

La mediana de una lista de números enteros

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

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

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

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

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 $O(n \cdot \log_2(n))$

- ▶ 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

- Recuerde que las listas son pasados por referencia en este curso

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

CalcularMediana($L[1 \dots n]$)

if $n < 2001$ **then**

Mergesort(L)

return $L[\lceil \frac{n}{2} \rceil]$

else

 sea R una lista de $\lceil n^{\frac{3}{4}} \rceil$ números enteros escogido con
 distribución uniforme y de manera independiente desde L

Mergesort(R)

Un algoritmo aleatorizado para el cálculo de la mediana

```
 $d := R[\lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \rfloor]$   
 $u := R[\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil]$   
 $S := \emptyset$   
 $m_d := 0$   
 $m_u := 0$   
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  $m_d \geq \lceil \frac{n}{2} \rceil$  or  $m_u \geq \lceil \frac{n}{2} \rceil$  or  
     $\text{Length}(S) > 4 \cdot \lfloor n^{\frac{3}{4}} \rfloor$  then return sin_resultado  
else  
    Mergesort( $S$ )  
    return  $S[\lceil \frac{n}{2} \rceil - m_d]$ 
```

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 $O(1)$
3. **CalcularMediana**(L) en el peor caso es $O(n)$, 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

- ▶ Vamos a demostrar esto

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:

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

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

La probabilidad de no retornar resultado

Lema

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

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

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:

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

Necesitamos entonces acotar superiormente $\Pr(Y_1 < \lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \rfloor)$, $\Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil - \lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil)$ y $\Pr(\text{Length}(S) > 4 \cdot \lfloor n^{\frac{3}{4}} \rfloor)$

- Vamos a ver dos desigualdades muy útiles para acotar probabilidades

La desigualdad de Markov

Teorema

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

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

Una demostración de la desigualdad de Markov

Tenemos que:

$$\begin{aligned} E(X) &= \sum_{r \in \mathbb{R}_0^+} r \cdot \Pr(X = r) \\ &= \left(\sum_{r \in \mathbb{R}_0^+ : r < a} r \cdot \Pr(X = r) \right) + \left(\sum_{s \in \mathbb{R}_0^+ : s \geq a} s \cdot \Pr(X = s) \right) \\ &\geq \sum_{s \in \mathbb{R}_0^+ : s \geq a} s \cdot \Pr(X = s) \\ &\geq \sum_{s \in \mathbb{R}_0^+ : s \geq a} a \cdot \Pr(X = s) \\ &= a \cdot \left(\sum_{s \in \mathbb{R}_0^+ : s \geq a} \Pr(X = s) \right) \\ &= a \cdot \Pr(X \geq a) \end{aligned}$$

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



La desigualdad de Chebyshev

Teorema

$$\Pr(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

Demostración. Utilizando la desigualdad de Markov obtenemos:

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



Utilizando la desigualdad de Chebyshev

Lema

$$Pr(Y_1 < \lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \rfloor) \leq n^{-\frac{1}{4}}$$

Demostración: para cada $i \in \{1, \dots, \lceil n^{\frac{3}{4}} \rceil\}$, definimos una variable aleatoria X_i de la siguiente forma:

$$X_i = \begin{cases} 1 & R[i] \leq m \\ 0 & R[i] > m \end{cases}$$

Tenemos que:

$$Y_1 = \sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} X_i$$

Utilizando la desigualdad de Chebyshev

Dado que la lista L no contiene elementos repetidos tenemos que:

$$\Pr(X_i = 1) = \frac{\lceil \frac{n}{2} \rceil}{n} = \frac{\frac{n-1}{2} + 1}{n} = \frac{1}{2} + \frac{1}{2 \cdot n}$$

