in \mathbb{N}_0$ otteniamo:

$$\mathcal{N}_k(\psi) = \bigcup_{m=1}^{\infty} \mathcal{N}_{k,m}(\psi),$$

ovvero lo spazio delle funzioni generabili tramite reti neurali.

Teorema 4.2 (Teorema di approssimazione universale). *Fissata $\psi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua e sigmoideale, ossia*

$$\psi(t) \xrightarrow{t \rightarrow -\infty} 0 \quad e \quad \psi(t) \xrightarrow{t \rightarrow +\infty} 1,$$

vale che $\mathcal{N}_k(\psi)$ è denso in $(C(X), \|\cdot\|_{\infty})$ per ogni $X \subset \mathbb{R}^k$ compatto.

Il teorema dimostra, dunque, che una rete neurale con un singolo strato nascosto con un numero sufficiente (e finito) di neuroni può approssimare qualsiasi funzione continua. Tuttavia, il teorema di approssimazione universale è un teorema di esistenza, cioè afferma che una rete con un singolo strato nascosto può approssimare qualsiasi funzione continua, ma non fornisce alcuna indicazione su come ottenerla, come costruirla e se un singolo strato sia la scelta più efficiente.

4.1.1 Algoritmo di discesa stocastica del gradiente

L'algoritmo di discesa stocastica del gradiente prevede di calcolare il gradiente v della funzione di perdita \mathcal{L} sugli elementi del training set (x_i, y_i) per individuare la direzione che minimizza la funzione, cioè l'errore. Lo scopo è raggiungere il minimo globale della funzione, quindi, una volta valutato il gradiente v , aggiorna il parametro θ , $\theta \mapsto \theta - \lambda v$ dove λ è detto tasso di apprendimento (*learning rate*). Poiché la funzione f non è lineare come funzione di θ , al quadrato non è convessa rispetto a θ , quindi, l'algoritmo potrebbe arrestarsi in presenza di un minimo locale, ma non globale; per evitare ciò si utilizza un fattore di "inerzia" μ , *momentum*. A questo punto l'algoritmo riparte finché la funzione di perdita assumerà un valore minore al criterio di arresto stabilito.

L'attributo *stocastica* deriva dal fatto che ad ogni iterazione non viene valutato l'intero training set, piuttosto viene scelto in modo casuale un suo sottoinsieme (*mini-batch*) $((x_i, y_i))_{i=1}^n$, $n < M$, con cui valutare il gradiente della funzione \mathcal{L} , ossia:

$$\nabla_{\theta} \frac{1}{n} \sum_{i=1}^n ((f(x_i; \theta) - y_i)^2) = v.$$

Per ogni iterazione, possiamo valutare il gradiente di \mathcal{L} su più di un mini-batch e ogni scelta di quest'ultimo viene chiamata *epoca*.

La scelta di questo algoritmo è, inoltre, volta a limitare due problemi di costo computazionale che dobbiamo fronteggiare: un'elevata dimensione dell'insieme dei parametri $\theta \in \Theta$ e del data set. La discesa del gradiente ci permette, a differenza di altri metodi iterativi come quello di Newton, di non calcolare l'inverso dell'Hessiano della funzione \mathcal{L} , che potrebbe essere assai costoso da un punto di vista computazionale; analogamente la scelta stocastica degli elementi su cui valutare il gradiente ci permette di generare un training set più vasto ed efficace per l'addestramento della rete neurale.

4.2 Stima dell'indice di Hurst

Il tipo di rete illustrato fin qui, rete neurale completamente connessa feed-forward, è una rete in cui:

- i neuroni sono organizzati in layer in cui non ci sono collegamenti a loop (né tra layers differenti, né all'interno dello stesso layer);

- non si fanno ipotesi a priori, quindi si collegano tutti i neuroni di un layer a tutti i neuroni del layer precedente.

Per la parte di sperimentazione numerica abbiamo, invece, utilizzato reti *convoluzionali*, che in generale non sono completamente connesse.

Dato un segnale $(s(t))_{t \in \mathbb{Z}}$, ad esempio un campione di rumore gaussiano frazionario, e un filtro $(f(t))_{t \in \mathbb{Z}}$ l'operazione di convoluzione tra s e t è definita come:

$$(s * f)(t) = \sum_{k \in \mathbb{Z}} f(k)s(t - k)$$

Se consideriamo, per ogni k , i fattori $s(t - k)$ e $f(k)$ rispettivamente come un dato di input e un parametro, l'operazione $(s * f)$ è una funzione lineare, esprimibile quindi tramite un prodotto matrice-vettore. Una rete convoluzionale si differenzia proprio per la scelta della funzione di pre-attivazione Φ_α , la quale, in almeno un layer della rete, viene codificata, invece che da una generica coppia (W, b) , da una matrice che corrisponde all'operazione di convoluzione tra il segnale e un filtro f .

