

# Relazione: LS Quicksort

Lorenzo Livio Vaccarecci (matr. 5462843)

A.A. 2023/2024

## 1 Codice

```
1 import random
2 from tqdm import tqdm
3 import matplotlib.pyplot as plt
4 import math
5
6 X = 0
7
8 def confronta(s, pos, Sequence):
9     global X
10    sequenceMin = []
11    sequenceMax = []
12    for i in range(len(Sequence)):
13        if (i != pos):
14            if Sequence[i] < s:
15                sequenceMin.append(Sequence[i])
16                X += 1
17            else:
18                sequenceMax.append(Sequence[i])
19                X += 1
20    return [sequenceMin, sequenceMax]
21
22 def LVQuickSort(Sequence):
23     if (len(Sequence) <= 1):
24         return Sequence
25     pos = random.randint(0, len(Sequence)-1)
26     s = Sequence[pos]
27     confronto = confronta(s, pos, Sequence)
28     sequenceMin = LVQuickSort(confronto[0])
29     sequenceMax = LVQuickSort(confronto[1])
30     sorted_sequence = sequenceMin
31     sorted_sequence.extend([s])
32     sorted_sequence.extend(sequenceMax)
33     return sorted_sequence
34
35 def random_array(n):
36     arr = []
37     for i in range(n):
38         arr.append(random.randint(0, n))
39     return arr
40
```

```

41 def valore_medio(R, Xr):
42     sommatoria = sum(Xr)
43     return (1/R) * sommatoria
44
45 def deviazione_standard(R, Xr, u):
46     sommatoria = 0
47     for i in Xr:
48         sommatoria += ((i-u)**2)
49     return (1/(R-1)) * sommatoria
50
51 def markov(mu, val):
52     return mu/(val*mu)
53
54 def chebyshev(mu, val, dev):
55     return dev / (((val-1)**2)*(mu**2))
56
57 def conta_frequenze(Xr, n):
58     k = 0
59     for x in Xr:
60         if x >= n*val_medio:
61             k += 1
62     return k
63
64 n = 10**4
65 R = 10**5
66 Xr = []
67 array = random_array(n)
68
69 for i in tqdm(range(R)):
70     X = 0
71     LVQuickSort(array)
72     Xr.append(X)
73
74 val_medio = valore_medio(R, Xr)
75 dev_standard = deviazione_standard(R, Xr, val_medio)
76
77 print("Valore medio: " + str(val_medio))
78 print("Varianza: " + str(dev_standard))
79 print("Deviazione standard: " + str(math.sqrt(dev_standard)))
80
81 plt.hist(Xr, edgecolor="black", bins=50)
82 plt.xlabel("Numero di confronti")
83 plt.ylabel("Frequenza")
84 plt.show()
85
86 v1 = 2
87 v2 = 3
88
89 print(markov(val_medio, v1))
90 print(markov(val_medio, v2))
91
92 print(chebyshev(val_medio, v1, dev_standard))
93 print(chebyshev(val_medio, v2, dev_standard))
94
95 print("Frequenza empirica di X per il doppio: ", conta_frequenze(Xr, 2)/R)
96 print("Frequenza empirica di X per il triplo: ", conta_frequenze(Xr, 3)/R)

```

## 2 Obiettivo

Determinare il numero medio di confronti effettuati dall'algoritmo Quicksort Las Vegas per ordinare un array di  $n$  elementi e calcolare la deviazione standard di tale valore. Inoltre, verificare la validità delle disuguaglianze di Chebyshev e Markov per il doppio e il triplo del valore medio confrontandole con la frequenza di valori superiori o uguali al doppio o al triplo del valore medio.

## 3 Formule utilizzate

### 3.1 Formule principali

#### 3.1.1 Valore medio e deviazione standard

$$\hat{\mu} = \frac{1}{R} \sum_{r=1}^R X_r \quad (1)$$

$$\hat{\sigma}^2 = \frac{1}{R-1} \sum_{r=1}^R (X_r - \hat{\mu})^2 \quad (2)$$

Dove  $X_r$  è il numero di confronti effettuati al passo  $r$ -esimo e  $R$  è il numero di esecuzioni dell'algoritmo.

#### 3.1.2 Disuguaglianze

Chebyshev:

$$Pr \{X \geq v\mu\} \leq \frac{\hat{\sigma}^2}{(v-1)^2 \hat{\mu}^2} \quad (3)$$

Markov:

$$Pr \{X \geq v\mu\} \leq \frac{\hat{\mu}}{v\hat{\mu}} \quad (4)$$

#### 3.1.3 Formule ausiliarie

$$\mathbb{E}[X] = \sum_{i=1}^{n-1} 2 \ln(n-i+1) = 2 \sum_{i=1}^{n-1} \ln(n-i+1) \quad (5)$$

