Ignacio Méndez Pérez
IIC2283 - 2022 - 2

Algoritmos Aleatorizados

Algoritmos aleatorizados:

- **Monte Carlo**: El algoritmo siempre entrega un resultado, pero hay una probabilidad de que sea **incorrecto**.
- **Las Vegas**: Si el algoritmo entrega un resultado es correcto, pero hay una probabilidad de **no entregue resultado**.

Equivalencia de Polinomios

Suponga que:

$$p(x) = \sum_{i=1}^{k} \prod_{j=1}^{r_i} (a_{i,j}x + b_{i,j})$$
$$q(x) = \sum_{i=1}^{l} \prod_{j=1}^{s_i} (c_{i,j}x + d_{i,j})$$

EquivPolAleatorizado$(p(x),\ q(x))$
  $K := 1 + \max\{r_1, \ldots, r_k, s_1, \ldots, s_\ell\}$
  escoja al azar y con distribución uniforme un elemento a
    del conjunto de números naturales $\{1, \ldots, 100 \cdot K\}$
  if $p(a) = q(a)$ **then return** sí
  else return no

K = Máximo exponente entre ambos polinomios +1.

Solo hacemos $\mathcal{O}(n)$ operaciones, donde $n = |p(x)| + |q(x)|$, ya que necesita calcular $p(a)$ y $q(a)$. **Pero me puedo equivocar**.

¿Probabilidad de error?

- Si los polinomios $p(x)$ y $q(x)$ son equivalentes, entonces el algoritmo responde **SI** sin cometer errores.

- Si los polinomios $p(x)$ y $q(x)$ no son equivalentes, el algoritmo puede responder SI al sacar al azar un elemento $a \in \{1, \ldots, 100 \cdot K\}$ tal que $p(a) = q(a)$.

  Esto significa que a es una **raíz** del polinomio

$$r(x) = p(x) - q(x)$$

Sabemos que $r(x)$ NO es el polinomio nulo y que es de grado a lo más K.

- Por lo tanto $r(x)$ tiene a lo más K raíces en $\mathbb{Q}$.

Concluimos que:

$$\Pr(a \text{ sea raíz de } r(x)) \leq \frac{K}{100 \cdot K} = \frac{1}{100}$$

¿Cómo mejoramos esta probabilidad? → Repetimos el algoritmo:

```
EquivPolAleatorizado(p(x), q(x))
    K := 1 + max{r_1, ..., r_k, s_1, ..., s_ℓ}
    A := {1, ..., 100 · K}
    total := 0
    for i := 1 to 10 do
        escoja al azar y con distribución uniforme un elemento a en A
        if p(a) = q(a) then total = total + 1
    if total = 10 then return sí
    return no
```

Ahora la probabilidad será de: $\boxed{\frac{1}{100^{10}}}$

Definición:

Un **polinomio en varias variables** es una expresión de la forma:

$$p(x_1, \dots, x_n) = \sum_{i=1}^{l} \prod_{j=1}^{m_i} \left(\sum_{k=1}^{n} a_{i,j,k} x_k + b_{i,j} \right)$$

donde cada $a_{i,j,k} \in \mathbb{Q}$ y cada $b_{i,j} \in \mathbb{Q}$.

Lema de Schwartz-Zipel

Sea $p(x_1, \dots, x_n)$ un polinomio no nulo de grado k, y sea A un subconjunto finito y no vacío de $\mathbb{Q}$. Si $a_1, \dots, a_n$ son elegidos de manera uniforme e independientemente desde A, entonces:

$$\Pr(p(a_1, \dots, a_n) = 0) \leq \frac{k}{|A|}$$

Suponga que la entrada del algoritmo aleatorizado está dada por los siguientes polinomios:

$$p(x_1, \dots, x_n) = \sum_{i=1}^{l} \prod_{j=1}^{r_i} \left(\sum_{k=1}^{n} a_{i,j,k} x_k + b_{i,j} \right)$$
$$q(x_1, \dots, x_n) = \sum_{i=1}^{m} \prod_{j=1}^{s_i} \left(\sum_{k=1}^{n} c_{i,j,k} x_k + d_{i,j} \right)$$

