Chapter 7
Randomization

Algorithm Theory
WS 2018/19

Fabian Kuhn

Randomization

Randomized Algorithm:

- An algorithm that uses (or can use) random coin flips in order to make decisions

We will see: randomization can be a powerful tool to

- Make algorithms faster
- Make algorithms simpler
- Make the analysis simpler
  - Sometimes it's also the opposite…
- Allow to solve problems (efficiently) that cannot be solved (efficiently) without randomization
  - True in some computational models (e.g., for distributed algorithms)
  - Not clear in the standard sequential model

Randomized Quicksort

Quicksort:

S	v

$S_\ell < v$	v	$S_r > v$

function Quick (S: sequence): sequence;

{returns the sorted sequence S}

begin

  if $\#S \leq 1$ then **return** S

  else { choose pivot element v in S;

    partition S into S_ℓ with elements $< v$,

    and S_r with elements $> v$

    return | Quick(S_ℓ) | v | Quick(S_r) |

end;

Randomized Quicksort Analysis

Randomized Quicksort: pick uniform random element as pivot

Running Time of sorting n elements:

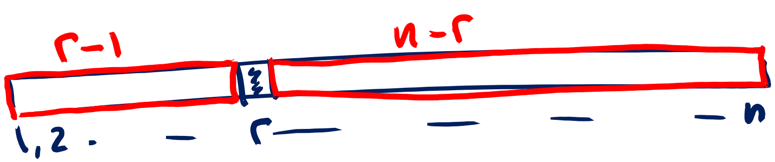
- Let's just count the number of comparisons
- In the partitioning step, all $n-1$ non-pivot elements have to be compared to the pivot
- **Number of comparisons:**

depends on partition

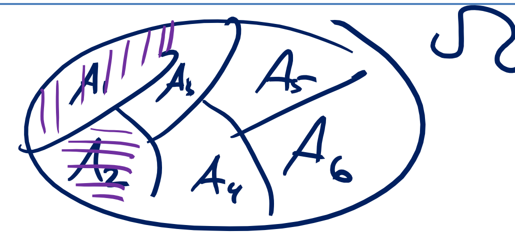
random variable

$$\boldsymbol{n-1\ +\ \#\text{comparisons in recursive calls}}$$

- **If rank of pivot is r:**
  recursive calls with $\boldsymbol{r-1}$ and $\boldsymbol{n-r}$ elements



Law of Total Expectation



- Given a random variable X and
- a set of events $A_1, \dots, A_k$ that partition Ω
  - E.g., for a second random variable Y, we could have
    $A_i \coloneqq \{\omega \in \Omega : Y(\omega) = i\}$

Law of Total Expectation

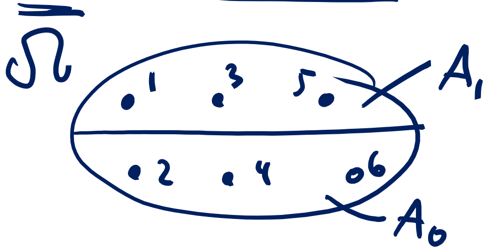
$\mathbb{R}$

$$\mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{P}(A_i) \cdot \mathbb{E}[X \mid A_i] = \sum_{y} \mathbb{P}(Y = y) \cdot \mathbb{E}[X \mid Y = y]$$

Example:

$\mathbb{E}[X] = 3{,}5$

- X: outcome of rolling a die
- $A_0 = \{X \text{ is even}\}, \quad A_1 = \{X \text{ is odd}\}$



$\mathbb{E}[X] = \mathbb{P}(A_0) \cdot \mathbb{E}[X \mid A_0] + \mathbb{P}(A_1) \cdot \mathbb{E}[X \mid A_1]$

$\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 3 = 3{,}5$

Randomized Quicksort Analysis

Random variables:

$\mathbb{E}[C] = \mathbb{E}[n-1 + C_\ell + C_r]$

- C: total number of comparisons (for a given array of length n)
- R: rank of first pivot $R \in \{1, \dots, n\}$
- C_ℓ, C_r: number of comparisons for the 2 recursive calls

$$\mathbb{E}[C] = n - 1 + \mathbb{E}[C_\ell] + \mathbb{E}[C_r]$$

Law of Total Expectation:

$$\mathbb{E}[C] = \sum_{r=1}^{n} \mathbb{P}(R = r) \cdot \mathbb{E}[C | R = r]$$

$$= \sum_{r=1}^{n} \mathbb{P}(R = r) \cdot (n - 1 + \underbrace{\mathbb{E}[C_\ell | R = r]}_{\text{sorting an array of length } r-1} + \underbrace{\mathbb{E}[C_r | R = r]}_{\text{sorting an array of length } n-r})$$

