Relazione d’esame
Computational Statistics

Cecci Riccardo
Matr: 20023915
UPO, Laurea Magistrale in Informatica

Esercizio 1

Consegna

Calcolo dei momenti e verifica relazioni tra centrati e non centrati

Sviluppo

Avendo a disposizione un elenco di valori discreti, stiamo trattando una variabile discreta; quindi, i momenti possono essere calcolati tramite le seguenti formule:

1. *Momento semplice di ordine n:*

$$m_n = \frac{1}{k}\sum_{i=0}^{k} x_i^n$$

   Con K = numero dei dati

   Da questa formula possiamo derivare che il momento di ordine 1 è pari alla media (somma degli elementi / numero degli elementi)

2. *Momento centrato di ordine n:*

$$\mu_n = \frac{1}{k}\sum_{i=0}^{k} \left(x_i - \eta\right)^n$$

   Con K = numero dei dati

   Il momento centrato del secondo ordine rappresenta la varianza

Tra momento semplice e momento centrato esiste una relazione, definita dalle seguenti formule:

- *Momento centrato da momento semplice*

$$\mu_n = \sum_{k=0}^{n} \binom{n}{k} m_k (-\eta)^{n-k}$$ con η la media e m_k momento semplice di ordine k.

- *Momento semplice da momento centrato*

$$m_n = \sum_{k=0}^{n} \binom{n}{k} \mu_k \eta^{n-k}$$ in cui η è la media e μ_k è il momento centrato di ordine k.

Risultati ottenuti:

Per mettere in evidenza le relazioni fra i momenti, si sono applicate le formule sopra descritte
Si è preso in considerazione il giorno 10 settembre e le temperature di quella giornata e i risultati ottenuti son stati i seguenti:

- *Momento semplice di ordine 1:*
  $m_1 = 24.191$
  Lo si è confrontato, essendo il momento di ordine 1 corrispondente alla media, con la funzione *mean* di r, ottenendo lo stesso valore
- *Momento semplice di ordine 1 da momento centrato:*
  $m_1' = 24.191$
  Si nota la corrispondenza fra questo valore e quello calcolato tramite il metodo del calcolo "diretto" del momento semplice
- *Momento centrato di ordine 2:*
  $\mu_2 = 22.635$
- *Momento centrato di ordine 2 da momento semplice:*
  $\mu_2' = 22.635$
  Anche in questo caso si nota la corrispondenza fra questo valore e quello calcolato tramite il metodo del calcolo "diretto" del momento centrato
- Essendo il momento semplice di ordine 1 e il momento centrato di ordine 2 corrispondenti rispettivamente alla media e alla varianza dei dati, si sono potuti confrontare i valori ottenuti con le funzioni *mean()* e *var()* di R; le quali hanno riportato gli stessi valori

Allo stesso modo si possono definire momenti, semplici e centrati, di ordine k

Esercizio 2

Consegna

Calcolo del giorno tipo per la temperatura: media su un mese per ogni minuto del giorno.

Sviluppo

Dato che si richiede la “giornata tipo” e l’andamento della temperatura per tale giornata, si è deciso di lavorare sulle temperature dello stesso minuto di ogni giorno e farne la media. Ripetendo questo procedimento ad ogni minuto della giornata per ciascun giorno del mese.

Risultati ottenuti:

Per una migliore interpretazione dei dati ho generato un grafico della “giornata tipo” tramite la funzione *plot*, ottenendo il seguente risultato, rappresentato in Figura 1:

