Algoritmi randomizzati / 26-04

Ne esistono due classi: Las Vegas e Montecarlo
Esistono solo pseudocodice che restituisce numeri casuali (stesso input = stesso output)

Las Vegas

Hanno una proprietà che li rende simili ai deterministici: fissato un input l'algoritmo restituisce lo stesso output

La randomizzazione sta nel tempo di risposta: il tempo è una variabile casuale. Possiamo rilassare il concetto di caso peggiore.

Quicksort caso ottimo e medio $(n\log n)$, peggiore (n^2)

Dato una sequenza input l'alg Las Vegas Quicksort la mischia casualmente più e più volte, per $n\log n$ volte.

$$E\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E[X_i]$$ trasporre linearità

$N = 10^6$ $\quad X_i$ dipendenti

$X_1 = x_1, \dots, x_n$
valori piccoli
$p(x_1) \quad p(x_n)$

$E[X_1] = \sum_i p(x_i)\, x_i$ media pesata

$E[X_1, \dots, X_{1 milione}]$
$p($1 milione di argomenti$)$ è utile dunque il teorema sopra
non è possibile calcolarla

Esempio

Valore atteso di quante persone ricevono il proprio cappello dopo averlo depositato in una cesta comune e ottenendolo casualmente fra quelli
Totale 100 persone

n persone n	contento
1	1
2	1
3	1

unica persona

$\boxed{1\ 2}$ $\boxed{2\ 1}$ eventi indip, somma $\frac{1}{2}+\frac{1}{2}=1$

1 2 3, 1 3 2, 2 1 3, 2 3 1, 3 1 2, 3 2 1

↳ 6 casi, ogni evento ha
$p(1,2,3) = p(1,3,2) = \dots = \frac{1}{6}$

contenti ↓ ↗ marginali

$p(0) = \frac{2}{6} = \frac{1}{3} = p(0,0,0)$
$p(1) = \frac{1}{2} = p(1,0,0) + p(0,1,0) + p(0,0,1)$
$p(2) = 0$
$p(3) = \frac{1}{6}$

Il crescere delle persone è ingestibile, sfrutto la linearità e la variabile casuale indicatrice (di bernoulli, assume due valori)

$p_X(1) = p$
$p_X(0) = 1-p$

$E[X] = 1 \cdot p + 0(1-p) = p$

Con 100 persone, p ottenere il proprio cappello è $p = \frac{1}{n}$

$$E\left[\sum_i X_i\right] = \sum_i E[X_i] = \sum_i^n \frac{1}{n} = 1$$

Pseudocodice

LVQS(S)
if $|S| \le 1$
then return S

else
1) campiona s da S con prob uniforme
2) forma $S_<$ $S_\geq$ (confrontando s con $S \setminus s$)
3) LVQS ($S_\geq$)
4) LVQS ($S_<$)
5) return concatenazione di $S_<$ e $S_\geq$

Dati $i, j = 1, \dots, n$ supposto $i < j$
Variabile indicatrice I

$$I(i,j) = \begin{cases} 1 & \text{se } i \text{ è confrontato con } j \text{ (es. } j\text{=6 è già stato confrontato con 5, 4 ...)} \\ 0 & \text{altrimenti} \end{cases}$$

probabilità di $P_I(1) = p$ e $P_I(0) = 1-p$
cioè p di confrontare alcune coppie

ad esempio 2 e 28 non sono confrontati se $1\ 2\ 3 < 5$ e $28 > 5$ preso $s = 5$, mentre 6 e 28 e 2 e 3 può essere

Una coppia viene confrontata se presenti nella stessa sottosequenza
es 6 e 11

→ 6 casi sono determinanti
- 2 casi su 6 sono confrontabili
- nei restanti rendo 6 e 11 mai confrontabili

$$p(i) = \frac{2}{j-i+1}$$

↓
presto perché li separo in due sottosequ. diverse

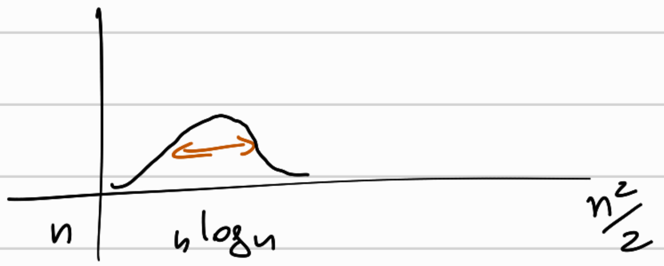
$$\mathbb{E}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} I(i,j)\right] = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{2}{j-i+1} \approx c\,(n \ln n)$$

costante

valore atteso di confronti = scambi $O(n \ln n)$

Esercizio

Implementare QS, lanciarlo tante volte x tracciarne un istogramma. Osservare che il valore atteso di confronti è $n \ln n$



Per diversi input avrò diversi tempi di esecuzione

es. 1 milione di lanci e osservo istogramma con i tempi di esecuzione

lo scostamento si riduce secondo $\sqrt{n}$ (vedi Chebyshev)