Come funzioni di attivazione utilizzeremo le *RELU*: $RELU(x) = x^+ = \max(0, x)$.

4.2.1 Modello

L'utilizzo della struttura di rete neurale per stimare l'indice di Hurst di un campione di rumore gaussiano frazionario, in modo alternativo rispetto ai metodi classici analizzati nel Capitolo 3, ci è stato suggerito dall'articolo *Long-Range Dependent Traffic Classification with Convolutional Neural Networks Based on Hurst Exponent Analysis*, pubblicato il 15 ottobre 2020 sulla rivista accademica "Entropy" [2].

Gli autori testano l'efficienza delle reti neurali convoluzionali, da loro costruite, come metodo di stima dell'indice di Hurst di un campione di rumore gaussiano frazionario, al variare dell'architettura della rete e degli algoritmi iterativi per la minimizzazione della funzione di perdita. La scelta di quest'ultima ricade sulla MSE, sfruttando la dimostrazione di Rea [11] secondo cui la media dell'errore quadratico decresce rispetto ad un aumento della taglia del campione o della serie storica in input. Il risultato finale non consiste esattamente nel valore della stima ottenuta per l'indice di Hurst, bensì in un intervallo: l'insieme $(0, 1)$ viene partizionato in nove sottointervalli e la rete restituisce come output l'intervallo di appartenenza dell'indice H stimato.

Vengono testate reti con 2, 3 e 4 hidden layers, di cui tuttavia non è specificato il numero di neuroni, allenate su Training set di 43,200 campioni di taglia $N = 10000$ e valutate su validation set e test set di 5400 elementi ciascuno. Considerando il tempo come una dimensione spaziale, sono state costruite reti che prendono in input campioni bidimensionali, ossia non di taglia 10000×1 , bensì 50×200 e 1000×10 .

4.2.2 Sperimentazione numerica

Guidati dall'articolo indicato sopra, abbiamo creato un data set di 7000 campioni di rumore gaussiano frazionario di taglia $N = 10000$, di cui il 30% destinato a costituire il test set, con cui

abbiamo allenato reti neurali convoluzionali, che differiscono tra loro per numero di layers e dimensione dell'input, usando l'errore quadratico medio (4.1.1) come funzione di perdita, minimizzato tramite l'algoritmo di discesa stocastica del gradiente per cui abbiamo lasciato fisso il numero epoche, 24, e la cardinalità del mini-batch, 32. Abbiamo considerato 24 epoche poiché, come possiamo riportare in Figura 4.2, abbiamo osservato che il comportamento della funzione di perdita in fase di addestramento si stabilizzava attorno alla 20-esima epoca; per il mini-batch, invece, non abbiamo riscontrato differenze significative rispetto alla variazione della cardinalità.