EquivPolAleatorizado$(p(x_1, \dots, x_n),\ q(x_1, \dots, x_n))$
    $K := 1 + \max\{r_1, \dots, r_\ell, s_1, \dots, s_m\}$
    $A := \{1, \dots, 100 \cdot K\}$
    sea $a_1, \dots, a_n$ una secuencia de números elegidos de
        manera uniforme e independiente desde A
    if $p(a_1, \dots, a_n) = q(a_1, \dots, a_n)$ **then return** sí
    else return no

Utilizando el lema de Schwartz-Zipel

Vamos a calcular la **probabilidad de error** del algoritmo:

- Si los polinomios $p(x_1, \dots, x_n)$ y $q(x_1, \dots, x_n)$ son equivalentes, entonces el algoritmo responde **SI** sin cometer error.

- Si los polinomios $p(x_1, \dots, x_n)$ y $q(x_1, \dots, x_n)$ NO son equivalentes, el algoritmo puede responder SI al escoger una secuencia de números $a_1, \dots, a_n$ desde A tales que $p(a_1, \dots, a_n) = q(a_1, \dots, a_n)$.
    - Donde $A = \{1, \dots, 100 \cdot K\}$.

  Esto significa que $(a_1, \dots, a_n)$ es una raíz del polinomio
  $$r(x_1, \dots, x_n) = p(x_1, \dots, x_n) - q(x_1, \dots, x_n)$$

  $r(x_1, \dots, x_n)$ NO es el polinomio nulo y es de grado t con $t < K$.
    - Dado que $K = 1 + \max\{r_1, \dots, r_l, s_1, \dots, s_m\}$.

Utilizando el lema de Schwartz-Zippel obtenemos:

$$\Pr(r(a_1, \dots, a_n) = 0) \leq \frac{t}{|A|} < \frac{K}{|A|} = \frac{K}{100 \cdot K} = \frac{1}{100}$$

Ignacio Méndez Pérez
IIC2283 - 2022 - 2

Demostración del lema de Schwartz-Zippel

Si es igual a $c \in (\mathbb{Q}\backslash\{0\})$, entonces es una constante, y el grado del polinomio debe ser 0.

Si $p(x_1, \ldots x_{n+1})$ en su forma canónica es igual a $c \in (\mathbb{Q} \setminus \{0\})$, entonces el lema se cumple trivialmente ya que:

$$\Pr(p(a_1, \ldots a_{n+1}) = 0) = 0$$

Suponemos entonces que $p(x_1, \ldots x_{n+1})$ en su forma canónica no es igual a $c \in \mathbb{Q}$.

- Puesto que además sabemos que $p(x_1, \ldots, x_{n+1})$ NO es nulo.

Por el lema de Schwartz

Tenemos que $p(x_1, \ldots, x_{n+1})$ en su forma canónica contiene un monomio de la forma:

$$c x_1^{l_1} x_2^{l_2} \cdots x_{n+1}^{l_{n+1}}$$

Porque es polinomio de varias variables.

donde $c \neq 0$ y $l_i > 0$ para algún $i \in \{1, \ldots, n+1\}$.

Al menos un l_i debe ser > 0.

Sin pérdida de generalidad suponemos que en el monomio anterior $l_1 > 0$
Tenemos que:

$$p(x_1, x_2, \ldots, x_{n+1}) = \sum_{i=0}^{k} x_1^i p_i(x_2, \ldots, x_{n+1})$$

Sacamos la variable factorizando.

donde cada $p_i(x_2, \ldots, x_{n+1})$ es un polinomio y al menos uno de ellos NO es nulo.