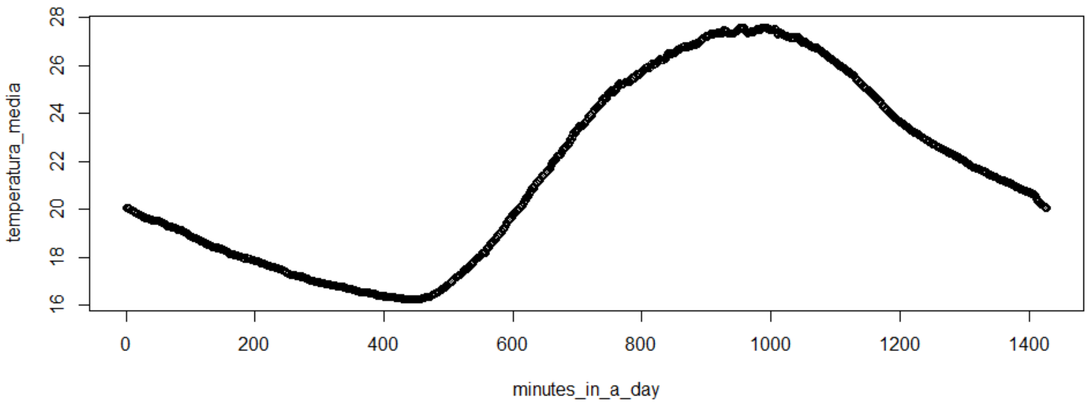

Figura 1 Giornata tipo per la temperatura

Si nota che il grafico segue perfettamente quello che ci si aspetta possa essere l'andamento delle temperature in una tipica giornata di settembre.

Temperature "basse" nelle ore teoricamente più fredde (durante la notte e di prima mattina) seguite da un graduale incremento man mano che ci si avvicina alle ore teoricamente più calde (comprese tra la tarda mattinata e il pomeriggio) durante le quali si verifica il picco massimo, per poi avere una decrescita man mano che ci si avvicina alla notte.

Esercizio 3

Consegna

Verifica grafica della PDF delle fluttuazioni di velocità per diversi periodi temporali.

Sviluppo

Le fluttuazioni indicano quanto si discosta un dato dalla media, pertanto per calcolarle è stato sufficiente, in un primo momento, raccogliere i dati sulle velocità; successivamente, si è calcolata la media di tali valori che infine è stata sottratta dai valori raccolti precedentemente.

Risultati ottenuti

Si sono presi in considerazione le misurazioni di:

- Giorno 5 e giorno 20
- Mese
- Ore diurne (6 – 20) e notturne (20 – 6) del giorno 5

La Figura 2 mostra le PDF delle fluttuazioni della velocità del vento nei giorni 5 e 20.

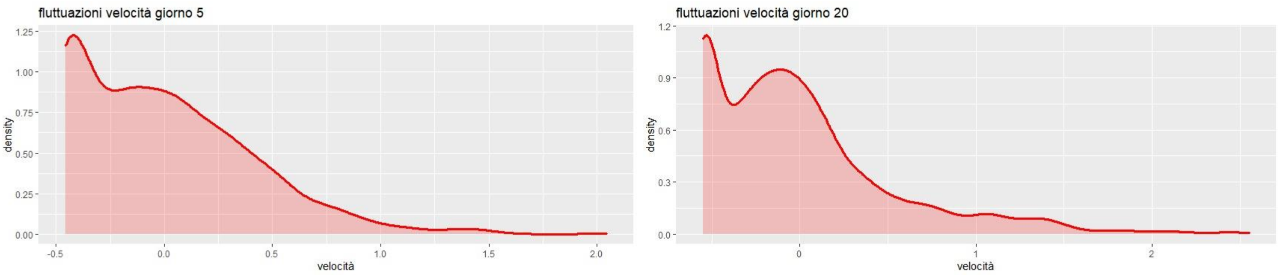

Figura 2 PDF delle fluttuazioni delle velocità nel giorno 5 (sinistra) e nel giorno 20 (destra)

La Figura 3 mostra la PDF delle fluttuazioni delle velocità del vento nell'arco di tutto il mese