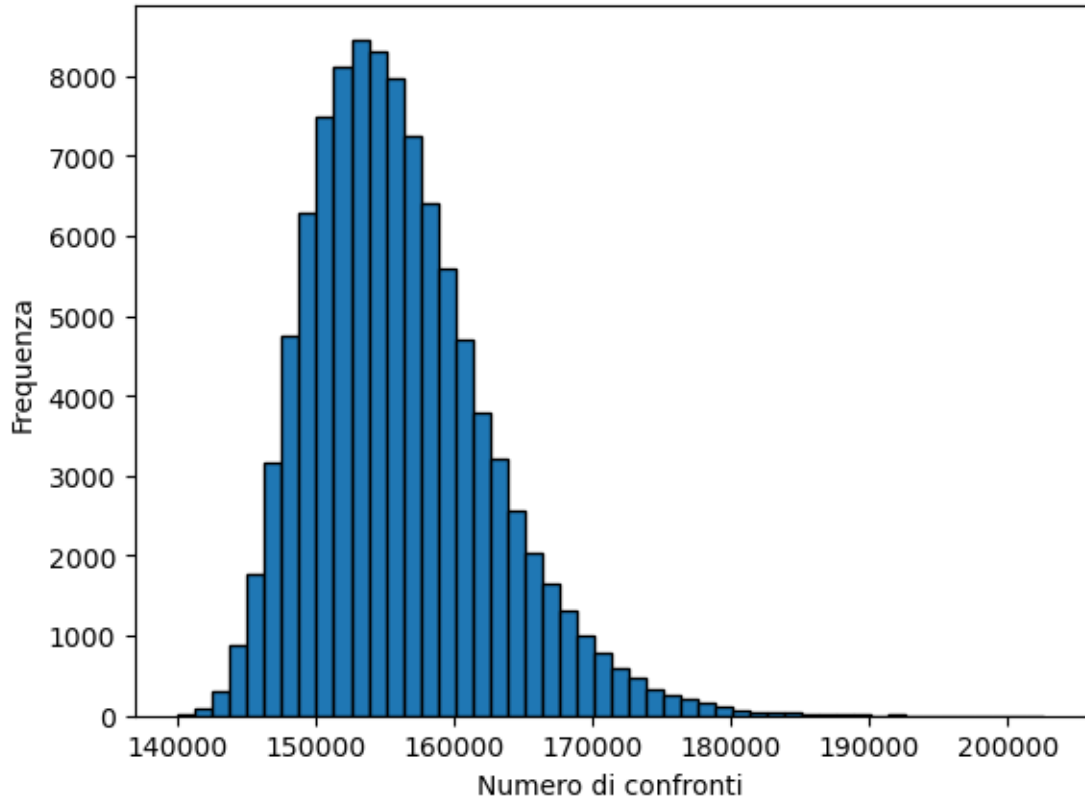
$$\mathbb{E}[X] = 2n \ln(n) \quad (6)$$

## 4 Risultati e Conclusioni

Dopo aver implementato l'algoritmo Quicksort Las Vegas, si è utilizzata una sequenza di numeri casuali di cardinalità  $10^4$  eseguendo l'algoritmo per  $10^5$  volte e si sono ottenuti i seguenti valori con la formula (1) e la radice della formula (2):

- Valore medio ( $\hat{\mu}$ )  $\simeq 156533.23$
- Deviazione standard ( $\hat{\sigma}$ )  $\simeq 6498.49$

Usando il numero di confronti  $X$  effettuati si può generare il seguente istogramma:



Il grafico mostra che la distribuzione dei valori è ampia (questo conferma la deviazione standard elevata), il picco si trova in un valore vicino a 150000 confermando il valore medio ottenuto. Un'ulteriore conferma che il valore medio è quello atteso è data dalla formula (5):

$$2 \sum_{i=1}^{10^4-1} \ln(10^4 - i + 1) \simeq 164216.47$$

E in modo simile dalla formula (6):

$$2 \cdot 10^4 \ln(10^4) \simeq 184206.81$$

Si può notare che i valori medi ottenuti dalle formule teoriche e quello calcolato dall'implementazione dell'algoritmo sono molto simili. Questa discrepanza può essere attribuita a diversi fattori, tra cui la natura approssimativa delle formule teoriche e dalla scelta del pivot. Tuttavia, il valore medio ottenuto dal programma è ragionevole e conferma l'efficienza dell'algoritmo.

Calcolata la disuguaglianza di Chebyshev usando la formula (3) con  $v = 2$   $v = 3$  rispettivamente:

$$\frac{6498.49^2}{(2-1)^2 \cdot 156533.23^2} \simeq 0.0017 \simeq 0.17\%$$

$$\frac{6498.49^2}{(3-1)^2 \cdot 156533.23^2} \simeq 0.0004 \simeq 0.04\%$$

Calcolata anche la disuguaglianza di Markov per le stesse  $v$  usando la formula (4) si ottengono facilmente le due probabilità:

$$\frac{156533.23}{2 \cdot 156533.23} = 0.5 = 50\%$$

$$\frac{156533.23}{3 \cdot 156533.23} \simeq 0.33 \simeq 33\%$$

Secondo la disuguaglianza di Markov, la probabilità che il numero di confronti sia superiore a  $2\hat{\mu}$  può arrivare fino al 50%, mentre la probabilità che sia superiore a  $3\hat{\mu}$  fino al 33%. Questo significa che la disuguaglianza di Markov è molto meno cauta rispetto a quella di Chebyshev che, invece, fornisce un limite superiore alla probabilità che il numero di confronti sia superiore a  $2\hat{\mu}$  del 0.17% e del 0.04% per  $3\hat{\mu}$ . Le frequenze del numero di confronti superiore al doppio o al triplo di  $\hat{\mu}$  sono state pari a 0 ed effettivamente  $0 \leq 0.17\%, 0 \leq 0.04\%$  (per Chebyshev) e  $0 \leq 50\%, 0 \leq 33\%$  (per Markov).

## 5 Conclusioni

In conclusione, i risultati ottenuti dall'implementazione dell'algoritmo Quicksort Las Vegas confermano la sua efficienza per la risoluzione di problemi di ordinamenti di array molto grandi. Il valore medio e la deviazione standard calcolati indicano che la distribuzione dei valori è ampia ma concentrata attorno al valore medio come confermato dall'istogramma.

Le disuguaglianze di Chebyshev e Markov confermano che la probabilità che il numero di confronti sia superiore al doppio e al triplo del valore medio è molto bassa ed è stata confermata dalla frequenza dei valori superiori al doppio e al triplo del valore medio.