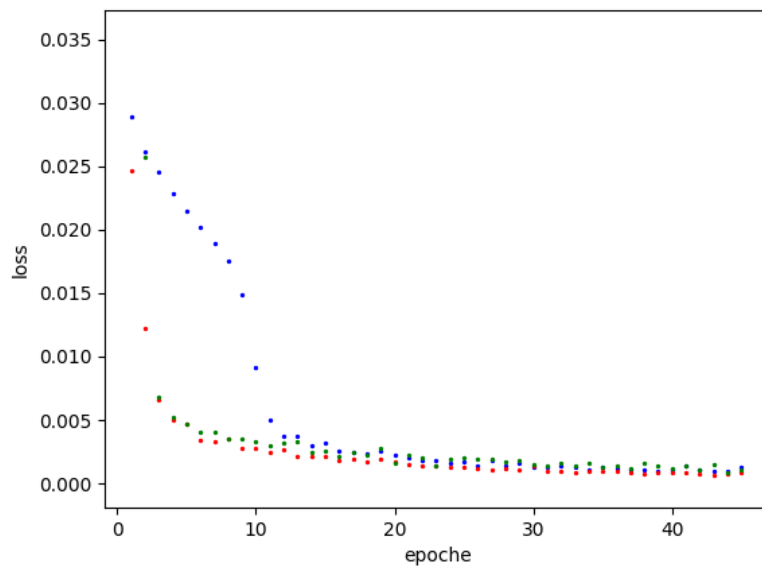


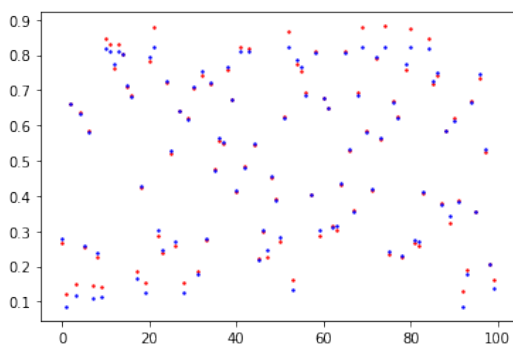
Figura 4.2: Riportiamo il comportamento della funzione loss al variare del numero di epoche: in blu è indicato il caso in cui la rete è composta da 2 hidden layers, in rosso il caso da 3 layers e in verde da 4.

Nelle tabelle di seguito abbiamo riportato l'errore quadratico medio (MSE) e l'errore assoluto medio (MAE) commessi dalle reti valutate sugli elementi del test set; ogni riga corrisponde una nuova rete e ad ogni colonna l'errore quadratico medio mse_k o l'errore assoluto medio mae_k commesso dalla rete, con k che ne indica il numero di layers da cui essa è costituita. Ad ogni tabella sono associati tre grafici in cui vengono confrontati i valori esatti dell'indice di Hurst (in rosso) con le stime ottenute sperimentalmente (in blu).

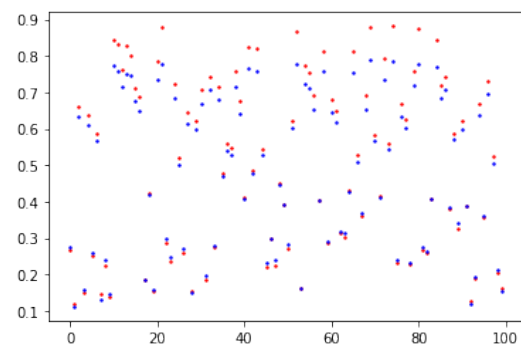
Tutte le simulazioni sono state condotte utilizzando le componenti della libreria Keras di Python.

| mse_2 | mae_2 | mse_3 | mae_3 | mse_4 | mae_4 |
|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1.4×10^{-3} | 2.6×10^{-2} | 1.4×10^{-4} | 9.3×10^{-3} | 8.0×10^{-4} | 2.0×10^{-2} |
| 2.8×10^{-4} | 1.2×10^{-2} | 2.7×10^{-3} | 3.7×10^{-2} | 4.1×10^{-4} | 1.4×10^{-2} |
| 7.6×10^{-4} | 2.1×10^{-2} | 1.6×10^{-3} | 2.9×10^{-2} | 1.9×10^{-4} | 1.0×10^{-2} |
| 5.6×10^{-4} | 2.0×10^{-2} | 5.2×10^{-2} | 2.0×10^{-1} | 7.7×10^{-4} | 1.9×10^{-2} |
| 9.5×10^{-4} | 2.2×10^{-2} | 3.8×10^{-3} | 4.4×10^{-2} | 8.6×10^{-4} | 2.2×10^{-2} |
| 3.4×10^{-4} | 1.1×10^{-2} | 1.3×10^{-3} | 2.9×10^{-2} | 6.4×10^{-4} | 2.0×10^{-2} |

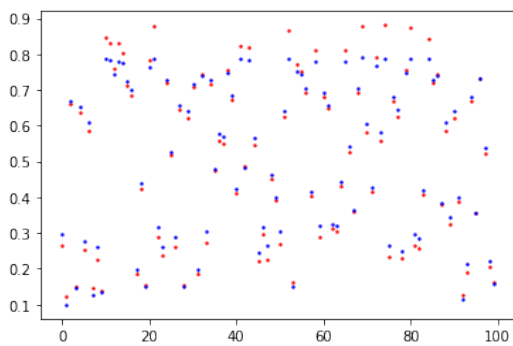
Tabella 4.1. Rete unidimensionale, 1 layer convoluzionale.



