Algoritmos Aleatorizados

Parte III

Segundo semestre 2022

IIC2283

Prof. Nicolás Van Sint Jan

Outline

Calculo de la mediana

Desigualdades de Markov y Chebyshev

La mediana de una lista de números enteros

Suponga que n es un número impar y $L[1 \ldots n]$ es una lista de números enteros sin elementos repetidos.

- Recuerde que usamos la notación $L[1 \ldots n]$ para indicar que L es una lista con n elementos

Definición

$L[i]$ es la **mediana** de L si:

$$\begin{aligned} |\{j \in \{1, \ldots, n\} \mid L[j] < L[i]\}| &= \left\lfloor \frac{n}{2} \right\rfloor \\ |\{k \in \{1, \ldots, n\} \mid L[k] > L[i]\}| &= \left\lfloor \frac{n}{2} \right\rfloor \end{aligned}$$

La mediana de una lista de números enteros

Ejercicio

Construya un algoritmo que calcule la mediana de una lista $L[1 \dots n]$ y que en el peor caso sea $\mathcal{O}(n \cdot \log_2(n))$

- Considere como la operación básica a contar la comparación de números enteros

Vamos a construir un algoritmo aleatorizado de tipo **Las Vegas** para este problema.

¡Este algoritmo funciona en **tiempo lineal**!

Un algoritmo aleatorizado para el cálculo de la mediana

Suponga que el procedimiento **Mergesort**(L) ordena una lista L utilizando el algoritmo Mergesort.

El siguiente procedimiento calcula la mediana de una lista de enteros $L[1 \ldots n]$ (suponiendo que n es impar y L no tiene elementos repetidos):

CalcularMediana($L[1 \ldots n]$)
  if $n < 2001$ **then**
    Mergesort(L)
    return $L[\lceil \frac{n}{2} \rceil]$
  else
    sea R una lista de $\left\lceil n^{\frac{3}{4}} \right\rceil$ números enteros escogido con
        distribución uniforme y de manera independiente desde L
    Mergesort(R)

Un algoritmo aleatorizado para el cálculo de la mediana

$d := R\left[\left\lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \right\rfloor\right]$

$u := R\left[\left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil\right]$

$S := \varnothing$

$m_d := 0$

$m_u := 0$

for $i := 1$ **to** n **do**

  if $d \leq L[i]$ **and** $L[i] \leq u$ **then** **Append**$(S, [L[i]])$

  else if $L[i] < d$ **then** $m_d := m_d + 1$

  else $m_u := m_u + 1$

if $m_d \geq \left\lceil \frac{n}{2} \right\rceil$ **or** $m_u \geq \left\lceil \frac{n}{2} \right\rceil$ **or**

  Length$(S) > 4 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor$ **then return** *sin_resultado*

else

  Mergesort(S)

  return $S\left[\left\lceil \frac{n}{2} \right\rceil - m_d\right]$

El algoritmo es correcto y eficiente

Ejercicios

Demuestre lo siguiente:

1. Si **CalcularMediana**(L) retorna un número entero m, entonces m es la mediana de L

2. Si se tiene un procedimiento **LanzarMoneda**() que retorna 0 ó 1 con probabilidad $\frac{1}{2}$, entonces existe un algoritmo para construir R que invoca a este procedimiento a los más $c \cdot n^{\frac{3}{4}} \cdot \log_2(n)$ veces, donde c es una constante fija y n es el largo de la lista de entrada
   - Podemos suponer que **LanzarMoneda**() en el peor caso es $\mathcal{O}(1)$

3. **CalcularMediana**(L) en el peor caso es $\mathcal{O}(n)$, suponiendo que n es el largo de L y considerando todas las operaciones realizadas

¿Cuál es la probabilidad de no retornar un resultado?

