Relazione: LS Quicksort

Lorenzo Livio Vaccarecci (matr. 5462843)

A.A. 2023/2024

```python
import random
from tqdm import tqdm
import matplotlib.pyplot as plt
import math

X = 0

def confronta(s, pos, Sequence):
    global X
    sequenceMin = []
    sequenceMax = []
    for i in range(len(Sequence)):
        if (i != pos):
            if Sequence[i] < s:
                sequenceMin.append(Sequence[i])
                X += 1
            else:
                sequenceMax.append(Sequence[i])
                X += 1
    return [sequenceMin, sequenceMax]

def LVQuickSort(Sequence):
    if (len(Sequence) <= 1):
        return Sequence
    pos = random.randint(0, len(Sequence)-1)
    s = Sequence[pos]
    confronto = confronta(s, pos, Sequence)
    sequenceMin = LVQuickSort(confronto[0])
    sequenceMax = LVQuickSort(confronto[1])

    # necessario per avere in output un array senza troncamenti
    sorted_sequence = sequenceMin
    sorted_sequence.extend([s])  # la funzione extend concatena
    sorted_sequence.extend(sequenceMax)

    return sorted_sequence

def random_array(n):
    arr = []
    for i in range(n):
        arr.append(random.randint(0, n))
    return arr

```

```
def valore_medio(R, Xr):
    sommatoria = sum(Xr)
    return (1/R) * sommatoria

def deviazione_standard(R, Xr, u):
    sommatoria = 0
    for i in Xr:
        sommatoria += ((i-u)**2)
    return (1/(R-1)) * sommatoria

def markov(mu, val):
    return mu/(val*mu)

def chebyshev(mu, val, dev):
    return dev / (((val-1)**2)*(mu**2))

n = 10**4
R = 10**5
Xr = []

for i in tqdm(range(R)):
    array = random_array(n)
    X = 0
    LVQuickSort(array)
    Xr.append(X)

val_medio = valore_medio(R, Xr)
dev_standard = deviazione_standard(R, Xr, val_medio)

print("Valore medio: " + str(val_medio))
print("Varianza: " + str(dev_standard))
print("Deviazione standard: " + str(math.sqrt(dev_standard)))

plt.hist(Xr, edgecolor="black", bins=50)
plt.xlabel("Numero di confronti")
plt.ylabel("Frequenza")
plt.show()

v1 = 2
v2 = 3

print(markov(val_medio, v1))
print(markov(val_medio, v2))

print(chebyshev(val_medio, v1, dev_standard))
print(chebyshev(val_medio, v2, dev_standard))
```

Dopo aver implementato l'algoritmo Quicksort Las Vegas, si sono utilizzate delle sequenze di numeri casuali di cardinalità 10^4 eseguendo l'algoritmo per $R = 10^5$ volte (con array generati in modo casuale) e si sono ottenuti i seguenti valori:

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

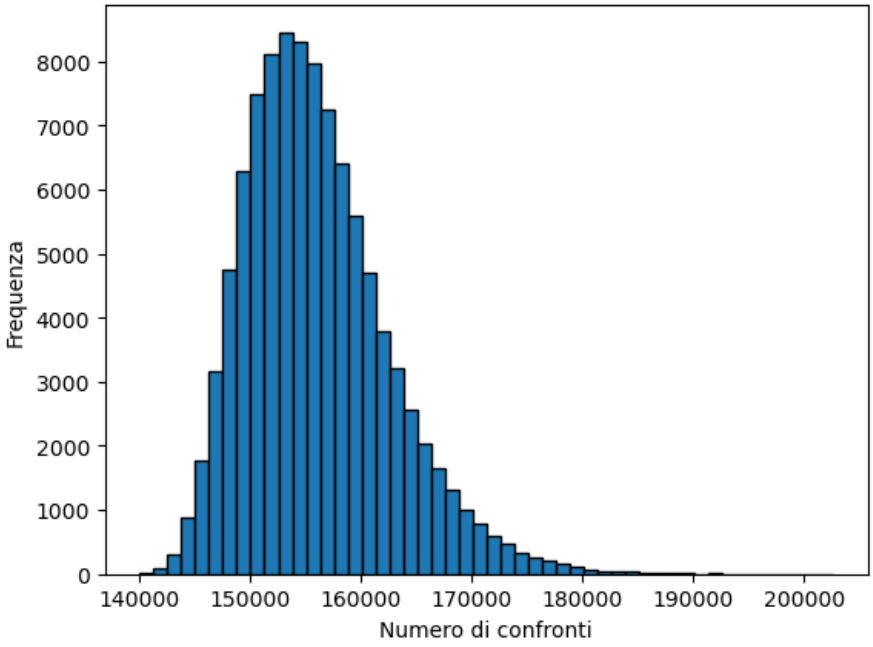
Per calcolare i valori sono state usate le formule:

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

Dove X_r è il numero di confronti effettuati al passo r-esimo e:

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

Usando il numero di confronti 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:

$$\mathbb{E}[X] = \sum_{i=1}^{n-1} 2\ln(n-i+1) = 2\sum_{i=1}^{n-1}\ln(n-i+1) = 2\sum_{i=1}^{10^4-1}\ln(10^4-i+1) \simeq 164216.47$$

In modo simile:

$$\mathbb{E}[X] = 2n\ln(n) = 2\cdot 10^4\ln(10^4) \simeq 184206.81$$

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 la presenza di numeri casuali. Tuttavia, il valore medio ottenuto dal programma è ragionevole e conferma l'efficienza dell'algoritmo.
Calcolata la disuguaglianza di Chebyshev usando la formula:

$$\frac{\hat{\sigma}^2}{(v-1)^2\hat{\mu}^2}$$

Abbiamo che per il doppio ($2\hat{\mu} \simeq 313066.46$) e il triplo ($3\hat{\mu} \simeq 469599.69$) del valore medio con $v = 2 \quad v = 3$ rispettivamente:

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

e

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

Calcolata anche la disuguaglianza di Markov per le stesse v usando la formula:

$$\frac{\hat{\mu}}{v\hat{\mu}}$$

Si ottengono facilmente le due probabilità:

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

e

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

Secondo la disuguaglianza di Markov, la probabilità che il numero di confronti sia superiore a $2\hat{\mu}$ è del 50%, mentre la probabilità che sia superiore a $3\hat{\mu}$ è del 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 $3\hat{\mu}$.
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.