(a) NN con 2 layers



(b) NN con 3 layers

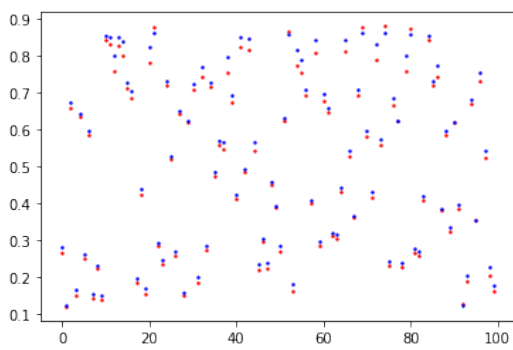


(c) NN con 4 layers

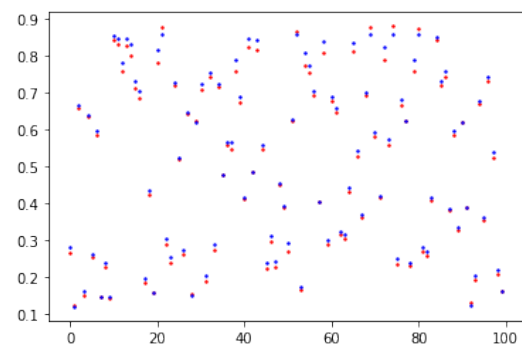
Figura 4.3: Rete unidimensionale, 1 layer convoluzionale.

| mse_2 | mae_2 | mse_3 | mae_3 | mse_4 | mae_4 |
|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 9.7×10^{-4} | 2.7×10^{-2} | 7.8×10^{-4} | 1.6×10^{-2} | 4.0×10^{-4} | 1.8×10^{-2} |
| 2.9×10^{-4} | 1.4×10^{-2} | 1.9×10^{-4} | 1.1×10^{-2} | 3.3×10^{-4} | 1.5×10^{-2} |
| 1.6×10^{-4} | 9.5×10^{-3} | 2.2×10^{-4} | 1.1×10^{-2} | 1.9×10^{-4} | 7.6×10^{-3} |
| 1.4×10^{-4} | 8.6×10^{-3} | 1.2×10^{-3} | 3.2×10^{-2} | 1.7×10^{-4} | 9.7×10^{-3} |
| 1.2×10^{-4} | 7.9×10^{-3} | 4.5×10^{-4} | 1.8×10^{-2} | 2.7×10^{-4} | 1.2×10^{-2} |
| 2.9×10^{-4} | 1.3×10^{-2} | 2.5×10^{-4} | 1.2×10^{-2} | 7.3×10^{-4} | 1.9×10^{-2} |

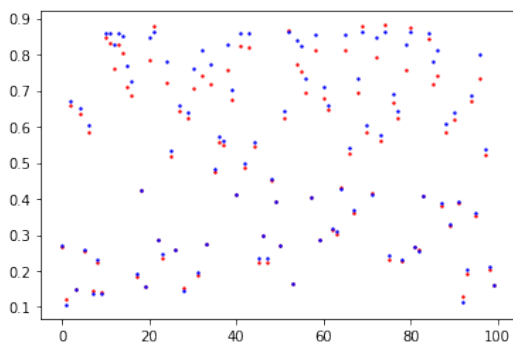
Tabella 4.2. Rete bidimensionale, dati di input 50×200 , 1 layer convoluzionale.



(a) NN con 2 layers



(b) NN con 3 layers



(c) NN con 4 layers

Figura 4.4: Rete bidimensionale, dati di input 50×200 , 1 layer convoluzionale.

| mse_2 | mae_2 | mse_3 | mae_3 | mse_4 | mae_4 |
|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 3.1×10^{-4} | 1.6×10^{-2} | 6.4×10^{-4} | 1.7×10^{-2} | 2.9×10^{-4} | 1.3×10^{-2} |
| 1.7×10^{-4} | 1.0×10^{-2} | 1.3×10^{-4} | 8.3×10^{-3} | 1.3×10^{-3} | 2.8×10^{-2} |
| 3.6×10^{-4} | 1.3×10^{-2} | 9.9×10^{-5} | 8.4×10^{-3} | 1.4×10^{-3} | 3.5×10^{-2} |
| 3.5×10^{-4} | 1.3×10^{-2} | 9.6×10^{-4} | 2.4×10^{-2} | 1.2×10^{-3} | 2.7×10^{-2} |
| 3.6×10^{-4} | 4.2×10^{-2} | 1.8×10^{-4} | 9.0×10^{-3} | 3.6×10^{-4} | 1.2×10^{-2} |
| 2.9×10^{-4} | 1.6×10^{-2} | 8.3×10^{-5} | 6.4×10^{-3} | 1.1×10^{-4} | 8.7×10^{-3} |

Tabella 4.3. Rete bidimensionale, dati di input 1000×10 , tutti i layers sono convoluzionali.