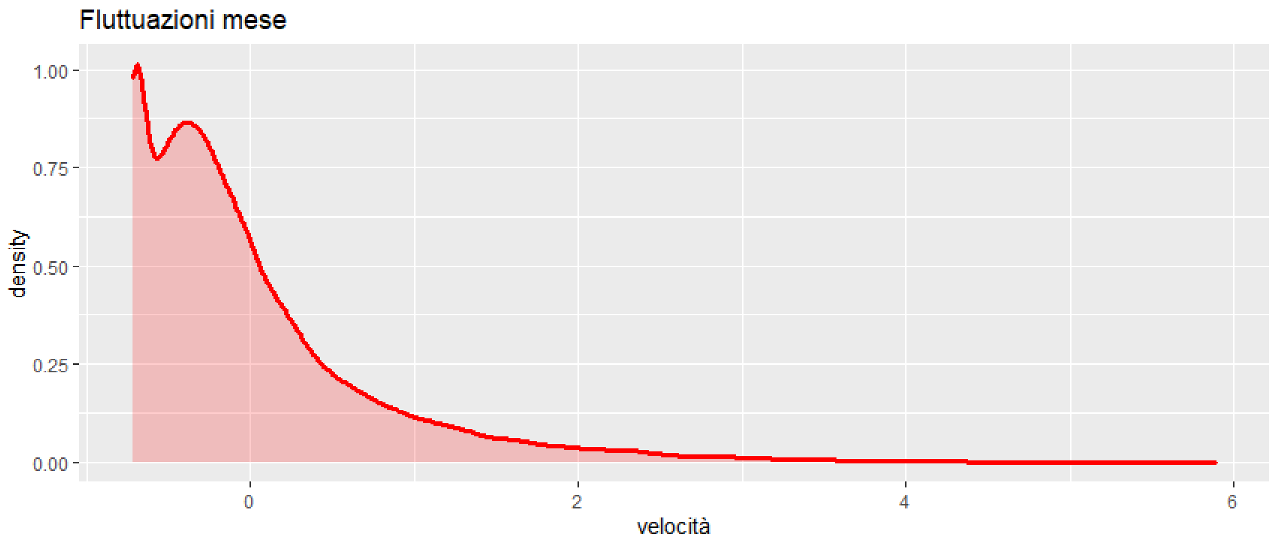

Figura 3 PDF delle fluttuazioni delle velocità in tutto il mese

La figura 4 mostra le PDF delle fluttuazioni di velocità del vento nelle ore diurne e notturne del giorno 5

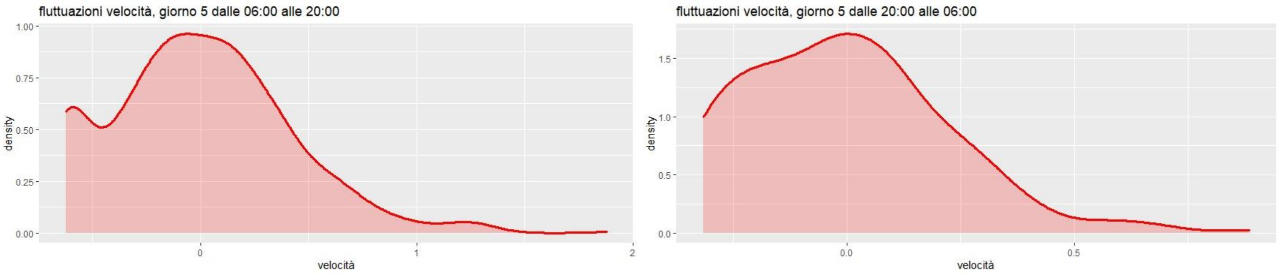

Figura 4 PDF delle fluttuazioni di velocità del giorno 5 nelle ore diurne (sinistra) e notturne (destra)

Esercizio 4

Consegna

Calcolo della PDF delle fluttuazioni di velocità verifica gaussianità tramite test chi-quadro sui seguenti intervalli temporali giorno, mese, e suddivisi giorno-notte

Sviluppo

Per verificare che le fluttuazioni di velocità del vento abbiano un andamento gaussiano è stato implementato il test chi-quadro. Tale test, a fronte di un confronto fra il valore calcolato di χ^2 e il valore *d* dei gradi di libertà, permette di accettare o rifiutare l'ipotesi
Per rigettare l'ipotesi dovrà verificarsi che

$$\chi^2 >> d$$

Dove $d = n - c$, in cui n è il numero totale di dati misurati e c è il numero di parametri determinati dai dati.
Dato che in questo esercizio tratta uno studio sulla densità gaussiana c sarà pari a 3, ovvero:

- Calcolo della standardizzazione
- Calcolo del valor medio
- Calcolo della varianza

La formula per il test chi-quadro è la seguente:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

dove:

- o_i rappresenta la frequenza osservata
- e_i = *n * (pnorm(estremo_superiore) – pnorm(estremo_inferiore))* e rappresenta la frequenza teorica attesa

