

HOMework 1

ALICE SCHIRINÀ

5 giugno 2018

Sono dati due file di testo, *data1.txt* e *data2.txt*, ciascuno dei quali organizzato in tre colonne di dati (per esempio T , U_1 e U_2) che rappresentano rispettivamente il tempo di misura e le misure delle osservabili U_i ($i = 1, 2, 3, 4$).

1 Analisi a blocchi

Supponiamo in un primo momento che i dati siano indipendenti e calcoliamone la media e l'errore:

$$\langle U_i \rangle_{ind} = \frac{1}{N} \sum_i^N U_i$$
$$\sigma_{i,ind} = \sqrt{\langle U_i^2 \rangle - \langle U_i \rangle^2}$$

dove N rappresenta il numero complessivo di dati per osservabile, in questo caso $N = 100000$. In questo modo otteniamo i seguenti risultati:

$\langle U_i \rangle_{ind}$	$\sigma_{i,ind}$
2.135503	0.001643
1.164563	0.000434
-0.069248	0.001884
2.031675	0.002164

Vogliamo ora eseguire un'analisi a blocchi per tenere in considerazione la correlazione tra i dati. Generiamo, quindi, i blocchi di dati definendo la relazione per ricorrenza:

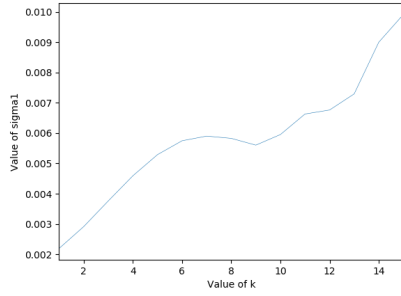
$$U^{(1)}(t) = \frac{1}{2}[U_i(2t-1) + U_i(2t)]$$
$$U_i^{(k)}(t) = \frac{1}{2}[U_i^{(k-1)}(2t-1) + U_i^{(k-1)}(2t)]$$

cosicché a ogni iterazione k il numero dei dati si dimezza ($N/2$, $N/4$, etc).

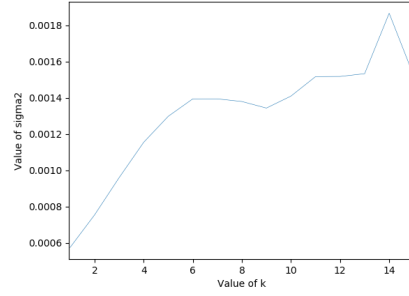
Calcoliamo nuovamente la media usuale e l'errore sui dati come segue:

$$\sigma_i(k) = \sqrt{\frac{\langle U_i^2 \rangle_k - \langle U_i \rangle_k^2}{N-1}} \quad (1)$$

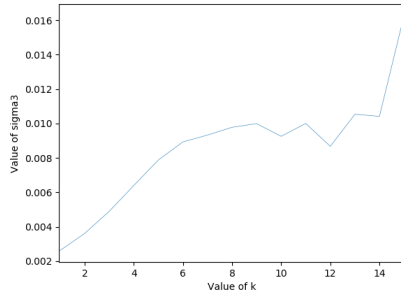
in cui la media ed N sono rispettivamente la media e il numero di dati nel blocco. Mostriamo in figura gli andamenti degli errori in funzione di k .



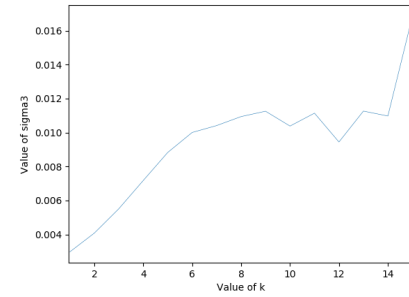
(a) Andamento σ_1 in funzione di k



(b) Andamento σ_2 in funzione di k



(c) Andamento σ_3 in funzione di k



(d) Andamento σ_4 in funzione di k

Come possiamo osservare l'errore si stabilizza intorno a $k = 6$. Scegliamo, quindi, questo valore di k per ottenere una stima dell'errore per ognuna delle osservabili.

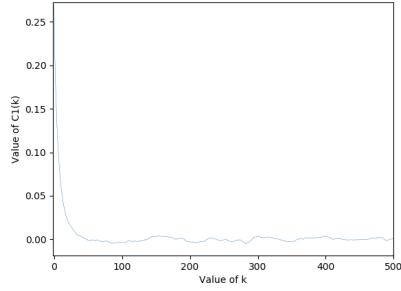
Osservabile	Stima di σ_i
U_1	0.005901
U_2	0.001394
U_3	0.009327
U_4	0.010939

2 Funzione di autocorrelazione

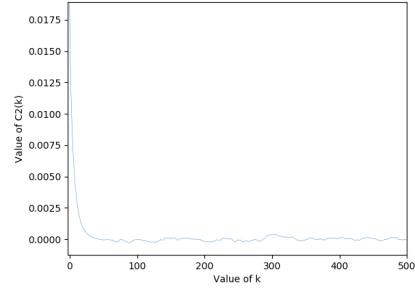
Calcoliamo ora la funzione di autocorrelazione per ognuna delle variabili:

$$C(k) = \frac{1}{N-k} \sum_{j=1}^{N-k} (U(j+k) - \bar{U})(U(j) - \bar{U}) \quad (2)$$

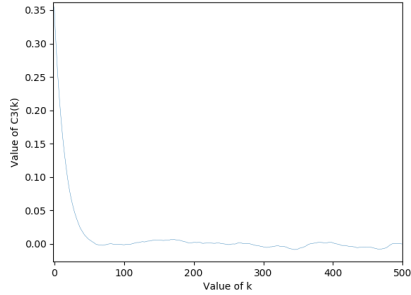
Abbiamo simulato l'andamento della funzione fino a $k = 500$.