De esto se deduce que:

$$\begin{aligned} E(X_i) &= \frac{1}{2} + \frac{1}{2 \cdot n} \\ \text{Var}(X_i) &= \left(\frac{1}{2} + \frac{1}{2 \cdot n} \right) \cdot \left(1 - \frac{1}{2} - \frac{1}{2 \cdot n} \right) \\ &= \left(\frac{1}{2} + \frac{1}{2 \cdot n} \right) \cdot \left(\frac{1}{2} - \frac{1}{2 \cdot n} \right) \\ &= \frac{1}{4} - \frac{1}{4 \cdot n^2} \end{aligned}$$

Utilizando la desigualdad de Chebyshev

Por lo tanto tenemos que:

$$\begin{aligned} E(Y_1) &= E\left(\sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} X_i\right) \\ &= \sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} E(X_i) \\ &= \sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} \left(\frac{1}{2} + \frac{1}{2 \cdot n}\right) \\ &= \lceil n^{\frac{3}{4}} \rceil \cdot \left(\frac{1}{2} + \frac{1}{2 \cdot n}\right) \end{aligned}$$

Utilizando la desigualdad de Chebyshev

Para $i, j \in \{1, \dots, \lceil n^{\frac{3}{4}} \rceil\}$ tal que $i \neq j$ se tiene que X_i es independiente de X_j

Concluimos entonces que:

$$\begin{aligned}\text{Var}(Y_1) &= \text{Var}\left(\sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} X_i\right) \\ &= \sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} \text{Var}(X_i) \\ &= \sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} \left(\frac{1}{4} - \frac{1}{4 \cdot n^2}\right) \\ &= \lceil n^{\frac{3}{4}} \rceil \cdot \left(\frac{1}{4} - \frac{1}{4 \cdot n^2}\right) \\ &\leq \frac{1}{4} \cdot \lceil n^{\frac{3}{4}} \rceil\end{aligned}$$

Utilizando la desigualdad de Chebyshev

Tenemos que:

$$\begin{aligned}\Pr(Y_1 < \lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \rfloor) &\leq \Pr(Y_1 < \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}}) \\ &\leq \Pr(Y_1 < \frac{1}{2} \cdot \lceil n^{\frac{3}{4}} \rceil - n^{\frac{1}{2}}) \\ &\leq \Pr(Y_1 < \lceil n^{\frac{3}{4}} \rceil \cdot (\frac{1}{2} + \frac{1}{2 \cdot n}) - n^{\frac{1}{2}}) \\ &= \Pr(Y_1 < E(Y_1) - n^{\frac{1}{2}}) \\ &= \Pr(n^{\frac{1}{2}} < E(Y_1) - Y_1) \\ &\leq \Pr(|Y_1 - E(Y_1)| > n^{\frac{1}{2}}) \\ &\leq \Pr(|Y_1 - E(Y_1)| \geq n^{\frac{1}{2}})\end{aligned}$$

Utilizando la desigualdad de Chebyshev

Por lo tanto, utilizando la desigualdad de Chebyshev concluimos que:

$$\begin{aligned}\Pr(Y_1 < \lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \rfloor) &\leq \Pr(|Y_1 - E(Y_1)| \geq n^{\frac{1}{2}}) \\ &\leq \frac{\text{Var}(Y_1)}{n} \\ &\leq \frac{1}{4} \cdot \frac{\lceil n^{\frac{3}{4}} \rceil}{n} \\ &\leq \frac{1}{4} \cdot \frac{n^{\frac{3}{4}} + 1}{n} \\ &\leq \frac{1}{4} \cdot \frac{2 \cdot n^{\frac{3}{4}}}{n} \\ &= \frac{1}{2} \cdot n^{-\frac{1}{4}} \\ &\leq n^{-\frac{1}{4}}\end{aligned}$$



Utilizando nuevamente la desigualdad de Chebyshev

Lema

$$\Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil - \lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil) \leq n^{-\frac{1}{4}}$$