Per standardizzare i dati si è usata la seguente formula

$$Z = \frac{X - \eta}{\sigma}$$

dove η è la media dei valori e σ è la deviazione standard

La frequenza teorica attesa è associata alla probabilità che ha un valore di ricadere in un determinato intervallo di probabilità. Gli intervalli dovranno essere in numero almeno pari a 5, tale numero dipenderà dall'ampiezza *a* degli stessi. Ogni intervallo rappresenta quante volte la variabile osservata assume un valore che si trova all'interno dell'intervallo corrispondente.
La quantità di intervalli viene calcolata nel seguente modo:

$$\left\lceil \frac{max - min}{a} \right\rceil$$

Dove *max* e *min* corrispondono rispettivamente al massimo e al minimo valore dei dati.
Gli estremi superiore e inferiore dell'intervallo sono definiti nel modo seguente:

- *estremo_inferiore = min + a(i – 1)*
- *estremo_superiore = min + ai*

con *i* che corrisponde all'indice dell'intervallo considerato

Risultati ottenuti

La Figura 5 mostra le PDF delle fluttuazioni standardizzate della velocità del vento calcolate nei giorni 5 e 20.

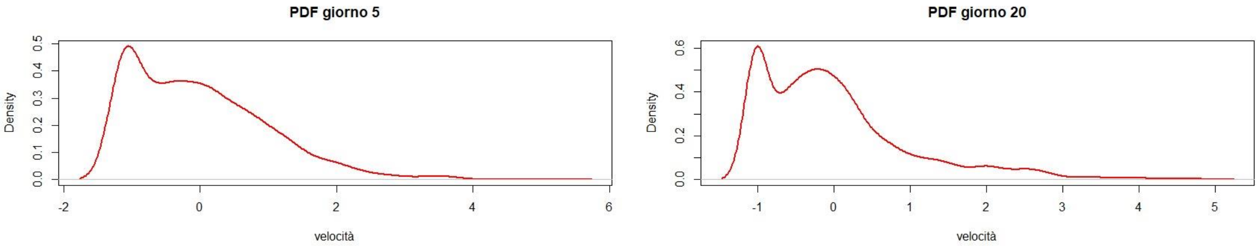

Figura 5 PDF delle fluttuazioni standardizzate delle velocità nei giorni 5 (sinistra) e 20 (destra)

Con a = 1.2 e 6 intervalli sono stati ricavati i seguenti valori per il χ^2 dei due giorni:

giorno 5: $\chi^2 = 1323.4892$

giorno 20: $\chi^2 = 364.2026$

dato che per entrambi vale $\chi^2 >> d$ si rigetta l'ipotesi di una distribuzione gaussiana

La Figura 6 mostra le PDF delle fluttuazioni standardizzate delle velocità del vento del giorno 5 in orario diurni (dalle 6 alle 20) e in orari notturni (dalle 20 alle 6)

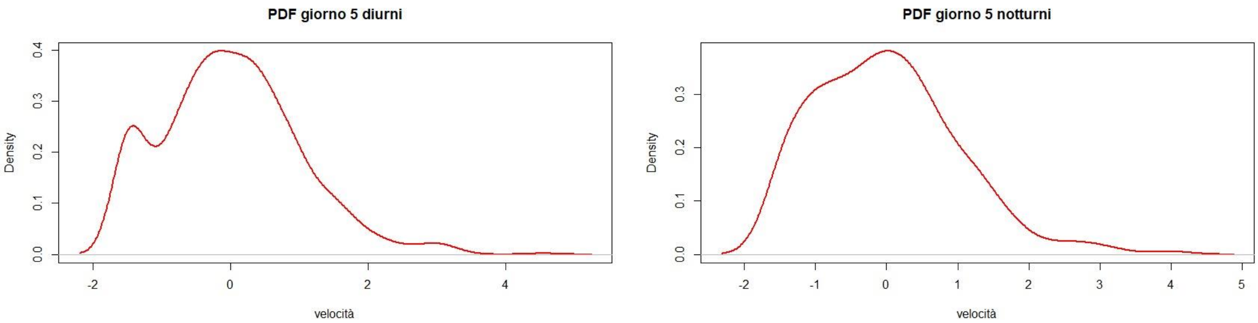

Figura 6 PDF delle fluttuazioni delle velocità nel giorno 5 in orari diurni (sinistra) e notturni (destra)