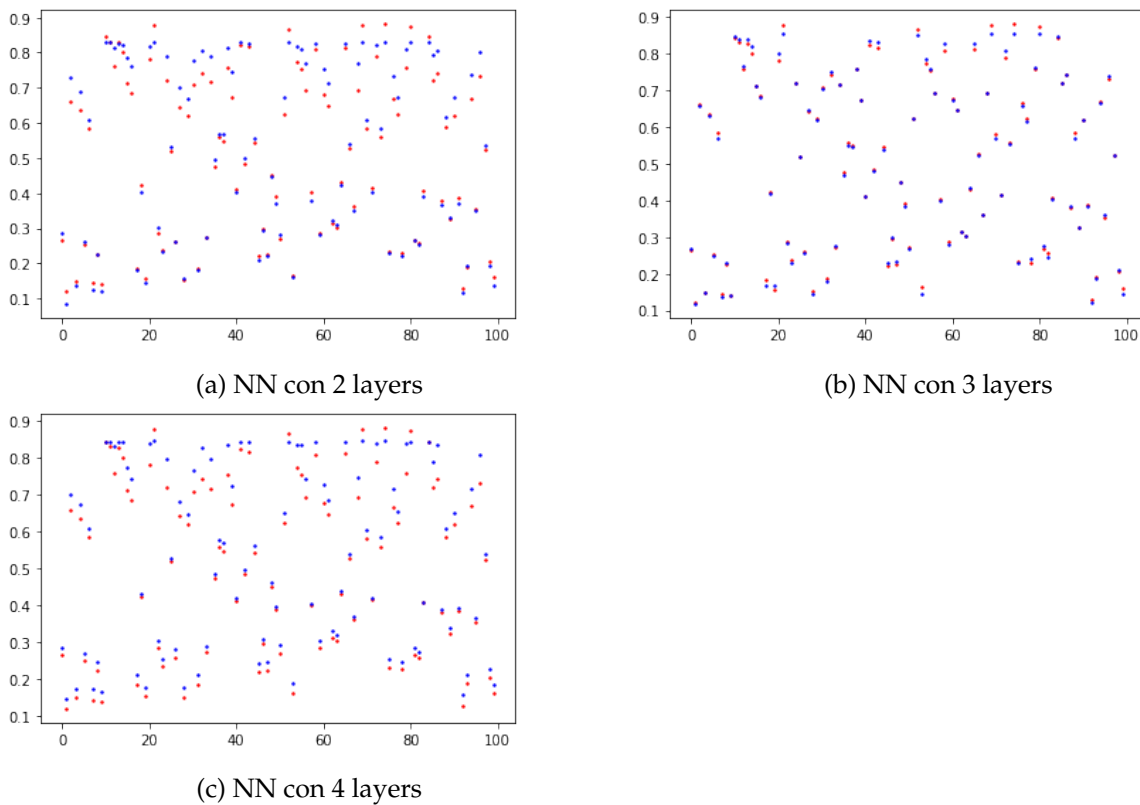


Figura 4.5: Rete bidimensionale, dati di input 50×200 , tutti i layers sono convoluzionali.

Dal confronto delle tabelle 4.2 e 4.3 osserviamo come la sostituzione dei layers classici con quelli di tipo convoluzionale abbia portato un leggero miglioramento nella stima dell'indice di Hurst per le reti con 2 e 3 layers; al contrario, invece, avviene per quelle con 4 livelli, anche se pure in questo caso il cambiamento non risulta così significativo.

Capitolo 5

Conclusioni

Riportiamo, infine, l'errore quadratico medio (MSE) e l'errore assoluto medio (MAE) commessi con il metodo del periodogramma valutato sugli stessi elementi del test set delle reti: $MSE = 9,1 \times 10^{-4}$; $MAE = 2,2 \times 10^{-2}$, in linea con i risultati ottenuti per le reti convoluzionali con un solo strato convoluzionale.

L'utilizzo delle reti permette, quindi, di affinare la precisione delle stime agendo sulle dimensioni dell'input e sul tipo di layers da cui sono costituite; allo stesso tempo, il metodo classico del periodogramma si rivela piuttosto affidabile, anche se sensibile alla scelta della regione del grafico su cui eseguire la regressione lineare.

Si potrebbe, quindi, creare una variante della struttura classica delle rete neurale sfruttando il punto di vista spettrale, ossia generare una rete che prende in input sia il segnale nel tempo che la sua trasformata di Fourier e nella fase di apprendimento sfrutta la precisione del metodo del periodogramma. In questo modo, a parità di efficienza e affidabilità, potremmo ottimizzare il numero di layers e di neuroni per ogni layers della rete, quindi anche il tempo e costo computazionale.

Appendice A

Codici Python

Riportiamo i codici utilizzati per la sperimentazione numerica eseguita sui metodi: varianze aggregate, periodogramma e R/S.