Randomized Quicksort Analysis

We have seen that: $\mathbb{P}(R=r) = \frac{1}{n}$

$$\underbrace{\mathbb{E}[C]}_{T(n)} = \sum_{r=1}^{n} \mathbb{P}(R=r) \cdot \left(n - 1 + \underbrace{\mathbb{E}[C_\ell | R = r]}_{T(r-1)} + \underbrace{\mathbb{E}[C_r | R = r]}_{T(n-r)}\right)$$

Define:

- $\boldsymbol{T(n)}$: expected number of comparisons when sorting n elements

$$\mathbb{E}[C] = T(n)$$
$$\mathbb{E}[C_\ell | R = r] = T(r-1)$$
$$\mathbb{E}[C_r | R = r] = T(n-r)$$

Recursion:

$$\boldsymbol{\underline{T(n)} = \sum_{r=1}^{n} \frac{1}{n} \cdot \big(n - 1 + T(r-1) + T(n-r)\big)}$$

$$\boldsymbol{T(0) = T(1) = 0}$$

Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof: (by induction on n)

$T(1) = 0$

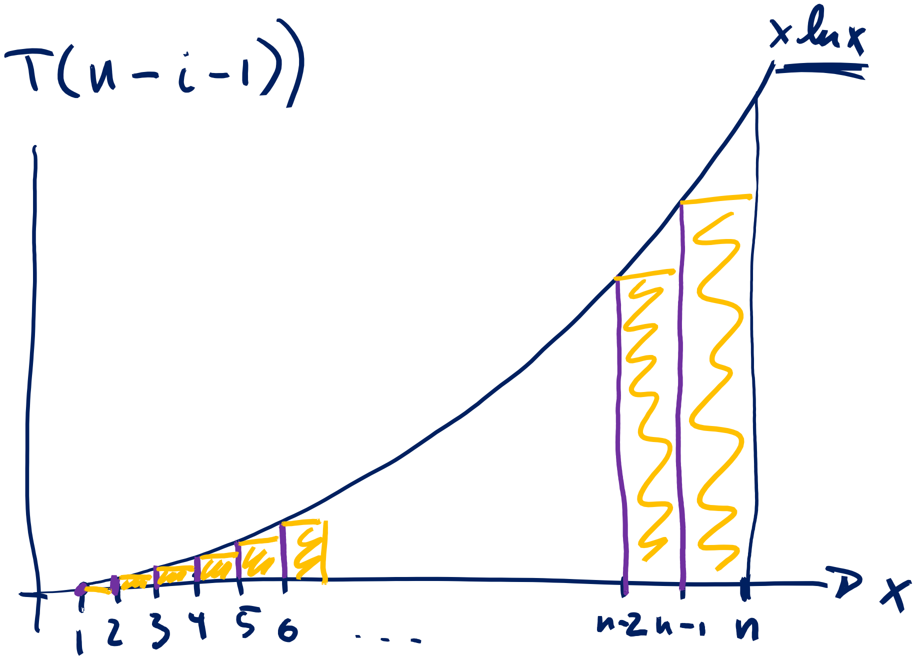
$$T(n) = \sum_{r=1}^{n} \frac{1}{n} \cdot \big(n - 1 + T(r-1) + T(n-r)\big), \qquad T(0) = 0$$

$$= n - 1 + \frac{1}{n} \cdot \sum_{i=0}^{n-1} \big(T(i) + T(n-i-1)\big)$$

$$= n - 1 + \frac{2}{n} \cdot \sum_{i=1}^{n-1} T(i)$$

$$\overset{\text{(I.H.)}}{\leq} n - 1 + \frac{4}{n} \cdot \sum_{i=1}^{n-1} i \cdot \ln(i)$$

$$\leq n - 1 + \frac{4}{n} \int_1^n x \ln(x)\, dx$$



Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof:

$$T(n) \leq n - 1 + \frac{4}{n} \cdot \int_1^n x \ln x \, dx$$

$$\int x \ln x \, dx = \frac{x^2 \ln x}{2} - \frac{x^2}{4}$$

$$T(n) \leq n-1 + \frac{4}{n}\left[\frac{n^2 \ln n}{2} - \frac{n^2}{4} + \frac{1}{4}\right]$$