Con a = 1.2 per gli orari diurni e 6 intervalli è stato ricavato il seguente valore per χ^2:

$$\chi^2 = 364.2026$$

Con a = 0.8 per gli orari notturni e 7 intervalli è stato ricavato il seguente valore per χ^2:

$$\chi^2 = 36.3709$$

In entrambi i casi si nota che $\chi^2 >> d$ e di conseguenza si rigetta l'ipotesi di una distribuzione gaussiana

La Figura 7 mostra la PDF delle fluttuazioni standardizzate delle velocità del vento nell'arco di tutto il mese

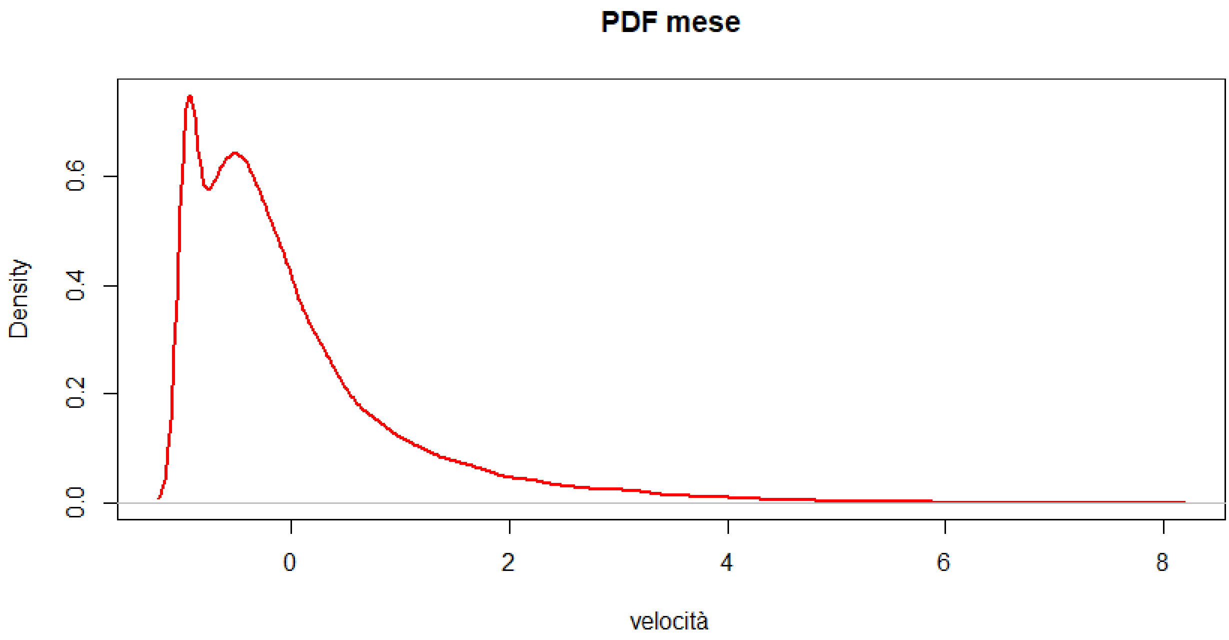

Figura 7 PDF delle fluttuazioni della velocità in tutto il mese

Con a = 1.5 e 6 intervalli è stato ricavato il seguente valore di χ^2:

$$\chi^2 = 76135900.4299$$

Dato che anche in questo caso $\chi^2 >> d$, si rigetta l'ipotesi di una distribuzione gaussiana per il mese

Esercizio 5

Consegna

Calcolo del tempo Lagrangiano delle fluttuazioni della velocità del vento tramite regressione esponenziale della funzione di autocorrelazione temporale

Sviluppo

Per calcolare il tempo Lagrangiano delle fluttuazioni della velocità del vento è necessario calcolare l'autocorrelazione temporale per ogni valore di τ, il che viene fatto attraverso la seguente formula:

$$R(\tau) = \frac{\overline{u(t) \times u(t+\tau)}}{\sigma^2}$$

Dove:

- *u(t)* è il valore della fluttuazione all'istante t
- σ^2 è la varianza

Usando questa formula su ogni valore di τ si ottiene un vettore di autocorrelazione temporale delle fluttuazioni della velocità del vento.