A.1 Metodo delle varianze aggregate

```
1  from fbm import FBM
2  import numpy as np
3  from sklearn.linear_model import LinearRegression
4  import matplotlib.pyplot as plt
5
6  #Genera il campione di rumore
7  N=2**14
8  H=0.9
9  f = FBM(N, H)
10 X = f.fgn()
11
12 #Crea lo stimatore della varianza
13 def var_estim(X, m, N):
14     M = N // m
15     X_m = np.zeros(M)
16     for k in range(M):
17         X_m[k] = np.mean(X[k*m:(k+1)*m])
18     stimatore_m = ((X_m - np.mean(X_m))**2).sum()
19     stimatore_m = stimatore_m / M
20     return stimatore_m
21
22 #Calcola lo stimatore
23 m_piccolo = 10*np.arange(1,250)
24 stimatore = np.zeros(len(m_piccolo))
25 for m in m_piccolo:
26     k = m//10
```

```

27     stimatore[k-1] = var_estim(X, m, N)
28
29     #Plot
30     plt.scatter(np.log10(m_piccolo), np.log10(stimatore), s=2, c='b')
31     plt.xlabel("Log m")
32     plt.ylabel("Log (stimatore varianza)")
33
34     #Calcola la retta di regressione
35     x = np.log10(m_piccolo).reshape(-1,1)
36     y = np.log10(stimatore)
37     retta = LinearRegression()
38     retta.fit(x, y)
39     #Plot retta di regressione
40     m = retta.coef_
41     x0 = retta.intercept_
42     x = np.log10(m_piccolo[:15])
43     plt.plot(x, m*x+x0, c='r')
44     stima = (2 + retta.coef_[0]) / 2
45     print(stima)
46     plt.show()

```

A.2 Metodo del periodogramma

```

1  from fbm import FBM
2  import numpy as np
3  from sklearn.linear_model import LinearRegression
4  import matplotlib.pyplot as plt
5
6  #Genera il campione di rumore
7  N = 2**14
8  H = 0.2
9  f = FBM(N, H)
10 X = f.fgn()
11
12 #Calcola il periodogramma con la formula equivalente
13 X_bar = np.mean(X)
14 I = np.fft.ifft(X-X_bar, norm="ortho") #orto divide per 1/sqrt(N)
15 lambda_k = 2*np.pi*np.arange(0,N)/N
16 I_module = np.abs(I)
17 I_sq = np.power(I_module,2)
18
19 #Plot
20 plt.scatter(np.log10(lambda_k), np.log10(I_sq), s=2, c='k')
21 plt.xlabel("Log ")

```

```

22 plt.ylabel("Log I()")
23
24 #Calcola la retta di regressione lineare sul 10% delle frequenze più basse
25 M=N//10
26 x = np.log10(lambda_k[1:M]).reshape(-1,1)
27 y = np.log10(I_sq[1:M])
28 retta = LinearRegression()
29 retta.fit(x, y)
30 #Plot retta di regressione
31 m = retta.coef_
32 x0 = retta.intercept_
33 x = np.log10(lambda_k[1:M])
34 plt.plot(x[1:],m*x[1:]+x0, c='green')
35 stima = (1 - retta.coef_[0])/2
36 print(stima)
37 plt.show()

```

A.3 Metodo R/S

```

1 from fbm import FBM
2 import numpy as np
3 from sklearn.linear_model import LinearRegression
4 import matplotlib.pyplot as plt
5
6 #Genera il campione di rumore e stabilisce il range
7 N = 2**14
8 H = 0.2
9 K = 3
10 f = FBM(N, H)
11 X = f.fgn()
12
13 #Crea la statistica R/S
14 def aux(X, N, K, r):
15     M = N // K
16     t = np.arange(K) * M
17     R = np.zeros(K)
18     S_square = R.copy()
19     for i in np.arange(K):
20         W = np.zeros(r)
21         if t[i] + r <= N :
22             for k in range(r):
23                 W[k] = X[t[i]:t[i] + k].sum() - k * np.mean(X[t[i]: t[i]+r])
24             R[i] = max(W) - min(W)
25             S_square[i] = np.mean((X[t[i]: t[i]+r])**2) - np.mean(X[t[i]: t[i]+r])**2