La llamada **CalcularMediana**(L) puede no retornar un resultado

- El procedimiento en este caso retorna *sin_resultado*

Para que **CalcularMediana** pueda ser utilizado en la práctica la probabilidad que no entregue un resultado debe ser baja

La probabilidad de no retornar resultado

Sea $L[1 \dots n]$ una lista de números enteros tal que $n \geq 2001$, n es impar y la mediana de L es m

Defina las siguientes variables aleatorias:

$$\begin{aligned} Y_1 &= \left| \left\{ i \in \left\{ 1, \ldots, \left\lceil n^{\frac{3}{4}} \right\rceil \right\} \mid R[i] \leq m \right\} \right| \\ Y_2 &= \left| \left\{ i \in \left\{ 1, \ldots, \left\lceil n^{\frac{3}{4}} \right\rceil \right\} \mid R[i] \geq m \right\} \right| \end{aligned}$$

Estas son variables aleatorias dado que R es construido escogiendo elementos de L con distribución uniforme (y de manera independiente)

¿ Qué representan Y_1 e Y_2 ?

La probabilidad de no retornar resultado

Lema

CalcularMediana(L) retorna sin_resultado si y sólo si alguna de las siguientes condiciones se cumple:

1. $Y_1 < \left\lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \right\rfloor$
2. $Y_2 \leq \left\lceil n^{\frac{3}{4}} \right\rceil - \left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil$
3. **Length**$(S) > 4 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor$

Ejercicio

Demuestre el lema.

La probabilidad de no retornar resultado

Tenemos entonces que la probabilidad de que **CalcularMediana**(L) retorne *sin_resultado* es igual a:

$$\Pr\left(Y_1 < \left\lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \right\rfloor \vee Y_2 \leq \left\lceil n^{\frac{3}{4}} \right\rceil - \left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil \vee \mathbf{Length}(S) > 4 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor\right)$$

Necesitamos entonces acotar superiormente

- $\Pr(Y_1 < \lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \rfloor)$
- $\Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil - \lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil)$
- $\Pr(\mathbf{Length}(S) > 4 \cdot \lfloor n^{\frac{3}{4}} \rfloor)$

¿ Que herramientas podemos usar para **acotar** estas probabilidades ?

Outline

La desigualdad de Markov

Teorema

Sea X una variable aleatoria no negativa. Para cada $a \in \mathbb{R}^+$ se tiene que:

$$\Pr(X \geq a) \quad \leq \quad \frac{\mathrm{E}(X)}{a}$$

Este resultado se conoce como la **desigualdad de Markov**.

Una demostración de la desigualdad de Markov

Demostración

Suponemos que el recorrido de X es un conjunto finito $\Omega \subseteq \mathbb{R}_0^+$:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{r\in\Omega} r\cdot \Pr(X=r) \\
&= \Bigg(\sum_{r\in\Omega\,:\,r<a} r\cdot \Pr(X=r)\Bigg) + \Bigg(\sum_{s\in\Omega\,:\,s\geq a} s\cdot \Pr(X=s)\Bigg) \\
&\geq \sum_{s\in\Omega\,:\,s\geq a} s\cdot \Pr(X=s) \\
&\geq \sum_{s\in\Omega\,:\,s\geq a} a\cdot \Pr(X=s) \\
&= a\cdot \Bigg(\sum_{s\in\Omega\,:\,s\geq a} \Pr(X=s)\Bigg) \\
&= a\cdot \Pr(X\geq a)
\end{aligned}$$

Concluimos que $\Pr(X\geq a)\leq \frac{\mathrm{E}(X)}{a}$. $\square$

La desigualdad de Chebyshev

Teorema

El siguiente resultado se conoce como la **desigualdad de Chebyshev**:

$$\Pr\left(|X - \mathrm{E}(X)| \geq a\right) \quad \leq \quad \frac{\mathrm{Var}(X)}{a^2}$$

Demostración

Utilizando la desigualdad de Markov obtenemos:

$$\begin{aligned}
\Pr(|X - \mathrm{E}(X)| \geq a) &= \Pr((X - \mathrm{E}(X))^2 \geq a^2) \\
&\leq \frac{\mathrm{E}((X - \mathrm{E}(X))^2)}{a^2} \\
&= \frac{\mathrm{Var}(X)}{a^2}
\end{aligned}$$

$\square$