La regressione esponenziale è definita come segue:

$$y = A + e^{Bx}$$

Ma, ponendo *z = ln(y)* si possono calcolare i due coefficienti A e B nella forma della regressione lineare nel seguente modo:

$$ln(y) = ln(A) + ln(e^{Bx}) = ln(A) + Bx$$

Il tempo Lagrangiani *Lt* dipende dal coefficiente di autocorrelazione temporale:

$$r = e^{-\frac{\tau}{Lt}}$$

Da cui si ricava:

$$ln(r) = ln(e^{-\frac{\tau}{Lt}}) = -\frac{\tau}{Lt}$$

E di conseguenza:

$$Lt = -\frac{\tau}{ln(r)}$$

Per la regressione lineare, si ottiene che il tempo Lagrangiano dipende dal coefficiente B. Tale dipendenza è definita dalla seguente relazione:

$$Lt = -\frac{1}{B}$$

Dove B viene calcolato applicando il logaritmo naturale e utilizzando la funzione *line()*, in questo modo si ottiene:

$$Lt = -\frac{1}{line(ln(r))}$$

Dove *r* è il vettore di autocorrelazione temporale. Dalla funzione *line()* si ricava una retta passante per i punti con coefficiente angolare pari a:

$$m = line(ln(r))$$

In R si può calcolare il tempo Lagrangiano utilizzando la funzione *acf*, ciò ha permesso un confronto più efficace ed immediato con i valori calcolati tramite le funzioni definite nello script.

Risultati ottenuti

La Figura 8 mostra il confronto fra il vettore di autocorrelazione temporale calcolato tramite funzione implementate *ad hoc* (linea rossa) con i valori ottenuti dalla funzione *acf* di R (linee verticali) relativamente al giorno 5.

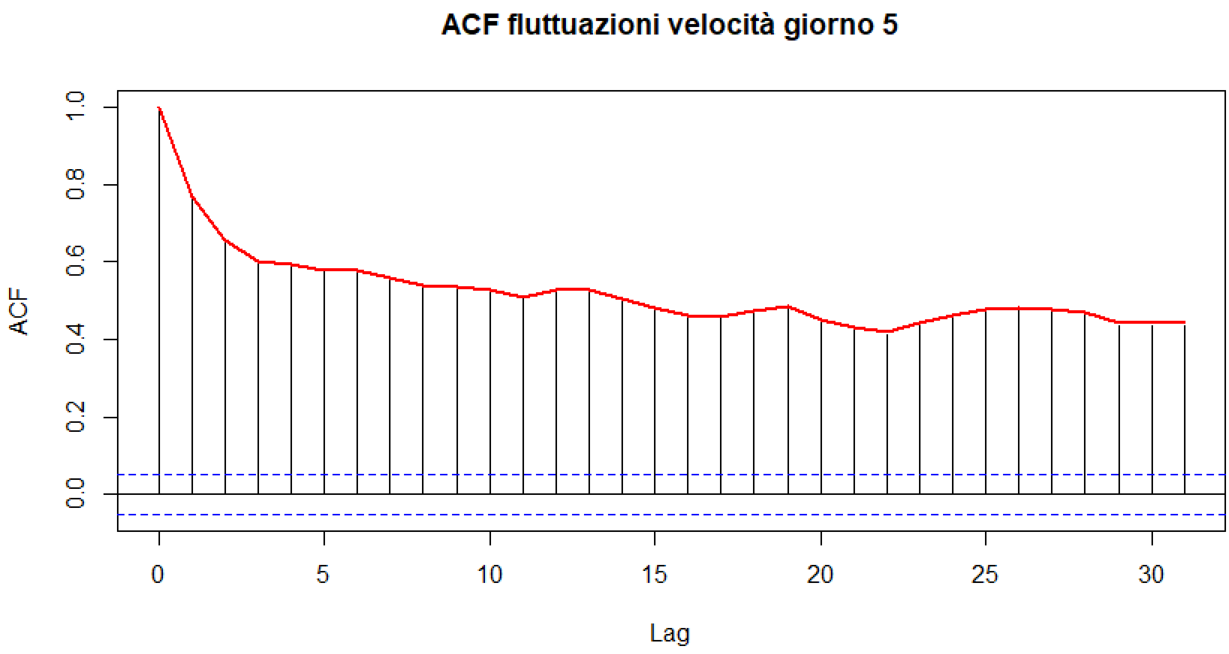

Figura 8 Confronto ACF delle fluttuazioni di velocità del giorno 5

A testimonianza della similarità dei risultati ottenuti per il giorno 5 si riportano anche i valori del tempo Lagrangiano ricavati tramite i due metodi già citati:

- $Lt = 0.5077$
- $acf = 0.52042$

La Figura 9 mostra il confronto fra il vettore di autocorrelazione temporale calcolato tramite funzione implementatate *ad hoc* (linea rossa) con i valori ottenuti dalla funzione *acf* di R (linee verticali) relativamente al giorno 20.