```

```
26     S = np.sqrt(S_square)
27     else:
28         return np.mean(R[:i]/S[:i])
29     return np.mean(R/S)
30
31     #Calcola la statistica R/S
32     ranges = 10 + 50*np.arange(198)
33     statistica = np.zeros(len(ranges))
34     for r in ranges:
35         i = (r-10)//50
36         statistica[i] = aux(X, N, K, r)
37
38     #Plot
39     plt.scatter(np.log10(ranges),np.log10(statistica), s=2, c='b')
40     plt.xlabel("Log r")
41     plt.ylabel("Log (R/S)")
42
43     #Calcola la retta di regressione
44     x = np.log10(ranges).reshape(-1,1)
45     y = np.log10(statistica)
46     retta = LinearRegression()
47     retta.fit(x, y)
48     #Plot retta di regressione
49     m = retta.coef_
50     x0 = retta.intercept_
51     x = np.log10(ranges)
52     plt.plot(x, m*x+x0, c='g')
53     stimea= retta.coef_[0]
54     print(stima)
55     plt.show()
```


A.4 Rete neurale convuluzionale 1-dimensionale

```
1 import random
2 from fbm import FBM
3 import numpy as np
4
5 # Genera il dataset
6 n_samples = 2000
7 length_samples = 10000
8 X = np.zeros((n_samples, length_samples))
9 Y = np.zeros(n_samples)
10 for i in range(0, n_samples):
11     H = random.uniform(0.1, 0.9)
12     f = FBM(length_samples, H)
13     X[i, :] = f.fgn()
14     Y[i] = H
15
16 # Split in dataset di training e di test
17 from sklearn.model_selection import train_test_split
18 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.33, random_state=42)
19
20 # Costruzione della rete
21 from keras.models import Sequential
22 from keras.layers import Dense
23 from keras.layers import Flatten
24 from keras.layers.convolutional import Conv1D
25
26 epochs, batch_size = 50, 32
27 n_timesteps = X_train.shape[1]
28 model = Sequential()
29 # Aggiunge uno strato convoluzionale
30 model.add(Conv1D(filters=32, kernel_size=3, activation='relu', input_shape=(n_timesteps, 1)))
31 model.add(Flatten()) #allinea l'output
32 # Aggiunge uno strato di neuroni classici
33 model.add(Dense(10, activation='relu'))
34 model.add(Dense(1, activation='sigmoid'))
35 model.compile(optimizer='rmsprop', loss='mse', metrics=['mse', 'mae'])
36
37 # Training della rete
38 model.fit(X_train, Y_train, epochs=epochs, batch_size=batch_size)
39
40 # Test della rete
41 from sklearn.metrics import mean_squared_error, mean_absolute_error
42 Y_pred = model.predict(X_test)
43 print("MSE sul dataset di test: ", mean_squared_error(Y_pred, Y_test))
44 print("MAE sul dataset di test: ", mean_absolute_error(Y_pred, Y_test))
```

A.5 Rete neurale convuluzionale bidimensionale

```

1 X_res=X.reshape(2000,50,200)
2 # Split in dataset di training e di test
3 from sklearn.model_selection import train_test_split
4 X_train, X_test, Y_train, Y_test = train_test_split(X_res, Y, test_size=0.33, random_state=42)
5
6 from keras.models import Sequential
7 from keras.layers import Dense
8 from keras.layers import Flatten
9 from keras.layers.convolutional import Conv1D, Conv2D
10 from keras.layers.convolutional import MaxPooling1D, MaxPooling2D
11
12 # Costruzione della rete
13 epochs, batch_size = 50, 32
14 n_timesteps = X_train.shape[1]
15 model = Sequential()
16
17 # Aggiunge uno strato convoluzionale
18 model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(50, 200, 1)))
19 model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(50, 200, 1)))
20 # Aggiunge uno strato di pooling
21 model.add(MaxPooling2D((2, 2)))
22 model.add(Flatten())
23 # Aggiunge uno strato di neuroni classici
24 model.add(Dense(10, activation='relu'))
25 model.add(Dense(1, activation='sigmoid'))
26 model.compile(optimizer='rmsprop', loss='mse', metrics=['mse', 'mae'])
27
28 # Training della rete
29 model.fit(X_train, Y_train, epochs=epochs, batch_size=batch_size)
30
31 # Test della rete
32 from sklearn.metrics import mean_squared_error, mean_absolute_error
33 Y_pred=model.predict(X_test)
34 print("MSE sul dataset di test: ", mean_squared_error(Y_pred, Y_test))
35 print("MAE sul dataset di test: ", mean_absolute_error(Y_pred, Y_test))

```

Appendice B

Codici Matlab

Riportiamo i codici del metodo di Davies e Harte per la simulazione del rumore gaussiano frazionario e di come, a partire dal rumore, generare il moto Browniano frazionario corrispondente. Per testare i codici abbiamo utilizzato MATLAB[®] R2020b su Mac.