De lo contrario tendriamos el polinomio nulo.

Sea $l = \max\{\, i \in \{0, \ldots, k\} \mid p_i(x_2, \ldots, x_{n+1}) \text{ no es nulo } \}$

- Tenemos que $l > 0$ ya que supusimos que $l_1 > 0$.

l será el máximo exponente de la variable x_1.

Lema de Schwartz

Dado que el grado de $p(x_1, x_2, \ldots, x_{n+1})$ es k, tenemos que el grado de $p_l(x_2, \ldots, x_{n+1})$ es m con $m \leq k - l$.

p_l esta siendo multiplicado por x_1^l

Sea A un subconjunto finito y NO vacío de $\mathbb{Q}$, y sea $a_1, \ldots, a_{n+1}$ una secuencia de números elegidos de manera uniforme e independiente desde A.

Por hipótesis de inducción tenemos que:

$$\Pr(p_l(a_2, \ldots, a_{n+1}) = 0) \leq \frac{m}{|A|} \leq \frac{k-l}{|A|}$$

Lema de Schwartz

Si $p_l(a_2, \ldots, a_{n+1}) \neq 0$, entonces por definición de l tenemos que $q(x_1) = p(x_1, a_2, \ldots, a_{n+1})$ es un polinomio de grado l.
Por lo tanto:

x_1 es de grado l, y el resto son grado 0 ($a_i \in \mathbb{Q}$).

$$\Pr\left(\overbrace{p(a_1, a_2, \ldots, a_{n+1})}^{\text{grado } l} = 0 \mid p_l(a_2, \ldots, a_{n+1}) \neq 0\right) \leq \frac{l}{|A|}$$

Recordar: $P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)$

Concluimos que:

$$
\begin{aligned}
&\Pr(p(a_1, a_2, \dots, a_{n+1}) = 0) = \\
&\qquad \Pr(p(a_1, a_2, \dots, a_{n+1}) = 0 \mid p_l(a_2, \dots, a_{n+1}) = 0) \cdot \Pr(p_l(a_2, \dots, a_{n+1}) = 0) + \\
&\qquad \Pr(p(a_1, a_2 \dots, a_{n+1}) = 0 \mid p_l(a_2, \dots, a_{n+1}) \neq 0) \cdot \Pr(p_l(a_2, \dots, a_{n+1}) \neq 0) \leq \\
&\qquad \Pr(p_l(a_2, \dots, a_{n+1}) = 0) + \\
&\qquad \Pr(p(a_1, a_2 \dots, a_{n+1}) = 0 \mid p_l(a_2, \dots, a_{n+1}) \neq 0) \cdot \Pr(p_l(a_2, \dots, a_{n+1}) \neq 0) \leq \\
&\qquad\qquad \frac{k-l}{|A|} + \frac{l}{|A|} = \frac{k}{|A|}
\end{aligned}
$$

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

$$\mathcal{O}\left(n^{\frac{3}{4}} \cdot \log\left(n^{\frac{3}{4}}\right)\right) = \mathcal{O}\left(n^{\frac{3}{4}} \cdot \frac{3}{4} \cdot \log(n)\right) = \mathcal{O}\left(n^{\frac{3}{4}} \cdot \log(n)\right) = \mathcal{O}(n)$$

Me tomará $\mathcal{O}(n)$ por el tamaño de R.

d y u son mis elementos que escojo de tal manera que la mediana m este entre estos 2 elementos, y el tamaño de los elementos que hay entre d y u no sea muy grande.

Con "no muy grande" no referimos a que el tamaño sea $o(n/\log n)$, de esta manera, podemos ordenar este conjunto en $\mathcal{O}(n)$.

$\frac{1}{2} \cdot n^{\frac{3}{4}} =$ La mitad de R. Entonces mi lista S tendrá tamaño $2\sqrt{n}$.

Agregamos a S todos los elementos que están entre d y u.

