Relazione Consenso Bizantino (Monte Carlo Trace)

Pezzano Enrico

L'algoritmo (randomizzato) implementato in questo laboratorio ci permette di calcolare la traccia di una matrice, o meglio, stimarla non essendo affidabile al 100%. Stima che dovrebbe essere sempre più precisa all'aumentare di M, numero di vettori di *Rademacher* (5, 10, 25, 100).

IL NOSTRO PROBLEMA
Abbiamo stimato la traccia del prodotto di una matrice 300x300 per la sua trasposta iterando un algoritmo di tipo Monte Carlo 100 volte per ogni M, con M numero di campioni.
Successivamente abbiamo calcolato la norma di Frobenius, la varianza campionaria della stima e infine confrontato la traccia stimata con quella reale, con la traccia reale ±*sigma m* e confrontato la varianza campionaria della stima con due volte la norma di Frobenius elevata alla seconda, fratto il numero dei campioni.

IL CODICE PER RISOLVERLO
Tramite la funzione *fill_matrix()* il programma in c riempi una matrice 300x300 con numeri float campionati dall'intervallo [0,1], calcola la trasposta con l'omonima funzione e la moltiplica per la matrice originale, ottenendo così la matrice A. Successivamente calcola la traccia di A utilizzando la funzione *trace(),* in modo da sommare gli elementi sulla diagonale della matrice e calcola la varianza della media campionaria attraverso la formula:

$$\frac{4}{M}\sum_{i=1}^{n}\sum_{r\leq i} A_{ir}^2$$

Dopo di che itera un ciclo for 100 volte il seguente algoritmo:
Genero un vettore di *Rademacher* (di dimensione uguale ad A contenente con uguale probabilità numeri 1 e -1), moltiplico A per tale vettore trasposto ed il risultato verrà di nuovo moltiplicato per il vettore iniziale, salvando il risultato in X (che sarà un numero); stimo la traccia seguendo la formula:

$$\langle X\rangle_m = \langle X\rangle_{m-1} + (X_m - \langle X\rangle_{m-1})/m$$

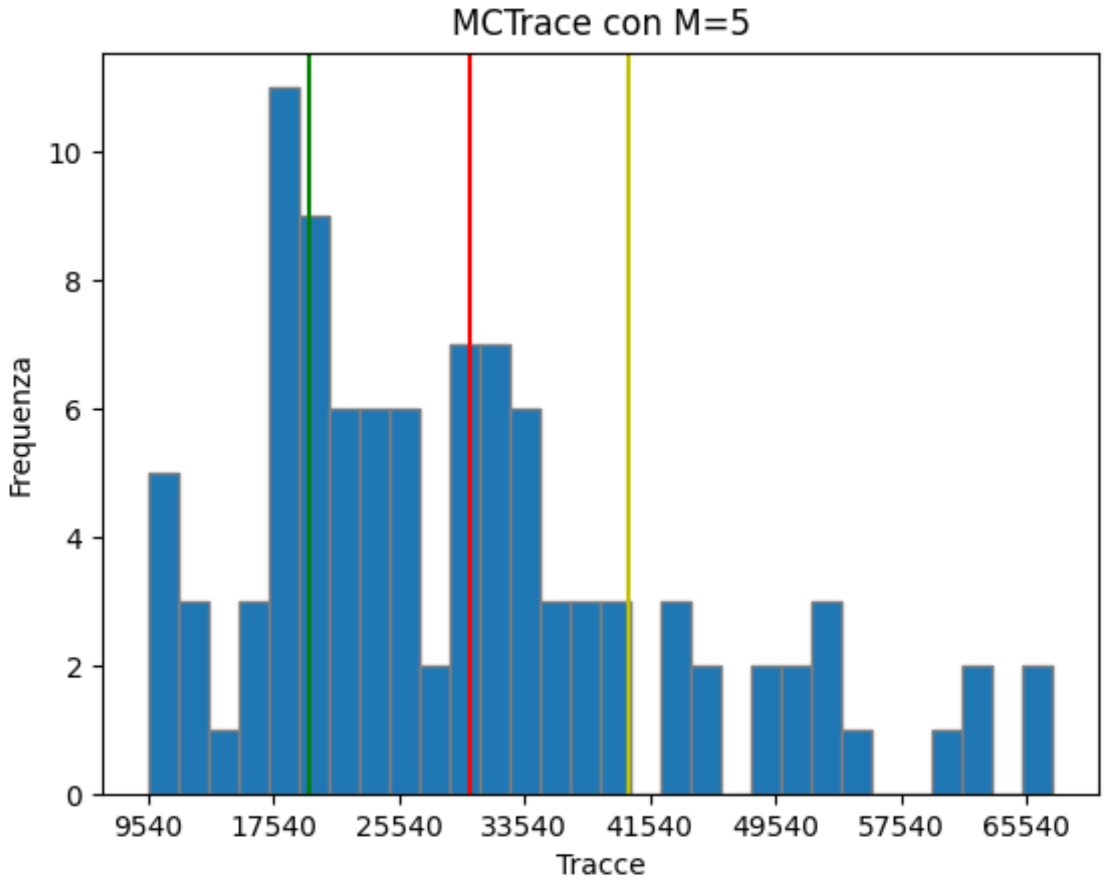
A questo punto il codice calcola la traccia stimata media e la media delle varianze campionarie delle stime che ottenute.
Alla fine stampo a terminale i risultati: la *norma di Frobenius* (prima elevata al quadrato), raddoppiata e divisa per *M* e finalmente si esegue il codice *python* per generare gli istogrammi.