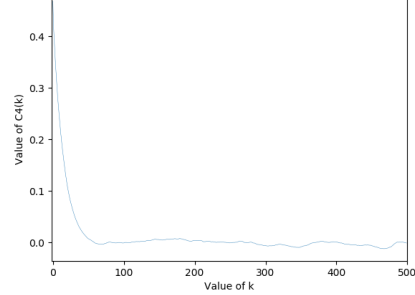
(e) Andamento $C_1(k)$ in funzione di k



(f) Andamento $C_2(k)$ in funzione di k



(g) Andamento $C_3(k)$ in funzione di k



(h) Andamento $C_4(k)$ in funzione di k

Come possiamo osservare, a partire da un certo k nell'intervallo tra 40 e 60, la curva tocca lo zero.

Calcoliamo poi il corrispondente τ_{int} secondo:

$$\tau_{int} = \frac{1}{2} + \sum_{k=1}^{k_{max}} \frac{C(k)}{C(0)} \quad (3)$$

dove per k_{max} abbiamo considerato l'ultimo k per cui la funzione $C(k)$ è ancora positiva.

Osservabile	k_{max}	τ_{int}
U_1	47	7.069802
U_2	43	5.712427
U_3	59	14.166401
U_4	58	13.406710

Utilizziamo i tempi di autocorrelazione per stimare gli errori sul campione medio di U_i :

$$\sigma = \sqrt{\frac{C(0)}{N}} 2\tau_{int} \quad (4)$$

e riportiamo in tabella il confronto tra gli errori appena ottenuti e quelli della sezione precedente per ogni variabile.

Osservabile	$\sigma(6)$	$\sigma_i(\tau_{int})$
U_1	0.005901	0.006177
U_2	0.001394	0.001466
U_3	0.009327	0.010029
U_4	0.010939	0.011206

3 Metodo Jackknife

Nell'analisi a blocchi ogni blocco contiene $N/2^k$ misure (con k numero dell'iterazione) e l'errore si stabilizza intorno a $k = 7$. Quindi, partendo dagli $N = 100000$ dati iniziali, dovremmo raggruppare 2^k misure con dimensione massima per blocco pari a 512 per avere blocchi di dimensione $N/2^k$. Riorganizziamo allora i dati in blocchi di 2000 (circa quattro volte 512), potendo assumere che essi siano scorrelati.

Su questi nuovi dati calcoliamo la variabile rapporto:

$$R_i = \frac{\langle U_i \rangle}{\langle U_1 \rangle} \quad (5)$$

dove $i = 2, 3, 4$. Quindi calcoliamo l'errore come segue:

$$\sigma_R = \left[\frac{N-1}{N} \sum_{k=0}^{N-1} (\hat{R}_{i,k} - R_i) \right]^{\frac{1}{2}} \quad (6)$$

dove R_i è il rapporto delle medie aritmetiche delle medie Jackknife, in formule:

$$\begin{aligned} \hat{R}_{ave} &= \frac{1}{N} \sum_{k=1}^N \hat{U}_k \\ R_i &= \frac{\hat{R}_{ave,i}}{\hat{R}_{ave,1}} \end{aligned}$$

e, infine, le medie Jackknife sono:

$$\hat{U}_i = \frac{1}{N-1} \sum_{\substack{k=1 \\ k \neq i}}^N U_k \quad (7)$$

Confrontiamo l'errore così ottenuto con quelli calcolati con l'*independent-error formula* e la *worst-error formula*:

$$\sigma_{IEF,R_i} = \sqrt{R_i^2 \left(\frac{\sigma_1^2}{\langle U_1 \rangle^2} + \frac{\sigma_i^2}{\langle U_i \rangle^2} \right)} \quad (8)$$

$$\sigma_{WEF,R_i} = \sqrt{R_i^2 \left(\frac{\sigma_1}{|\langle U_1 \rangle|} + \frac{\sigma_i}{|\langle U_i \rangle|} \right)^2} \quad (9)$$

Da queste formule otteniamo i seguenti valori per gli errori sulle R :

R_i	σ_{JK}	σ_{IEF}	σ_{WEF}
0.545334	0.001012	0.001720	0.002264
-0.032428	0.004321	0.004698	0.004790
0.951379	0.006699	0.005925	0.007999

Come ci aspettavamo gli errori ottenuti con la *worst-error formula* sono una sovrastima dell'errore ottenuto con il metodo Jackknife. Anche l'errore con l'*independent-error formula* in questo caso risulta essere una sovrastima.