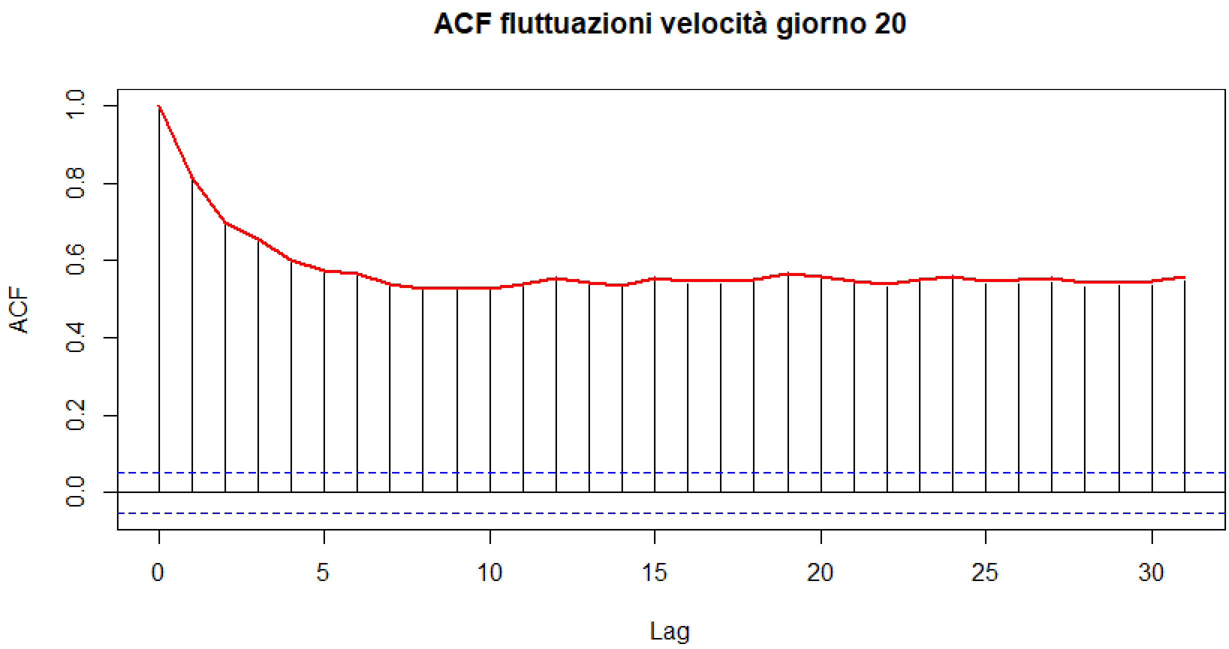

Figura 9 Confronto ACF delle fluttuazioni di velocità del giorno 20

Anche in questo caso, a testimonianza della similarità dei risultati ottenuti per il giorno 20 si riportano anche i valori del tempo Lagrangiano ricavati tramite i due metodi già citati:

- *Lt = 0.8194*
- *acf = 0.5749*

I due casi presi in esame, e i confronti attuati in ciascuno di essi, mostrano che i due metodi di calcolo sono in grado di ricavare soluzioni molto simili.

Esercizio 6

Consegna

Realizzazione del Modello di Langevin per il caso unidimensionale omogeneo e PDF gaussiana. Si disegnino le traiettorie in un piano (t,x). Si calcolino le statistiche delle velocità ottenute e si confrontino con il valor medio (nullo) e la deviazione standard usata in input.