Si esto ocurre, entonces la mediana estará afuera de S, y debemos retornar *sin_resultado*

Me tomará $\leq \mathcal{O}(n)$ por el tamaño de S.

Retornamos la mediana de S.

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.

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:

$$Y_1 = \left|\left\{i \in \left\{1, \dots, \left\lceil n^{\frac{3}{4}} \right\rceil\right\} \mid R[i] \leq m\right\}\right|$$

Cantidad de números de R que son **menores** o iguales a la mediana.

$$Y_2 = \left|\left\{i \in \left\{1, \dots, \left\lceil n^{\frac{3}{4}} \right\rceil\right\} \mid R[i] \geq m\right\}\right|$$

Cantidad de números de R que son **mayores** o iguales a la mediana.

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

Lema

CalcularMediana(L) retorna $sin\_resultado$ si y sólo si alguna de las siguientes condiciones se cumple:

1. $Y_1 < \left\lfloor \frac{1}{2} \cdot n^{\frac{3}{4}} - n^{\frac{1}{2}} \right\rfloor$

   La mediana está antes de d, y por lo tanto NO estará en S

2. $Y_2 \leq \left\lceil n^{\frac{3}{4}} \right\rceil - \left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil$

   La mediana está después de u, y por lo tanto NO estará en S

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

   De lo contrario nos podría llega a tomar más de $\mathcal{O}(n)$.

La desigualdad de Markov

Teorema

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

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

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

Demostración

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

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

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

Teorema

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

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

Demostración

$$\boxed{Var(X) = E\left[\left(X-E(X)\right)^2\right]}$$

Utilizando la desigualdad de Markov obtenemos:

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

Lema

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

Ignacio Méndez Pérez
IIC2283 - 2022 - 2

Demostración

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

$$X_i = \begin{cases} 1 & R[i] \leq m \\ 0 & R[i] > m \end{cases} \longrightarrow \boxed{\sim\text{Bernoulli}\left(\frac{1}{2} + \frac{1}{2 \cdot n}\right)}$$

Tenemos que:

$$Y_1 = \sum_{i=1}^{\lceil n^{3/4} \rceil} X_i \longrightarrow \boxed{\sim\text{Binomial}\left(\lceil n^{3/4} \rceil, \frac{1}{2} + \frac{1}{2 \cdot n}\right)}$$

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

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

De esto de deduce que:

$$E(X_i) = \frac{1}{2} + \frac{1}{2 \cdot n}$$

$$\boxed{Varianza_{binomial(n,p)} = np(1-p)}$$

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

Aquí hablamos de un único X_i, por eso $n = 1$.

Por lo tanto, tenemos que:

$$\begin{aligned} E(Y_1) &= E\left(\sum_{i=1}^{\lceil n^{3/4} \rceil} X_i\right) \\ &= \sum_{i=1}^{\lceil n^{3/4} \rceil} E(X_i) \\ &= \sum_{i=1}^{\lceil n^{3/4} \rceil} \left(\frac{1}{2} + \frac{1}{2 \cdot n}\right) \longrightarrow \boxed{\text{NO depende de } i} \\ &= \left\lceil n^{\frac{3}{4}} \right\rceil \cdot \left(\frac{1}{2} + \frac{1}{2 \cdot n}\right) \end{aligned}$$

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

Concluimos entonces que:

$$
\begin{aligned}
Var(Y_1) &= Var\left(\sum_{i=1}^{\lceil n^{3/4} \rceil} X_i\right) \\
&= \sum_{i=1}^{\lceil n^{3/4} \rceil} Var(X_i) \\
&= \sum_{i=1}^{\lceil n^{3/4} \rceil} \left(\frac{1}{4} - \frac{1}{4 \cdot n^2}\right) \qquad \longrightarrow \quad \boxed{\text{NO depende de } i} \\
&= \left\lceil n^{\frac{3}{4}} \right\rceil \cdot \left(\frac{1}{4} - \frac{1}{4 \cdot n^2}\right) \\
&\leq \frac{1}{4} \cdot \left\lceil n^{\frac{3}{4}} \right\rceil
\end{aligned}
$$