B.1 Algoritmo di Davies e Harte

Sia X il vettore di rumore gaussiano che vogliamo simulare e N la taglia richiesta per X . Dalla teoria sappiamo che $X = (X_1, \dots, X_N)$ è un vettore centrato e con matrice di covarianza

$$\Gamma_n = \begin{pmatrix} 1 & \gamma(1) & \gamma(2) & \dots & \gamma(n-2) & \gamma(n-1) \\ \gamma(1) & 1 & \gamma(1) & \dots & \gamma(n-3) & \gamma(n-2) \\ \gamma(2) & \gamma(1) & 1 & \dots & \gamma(n-4) & \gamma(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(n-2) & \gamma(n-3) & \gamma(n-4) & \dots & 1 & \gamma(1) \\ \gamma(n-1) & \gamma(n-2) & \gamma(n-3) & \dots & \gamma(1) & 1 \end{pmatrix},$$

quindi X può essere rappresentato come $X = SV$ dove $V = (V_1, \dots, V_N)$ è un vettore Guassiano standard e S una matrice tale che $SS^T = \Gamma_n$, ovvero per simulare X dobbiamo trovare una "radice quadrata" della matrice Γ_n .

Il metodo di Davies e Harte prevede di immergere Γ_n in una matrice circolante C e ricavare la "radice quadrata" di Γ_n dalla decomposizione spettrale di C . Si verifica [1], infatti, che data la decomposizione $C = Q\Lambda Q^H$, allora la matrice S che cerchiamo è esattamente $S = Q\Lambda^{\frac{1}{2}}Q^H$. Dunque, per ottenere X dobbiamo svolgere il prodotto $Q\Lambda^{\frac{1}{2}}Q^H V$.

```

1 function X=daviesharte(N,H)
2 % la funzione prende in input:
3 % H: parametro di Hurst
4 % N: N+1 taglia finale del campione
5 % e simula le "prime" N+1 variabili di un Fractional Gaussian
   Noise process: X={X(k) : k=0,1,2,...}
6
7 %passo 1
8 for k=1:N
9     alpha(k)= noise_covariance(k-1,H);
10    alpha(2*N-k+1)= noise_covariance(k,H);
11 end
12 lambda= fft(alpha);
13 lambda= real(lambda);
14
15 %passo 2
16 W(1)= randn(1);
17 W(N+1)= randn(1);
18 for i=2:N
19     V1(i)= randn(1);
20     V2(i)= randn(1);
21     W(i)= (V1(i) + sqrt(-1)*V2(i))/sqrt(2);
22     W(2*N-i+2)= (V1(i) - sqrt(-1)*V2(i))/sqrt(2);
23 end
24
25 %passo 3
26 w(1)= sqrt(lambda(1)/(2*N))*randn(1);
27 w(N+1)= sqrt(lambda(1)/(2*N))*randn(1);
28 for i=2:N
29     w(i)= sqrt(lambda(i)/(4*N))*(V1(i) + sqrt(-1)*V2(i));
30     w(2*N-i+2)= sqrt(lambda(2*N-i+2)/(4*N))*(V1(i) - sqrt(-1)*
       V2(i));
31 end
32 Z=fft(w);
33 Z=real(Z);
34 X=Z(1:N);
35 end

```

Bibliografia

- [1] Ton Dieker, *Simulation of fractional Brownian motion*, 2004.
- [2] Katarzyna Filus, Adam Domanski, Joanna Doman ska, Dariusz Marek, Jakub Szyguła, *Long-Range Dependent Traffic Classification with Convolutional Neural Networks Based on Hurst Exponent Analysis*, 2020.
- [3] Jean-Francois Coeurjolly, *Simulation of fractional Brownian motion*, 2001.
- [4] Romain François Peltier, Jacques Lévy Véhel, *A new method for estimating the parameter*, 2006.
- [5] Francesco Caporali, *Reti neurali profonde: capacità di approssimazione e convergenza a processi gaussiani*, 2021.
- [6] Timothy Graves, Robert Gramacy, Nicholas Watkins, Christian Franzke, *A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951–1980*, 2017.
- [7] Gabriella Ida Ghiara, *Approssimazioni del moto browniano frazionario e passeggiate aleatorie*, 2010.
- [8] Jan Beran, *Statistics for Long-Memory Processes*, 1994.
- [9] Andrea Basteri, Dario Trevisan, *Quantitative gaussian approximation of randomly initialized deep neural networks*.
- [10] M.B. Priestley, *Spectral analysis and time series*, 1981.
- [11] W. Rea, L. Oxley, M. Reale, J. Brown, *Not all estimators are born equal: The empirical properties of some estimators of long memory*, 2013.