Sviluppo

Il modello di Langevin è stato simulato attraverso un processo Markoviano. Un processo Markoviano è un processo stocastico in cui la probabilità di transizione ad uno stato dipende esclusivamente dallo stato precedente

Nel modello di Langevin, la transizione corrisponde al cambio di velocità della particella, definita da:

$$du' = -\frac{u'}{Tl}dt + \sqrt{C_o \varepsilon}dW$$

In cui dW è il processo di rumore bianco, e ε è il tasso medio di dissipazione; quest'ultimo è dato dalla formula seguente:

$$\varepsilon = \frac{2\sigma^2}{C_0 Tl}$$

Dato che ci interessa la velocità a due sitanti di tempo successivi, si è calcolato:

$$u'(t + \Delta t) - u'(t) = -\frac{u'(t)}{Tl}\Delta t + \sqrt{C_0 \varepsilon}dW$$

Di conseguenza si ricava che:

$$u'(t + \Delta t) = (1 - \frac{\Delta t}{Tl})u'(t) + \sqrt{C_0 \varepsilon}dW$$

Dove Δt = *0.1 * Lagrange_Time*

Dal punto di vista implementativo la particella è stata vista come un vettore, dove ad ogni indice ho una velocità relativa di tale particella ad un determinato istante di tempo. L'istante di tempo corrisponde a Δt * indice

Ovviamente, per ogni particella, l'istante iniziale ha valore 0 in quanto la particella si suppone parta da ferma

Come detto precedentemente il valore all'istante Δt+1 dipende solo dal valore dell'istante precedente Δt (processo di Markov)

Per ogni valore si calcola il rumore bianco tramite la funzione rnorm(), che simula variabili casuali aventi una distribuzione normale specificata, e Δt

Questo processo è stato ripetuto su una serie di particelle e, dato che la struttura utilizzata per la singola particella è il vettore, la scelta più oculata per memorizzare i valori calcolati è stata quella di usare una matrice.

Risultati ottenuti

Si sono prese in considerazione 20 particelle e, per quanto riguarda le variazioni di velocità, si è ottenuto il grafico riportato in Figura 10

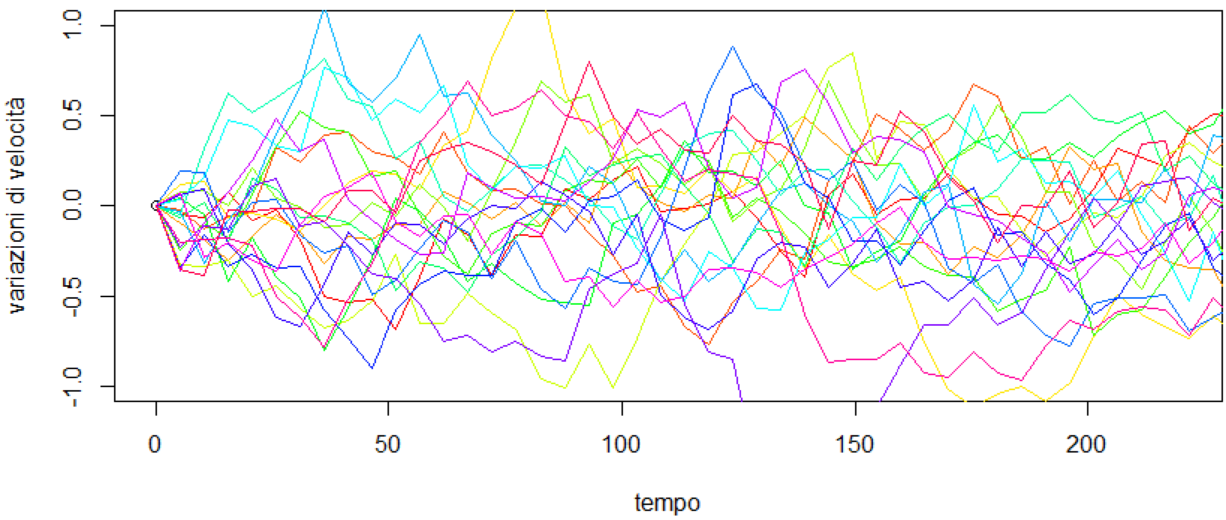

Figura 10 Modello di Langevin, velocità di 20 particelle

Il valor medio calcolato su tali particelle è 0.003068862, che si avvicina molto al valore teorico di media nulla

Anche la deviazione standard è molto simile tra quella calcolata sulle particelle e quella in input

- *Deviazione particelle: 0.4246568*
- *Deviazione input: 0.4146654*