Tenemos que:

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

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

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

Lema

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

Lema

$$\Pr\left(\mathbf{Length}(S) > 4\cdot\left\lfloor n^{\frac{3}{4}} \right\rfloor\right) \leq 4\cdot n^{-\frac{1}{4}}$$

Demostración

Si $\mathbf{Length}(S) > 4\cdot\left\lfloor n^{\frac{3}{4}}\right\rfloor$, entonces al menos una de las siguientes condiciones debe ser cierta:

(a) $|\{i \in \{1,\dots,\mathbf{Length}(S)\} \mid S[i] > m\}| \geq 2\cdot\left\lfloor n^{\frac{3}{4}}\right\rfloor$

Mediana estará a la izquierda de S.

(b) $|\{i \in \{1,\dots,\mathbf{Length}(S)\} \mid S[i] < m\}| \geq 2\cdot\left\lfloor n^{\frac{3}{4}}\right\rfloor$

Mediana estará a la derecha de S.

Vamos a demostrar que la probabilidad de que (a) ocurra es menor o igual a $2\cdot n^{-\frac{1}{4}}$.

De la misma forma se demuestra que la probabilidad que (b) ocurra es menor o igual a $2\cdot n^{-\frac{1}{4}}$.

- De esto se concluye que $\Pr\left(\mathbf{Length}(S) > 4\cdot\left\lfloor n^{\frac{3}{4}}\right\rfloor\right) \leq 4\cdot n^{-\frac{1}{4}}$.

Suponga que (a) es cierto, y sea l la posición de u en la lista L ordenada.

- Tenemos que $l \geq \left\lceil \frac{n}{2} \right\rceil + 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor$

Dado que $u = R\left[\left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil \right]$ al menos $\left\lceil n^{\frac{3}{4}} \right\rceil - \left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil$ elementos de R deben estar en posiciones mayores o iguales a l en la lista L ordenada.

- No podemos asegurar que estos elementos están en posiciones mayores a l puesto que R puede tener elementos repetidos.

- Vamos a acotar superiormente la probabilidad de que esto ocurra para obtener una cota superior para la probabilidad de que (a) ocurra.

Parar cada $i \in \left\{1, \dots, \left\lceil n^{\frac{3}{4}} \right\rceil\right\}$, definimos una variable aleatoria W_i de la siguiente forma:

$$W_i = \begin{cases} 1 & si \text{ la posición de } R[i] \text{ en la lista } L \text{ ordenada} \\ & \text{es mayor o igual a } \left\lceil \dfrac{n}{2} \right\rceil + 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor. \\ 0 & \text{en otro caso.} \end{cases}$$

Además, definimos la variable aleatoria W como $\sum_{i=1}^{\lceil n^{3/4} \rceil} W_i$

Dado que $l \geq \left\lceil \frac{n}{2} \right\rceil + 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor$, tenemos que la probabilidad de que (a) ocurra es menor o igual a:

$$\Pr\left(W \geq \left\lceil n^{\frac{3}{4}} \right\rceil - \left\lceil \frac{1}{2} \cdot n^{\frac{3}{4}} + n^{\frac{1}{2}} \right\rceil \right)$$

Como L no contiene elementos repetidos obtenemos:

$$\begin{aligned} \Pr(W_i = 1) &= \frac{n - \left(\left\lceil \frac{n}{2} \right\rceil + 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor \right) + 1}{n} \\ &= \frac{\left\lfloor \frac{n}{2} \right\rfloor - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor + 1}{n} \\ &= \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} \end{aligned}$$

Tenemos entonces que:

$$E(W_i) = \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n}$$

$$Var(W_i) = \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} \cdot \left(1 - \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} \right)$$