$$= n - 1 + 2n\ln n - n + \frac{1}{n}$$

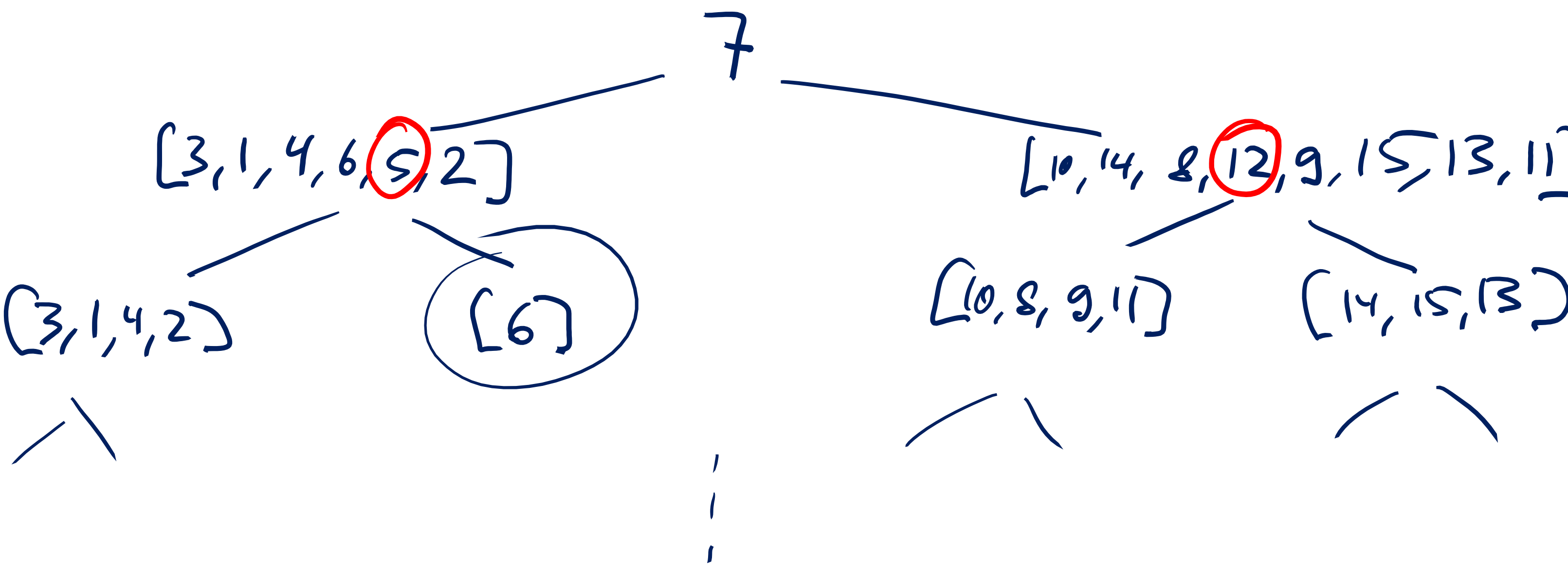
$$= 2n\ln n + \underbrace{\left(\frac{1}{n} - 1\right)}_{<0} < 2n\ln n \qquad \square$$

$$\Rightarrow E[C] \leq 2n\ln n$$

Alternative Analysis

Array to sort: [7 , 3 , 1 , 10 , 14 , 8 , 12 , 9 , 4 , 6 , 5 , 15 , 2 , 13 , 11]

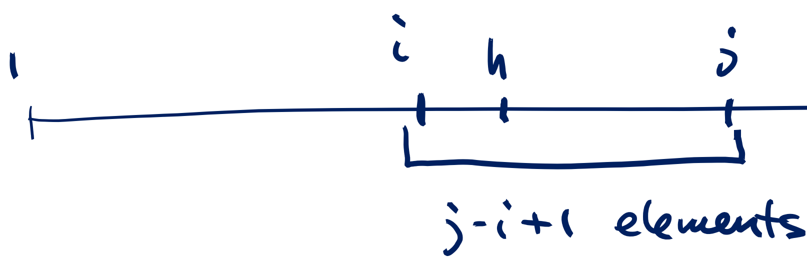
Viewing quicksort run as a tree:



Comparisons

- Comparisons are only between pivot and non-pivot elements
- Every element can only be the pivot once:
  → every 2 elements can only be compared once!