Demostración. En este caso tenemos que:

$$\begin{aligned} \Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil - \lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil) &\leq \Pr(Y_2 \leq n^{\frac{3}{4}} + 1 - \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}}) \\ &= \Pr(Y_2 \leq \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} + 1) \end{aligned}$$

Utilizando nuevamente la desigualdad de Chebyshev

Al igual que en el caso anterior, para cada $i \in \{1, \dots, \lceil n^{\frac{3}{4}} \rceil\}$, definimos una variable aleatoria Z_i de la siguiente forma:

$$Z_i = \begin{cases} 1 & R[i] \geq m \\ 0 & R[i] < m \end{cases}$$

Tenemos que:

$$Y_2 = \sum_{i=1}^{\lceil n^{\frac{3}{4}} \rceil} Z_i$$

Utilizando nuevamente la desigualdad de Chebyshev

Como L no contiene elementos repetidos nuevamente obtenemos:

$$\Pr(Z_i = 1) = \frac{\lceil \frac{n}{2} \rceil}{n} = \frac{1}{2} + \frac{1}{2 \cdot n}$$

Por lo que como en el caso anterior concluimos que:

$$\begin{aligned} E(Z_i) &= \frac{1}{2} + \frac{1}{2 \cdot n} \\ \text{Var}(Z_i) &= \frac{1}{4} - \frac{1}{4 \cdot n^2} \end{aligned}$$

Utilizando nuevamente la desigualdad de Chebyshev

Por lo tanto deducimos que:

$$\begin{aligned} E(Y_2) &= \lceil n^{\frac{3}{4}} \rceil \cdot \left(\frac{1}{2} + \frac{1}{2 \cdot n} \right) \\ \text{Var}(Y_2) &= \lceil n^{\frac{3}{4}} \rceil \cdot \left(\frac{1}{4} - \frac{1}{4 \cdot n^2} \right) \\ &\leq \frac{1}{4} \cdot \lceil n^{\frac{3}{4}} \rceil \end{aligned}$$

Utilizando nuevamente la desigualdad de Chebyshev

Finalmente tenemos que:

$$\begin{aligned}\Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil - \lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil) &\leq \Pr(Y_2 \leq \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} + 1) \\ &\leq \Pr(Y_2 \leq \frac{1}{2} \cdot \lceil n^{\frac{3}{4}} \rceil - n^{\frac{1}{2}} + 1) \\ &\leq \Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil \cdot (\frac{1}{2} + \frac{1}{2 \cdot n}) - n^{\frac{1}{2}} + 1) \\ &= \Pr(Y_2 \leq E(Y_2) - n^{\frac{1}{2}} + 1) \\ &= \Pr(n^{\frac{1}{2}} - 1 \leq E(Y_2) - Y_2) \\ &\leq \Pr(n^{\frac{1}{2}} - (1 - \frac{1}{\sqrt{2}}) \cdot n^{\frac{1}{2}} \leq E(Y_2) - Y_2) \\ &= \Pr(\sqrt{\frac{n}{2}} \leq E(Y_2) - Y_2) \\ &\leq \Pr(|Y_2 - E(Y_2)| \geq \sqrt{\frac{n}{2}})\end{aligned}$$

Utilizando nuevamente la desigualdad de Chebyshev

Por lo tanto, utilizando la desigualdad de Chebyshev concluimos que:

$$\begin{aligned}\Pr(Y_2 \leq \lceil n^{\frac{3}{4}} \rceil - \lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \rceil) &\leq \Pr(|Y_2 - E(Y_2)| \geq \sqrt{\frac{n}{2}}) \\ &\leq \frac{\text{Var}(Y_2)}{\frac{n}{2}} \\ &\leq \frac{1}{4} \cdot \frac{\lceil n^{\frac{3}{4}} \rceil}{\frac{n}{2}} \\ &\leq \frac{1}{4} \cdot \frac{2 \cdot n^{\frac{3}{4}}}{\frac{n}{2}} \\ &= n^{-\frac{1}{4}}\end{aligned}$$

□