CONSIDERAZIONI FINALI

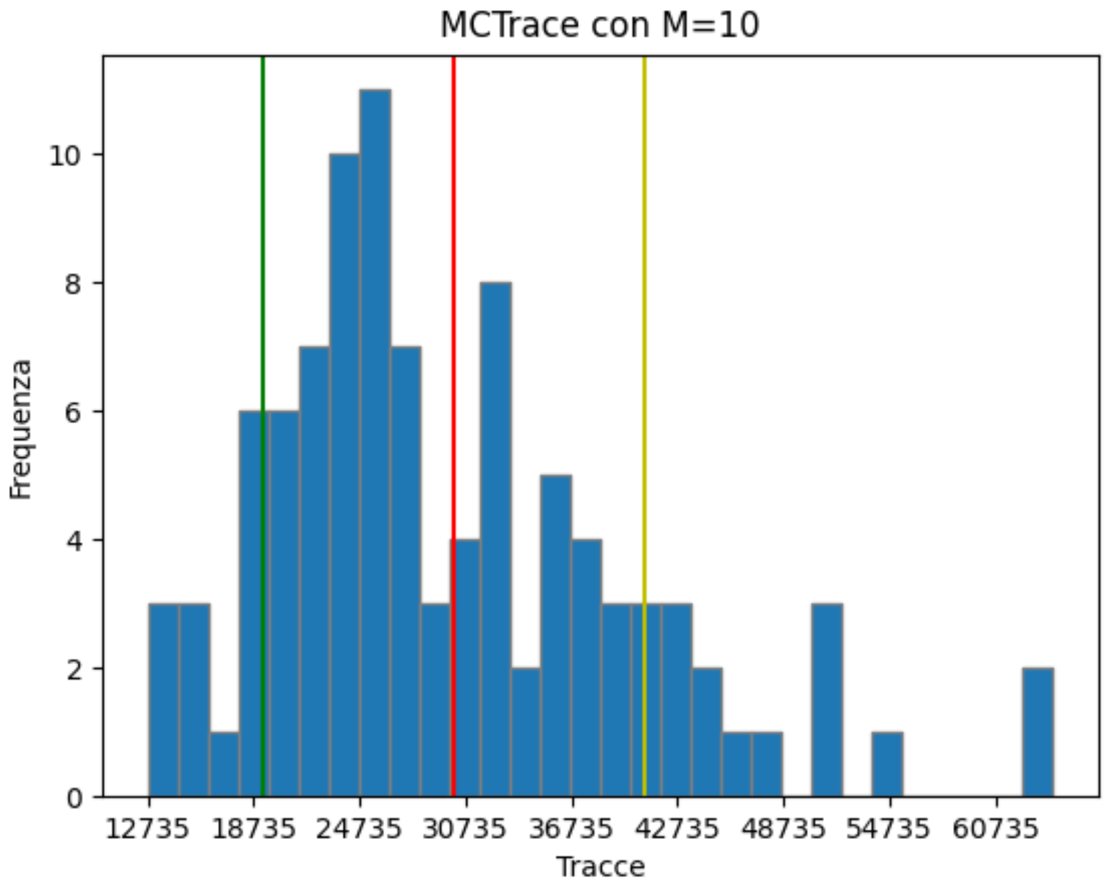
```
La traccia effettiva della matrice A è: 30105.9



Con M = 5:
La varianza della media campionaria è: 2.04709e+08
La traccia stimata media è: 29822.2
Il quadrato della norma di Frobenius è: 5.14807e+08
La varianza campionaria media della stima è: 2.0896e+06
Due volte il quadrato della norma di Frobenius fratto M equivale a: 2.05923e+08
Fine calcolo. Salvo uno dei 100 valori della varianza campionaria per poi confrontarlo nell'istogramma.


Con M = 10:
La varianza della media campionaria è: 1.02355e+08
La traccia stimata media è: 28828
Il quadrato della norma di Frobenius è: 5.14807e+08
La varianza campionaria media della stima è: 940930
Due volte il quadrato della norma di Frobenius fratto M equivale a: 1.02961e+08
Fine calcolo. Salvo uno dei 100 valori della varianza campionaria per poi confrontarlo nell'istogramma.


Con M = 25:
La varianza della media campionaria è: 4.09419e+07
La traccia stimata media è: 30275.8
Il quadrato della norma di Frobenius è: 5.14807e+08
La varianza campionaria media della stima è: 2.99524e+07
Due volte il quadrato della norma di Frobenius fratto M equivale a: 4.11845e+07
Fine calcolo. Salvo uno dei 100 valori della varianza campionaria per poi confrontarlo nell'istogramma.


Con M = 100:
La varianza della media campionaria è: 1.02355e+07
La traccia stimata media è: 30223
Il quadrato della norma di Frobenius è: 5.14807e+08
La varianza campionaria media della stima è: 1.13313e+07
Due volte il quadrato della norma di Frobenius fratto M equivale a: 1.02961e+07
```