- W.l.o.g., assume that the elements to sort are $1,2,\dots,n$
- Elements i and j are compared if and only if either i or j is a pivot before any element h: $i < h < j$ is chosen as pivot
  - i.e., iff i is an ancestor of j or j is an ancestor of i



$$\mathbb{P}(\text{comparison betw. } i \text{ and } j) = \frac{2}{j-i+1}$$

Counting Comparisons

Random variable for every pair of elements (i, j):

$$X_{ij} = \begin{cases} 1, & \text{if there is a comparison between } i \text{ and } j \\ 0, & \text{otherwise} \end{cases}$$

$$P(X_{ij}=1) = \frac{2}{j-i+1} \qquad \mathbb{E}[X_{ij}] = \frac{2}{j-i+1}$$

Number of comparisons: X

$$X = \sum_{i<j} X_{ij}$$

- What is $\mathbb{E}[X]$?

Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof:

- **Linearity of expectation:**
  For all random variables $X_1, \dots, X_n$ and all $a_1, \dots, a_n \in \mathbb{R}$,

$$\mathbb{E}\left[\sum_i^n a_i X_i\right] = \sum_i^n a_i \mathbb{E}[X_i].$$

$$X = \sum_{i<j} X_{ij} \qquad \mathbb{E}[X] = \mathbb{E}\Big[\sum_{i<j} X_{ij}\Big]$$

$$= \sum_{i<j} \mathbb{E}[X_{ij}]$$

$$= \sum_{i<j} \frac{2}{j-i+1} = \underline{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}}$$

Randomized Quicksort Analysis

Theorem: The expected number of comparisons when sorting n elements using randomized quicksort is $T(n) \leq 2n \ln n$.

Proof:

$$\mathbb{E}[X] = 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{1}{j-i+1} = 2\sum_{i=1}^{n-1}\sum_{k=2}^{n-i+1}\frac{1}{k}$$

$$\leq 2\sum_{i=1}^{n-1}\underbrace{\sum_{k=2}^{n}\frac{1}{k}}_{H(n)-1}$$

$$= 2(n-1)\underbrace{(H(n)-1)}_{\leq \ln n}$$

$$< 2n \ln n \quad \square$$

Harmonic series:

$$H(n) := \sum_{i=1}^{n}\frac{1}{i}$$

$$H(n) \leq 1 + \ln n$$

Quicksort: High Probability Bound

- We have seen that the number of comparisons of randomized quicksort is $O(n \log n)$ in expectation.

- Can we also show that the number of comparisons is $O(n \log n)$ with high probability? $1 - \frac{1}{n^c}$

- **Recall:**

  On each recursion level, each pivot is compared once with each other element that is still in the same "part"

Counting Number of Comparisons

- We looked at 2 ways to count the number of comparisons
  - recursive characterization of the expected number
  - number of different pairs of values that are compared

Let's consider yet another way:

- Each comparison is between a pivot and a non-pivot
- How many times is a specific array element x compared as a non-pivot?

Value x is compared as a non-pivot to a pivot once in every recursion level until one of the following two conditions apply:

1. x is chosen as a pivot
2. x is alone

#comp. of x as non-pivot
= depth when x becomes
pivot / alone

Successful Recursion Level $K_1 = n$

- Consider a specific recursion level ℓ $\underbrace{[\quad x \quad\quad\quad]}_{K_\ell}$

- Assume that at the beginning of recursion level ℓ, element x is in a sub-array of length K_ℓ that still needs to be sorted.

- If x has been chosen as a pivot before level ℓ, we set $K_\ell := 1$

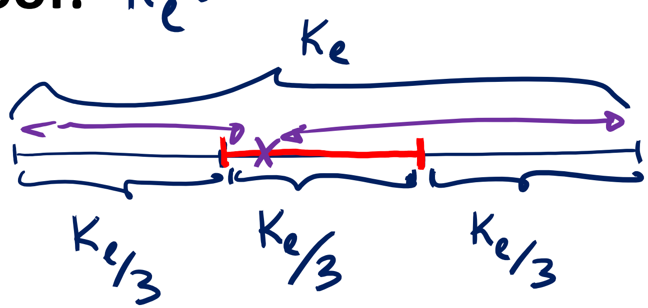
Definition: We say that recursion level ℓ is successful for element x iff the following is true:

$$K_{\ell+1} = 1 \quad \text{or} \quad K_{\ell+1} \leq \frac{2}{3} \cdot K_\ell$$

Successful Recursion Level

Lemma: For every recursion level ℓ and every array element x, it holds that level ℓ is successful for x with probability at least $1/3$, independently of what happens in other recursion levels.