Por lo tanto:

$$\begin{aligned}
E(W) &= E\left(\sum_{i=1}^{\lceil n^{3/4} \rceil} W_i \right) \\
&= \sum_{i=1}^{\lceil n^{3/4} \rceil} E(W_i) \\
&= \sum_{i=1}^{\lceil n^{3/4} \rceil} \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} \\
&= \left\lfloor n^{\frac{3}{4}} \right\rfloor \cdot \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n}
\end{aligned}$$

Pero tenemos que:

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

Concluimos que:

$$E(W) \leq \frac{1}{2} \cdot n^{\frac{3}{4}} - 2 \cdot n^{\frac{1}{2}} + 1$$

Para $i, j \in \left\{1, \dots, \left\lceil n^{\frac{3}{4}} \right\rceil \right\}$ tal que $i \neq j$ se tiene que W_i es independiente de W_j.

Concluimos entonces que:

$$\begin{aligned}
Var(W) &= Var\left(\sum_{i=1}^{\lceil n^{3/4} \rceil} W_i\right) \\
&= \sum_{i=1}^{\lceil n^{3/4} \rceil} Var(W_i) \\
&= \sum_{i=1}^{\lceil n^{3/4} \rceil} \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} \cdot \left(1 - \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n}\right) \\
&= \left\lceil n^{\frac{3}{4}} \right\rceil \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} \cdot \left(1 - \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n}\right) \\
&\leq \left\lceil n^{\frac{3}{4}} \right\rceil \cdot \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n}
\end{aligned}$$

Además, tenemos que:

$$\begin{aligned}
\left\lceil n^{\frac{3}{4}} \right\rceil \cdot \frac{\left\lceil \frac{n}{2} \right\rceil - 2 \cdot \left\lfloor n^{\frac{3}{4}} \right\rfloor}{n} &\leq \frac{1}{2} \cdot n^{\frac{3}{4}} - 2 \cdot n^{\frac{1}{2}} + 1 \\
&\leq \frac{1}{2} \cdot n^{\frac{3}{4}} + 1 \\
&\leq n^{\frac{3}{4}}
\end{aligned}$$

Concluimos que:

$$Var(W) \leq n^{\frac{3}{4}}$$

Finalmente tenemos que:

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

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

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

Concluimos que la probabilidad de que (a) ocurra es menor o igual a $2 \cdot n^{-\frac{1}{4}}$.

Recuerde que estamos considerando una lista $L[1 \dots n]$ de números enteros donde n es impar y mayor o igual a 2001.

Para la lista L demostramos lo siguiente:

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

Tenemos entonces que:

$$\Pr(\mathbf{CalcularMediana}(L) \text{ retorne } \mathit{sin\_resultado}) \leq 6 \cdot n^{-\frac{1}{4}}$$

Dado que $n \geq 2001$ concluimos que:

$$\begin{aligned} \Pr(\mathbf{CalcularMediana}(L) \text{ retorne } \mathit{sin\_resultado}) &\leq 6 \cdot 2001^{-\frac{1}{4}} \\ &< \frac{9}{16} \end{aligned}$$

Sea p la probabilidad de que $\mathbf{CalcularMediana}(L)$ retorne resultado.

- Tenemos que $p > \frac{1}{10}$

Sea T una variable aleatoria tal que para cada $i \geq 1$:

$$\Pr(T = i) = (1 - p)^{i-1} \cdot p$$

Vale decir, T representa el número de llamadas a $\mathbf{CalcularMediana}(L)$ hasta obtener un resultado.

Dado que T tiene una distribución geométrica de parámetro p, concluimos que

$$E(T) = \frac{1}{p} \leq \frac{1}{10}$$

Concluimos que en promedio se debe llamar 10 veces a $\mathbf{CalcularMediana}(L)$ para obtener la mediana de la lista L.