Considerando i risultati soprastanti, osserviamo che la stima della traccia all'aumentare di *M* si avvicina sempre di più al valore reale, ma dalla disuguaglianza di *Chebyshev* si intuisce che *M* dipende da $1/\varepsilon^2$; quindi, per ottenere una stima molto precisa servirà un *M* molto grande.
La traccia della matrice rimane ad ogni esecuzione del programma intorno a 30000.
La varianza della media campionaria e la varianza campionaria media della stima risultano sempre minori o uguali del doppio della *norma di Frobenius* al quadrato fratto M.

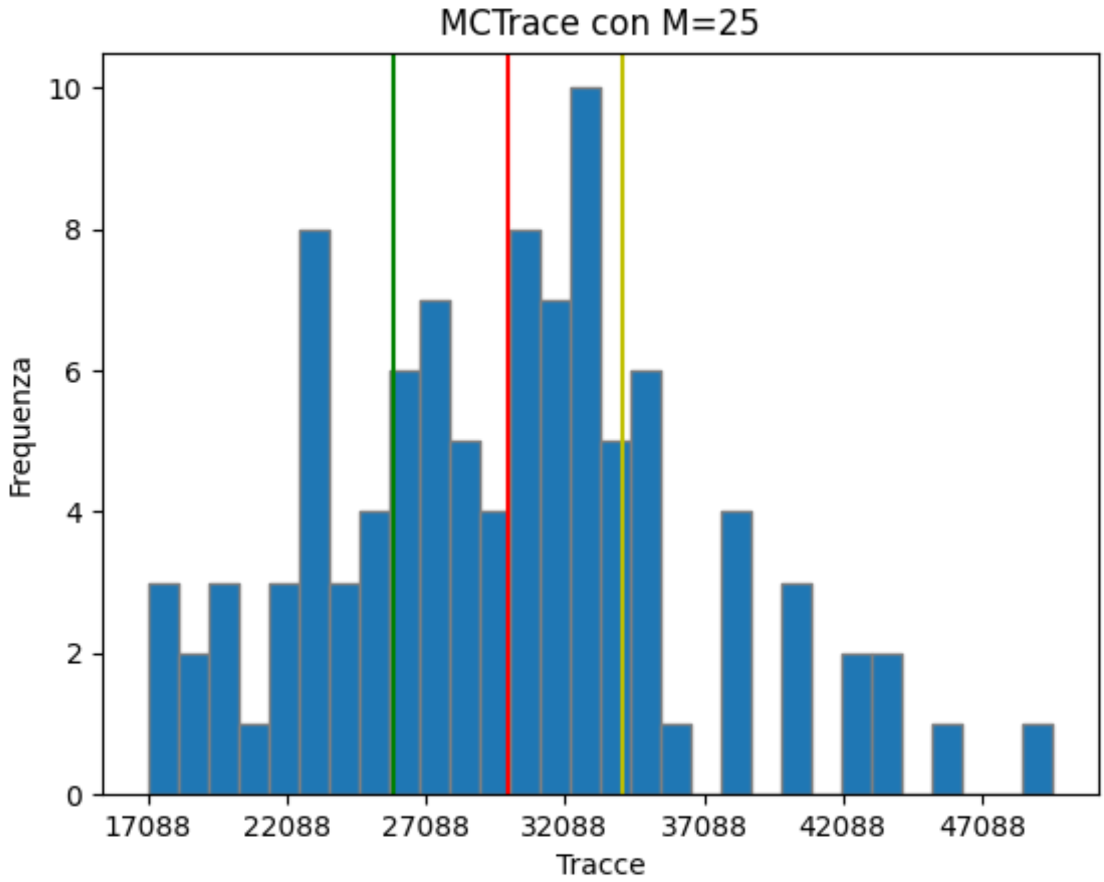
Adesso analizziamo gli istogrammi ottenuti:

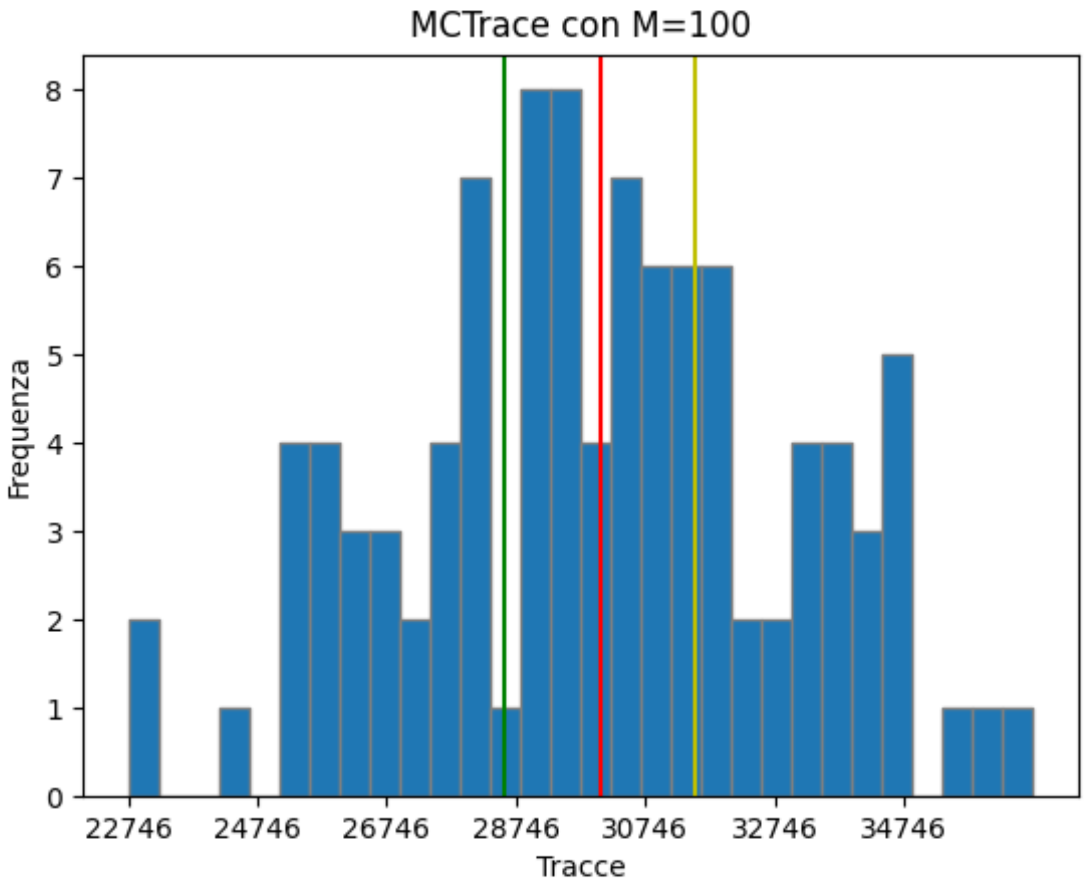


MCTrace con M=5
Frequenza
0
2
4
6
8
10
9540
17540
25540
33540
41540
49540
57540
65540
Tracce



MCTrace con M=10
Frequenza
0
2
4
6
8
10
12735
18735
24735
30735
36735
42735
48735
54735
60735
Tracce


MCTrace con M=25
Frequenza
Tracce
0
2
4
6
8
10
17088
22088
27088
32088
37088
42088
47088


MCTrace con M=100
Frequenza
Tracce
0
1
2
3
4
5
6
7
8
22746
24746
26746
28746
30746
32746
34746

Le linee gialle e verdi sono rispettivamente *Tr(**A**)* - σ_M e *Tr(**A**)* + σ_M mentre quella rossa è la traccia reale di *A*.
Si può osservare dai grafici che all'aumentare di *M* (e dei relativi *vettori di Rademacher*) le stime della traccia di A vicine alla traccia reale si ripetono più volte e che le tracce stimate con maggior frequenza stanno all'interno delle righe gialle e verdi; quindi, si avrà una stima dei valori ottenuti sempre più vicina alla traccia effettiva.

Infine, è interessante notare che la stima è effettivamente controllata superiormente dal quadrato della *norma di Frobenius*, che risulta sempre maggiore della varianza; inoltre la varianza diminuisce con l'aumentare di *M*, per l'aumento della precisione che porta l'intervallo della stima a diminuire.