Proof: $K_\ell > 1$



if pivot is in the middle part

$\Longrightarrow$ both remaining arrays have size $\leq \frac{2}{3} K_\ell \Longrightarrow K_{\ell+1} \leq \frac{2}{3} \cdot K_\ell$

$\Longrightarrow$ probability for this $\geq \frac{1}{3}$

$K_\ell = 1$

$\hookrightarrow K_{\ell+1} = 1$ ✓

Number of Successful Recursion Levels

Lemma: If among the first ℓ recursion levels, at least $\log_{3/2}(n)$ are successful for element x, we have $K_{\ell+1} = 1$.

Proof:

$K_1 = n$, $K_{i+1} \leq K_i$ if level i succ.: $K_{i+1} \leq \frac{2}{3} \cdot K_i$

$$K_{\ell+1} \leq n \left(\frac{2}{3}\right)^{\#\text{succ. levels}} \leq n \left(\frac{2}{3}\right)^{\log_{3/2} n} = n \cdot \frac{1}{n} = 1$$

Chernoff Bounds

- Let $X_1, \dots, X_n$ be independent 0-1 random variables and define $p_i := \mathbb{P}(X_i = 1)$.

  if $p_i = p$ for all i $X \sim \mathrm{Bin}(n,p)$
- Consider the random variable $X = \sum_{i=1}^{n} X_i$
- We have $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p_i$

$\mathbb{P}(X < \mu/2) < e^{-\mu/8}$

Chernoff Bound (Lower Tail):

$(\delta < 1)$

$$\boxed{\forall \delta > 0:\ \mathbb{P}(X < (1-\delta)\mu) < e^{-\delta^2\mu/2}}$$

Chernoff Bound (Upper Tail):

$$\forall \delta > 0:\ \mathbb{P}(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} < e^{-\delta^2\mu/3}$$

holds for $\delta \leq 1$

Chernoff Bounds, Example

Assume that a fair coin is flipped n times. What is the probability to have $p_i = \frac{1}{2}$ $\mu = \mathbb{E}[X] = n/2$

1. less than $n/3$ heads?

$$\mathbb{P}(X < n/3) = \mathbb{P}\left(X < \left(1 - \tfrac{1}{3}\right) \cdot \tfrac{n}{2}\right) < e^{-\frac{1}{9} \cdot \frac{1}{2} \cdot \frac{n}{2}} = e^{-n/36}$$

2. more than $0.51n$ tails?

$$\mathbb{P}\left(X > (1 + 0.02) \cdot \tfrac{n}{2}\right) < e^{-\frac{0.02^2}{3} \cdot \frac{n}{2}}$$

$$\boxed{e^{-\delta^2 \mu / 2}}$$

3. less than $n/2 - \sqrt{c \cdot n \ln n}$ tails?

$$\mathbb{P}\left(X < \left(1 - \frac{2\sqrt{c n \ln n}}{n}\right)\frac{n}{2}\right) \leq e^{-\frac{4 c \cdot n \cdot \ln n}{n^2 \cdot 2} \cdot \frac{n}{2}} = e^{-c \ln n} = \frac{1}{n^c}$$

w.h.p. #tails/#heads $= \frac{n}{2} \pm O(\sqrt{n \log n})$

Proof of Chernoff Bound

- Independent Bernoulli random variables $X_1, X_2, \dots, X_n$
- $\mathbb{P}(\boldsymbol{X_i} = \mathbf{1}) \geq \boldsymbol{p_i}, \boldsymbol{X} \coloneqq \sum_{i=1}^{n} \boldsymbol{X_i}, \boldsymbol{\mu} \coloneqq \sum_{i=1}^{n} \boldsymbol{p_i} \geq \mathbb{E}[\boldsymbol{X}]$

Chernoff Lower Tail: $\mathbb{P}(\boldsymbol{X} < (\mathbf{1} - \boldsymbol{\delta})\boldsymbol{\mu}) < \boldsymbol{e}^{-\delta^2 \mu / 2}$

Recall

- Markov Inequality: Given non-negative rand. var. $Z \geq 0$

$$\forall t > 0: \ \mathbb{P}(Z > t) < \frac{\mathbb{E}[Z]}{t}$$

$\mathbb{P}(Z > c \cdot \mathbb{E}[Z]) < \frac{1}{c}$

- Independent random variables Y, Z:

$$\mathbb{E}[Y \cdot Z] = \mathbb{E}[Y] \cdot \mathbb{E}[Z]$$

Proof of Chernoff Bound

$s > 0$

$a > b$

$e^{sa} > e^{sb}$

- $\mathbb{P}(X_i = 1) \geq p_i$, $X \coloneqq \sum_{i=1}^{n} X_i$, $\mu \coloneqq \sum_{i=1}^{n} p_i \geq \mathbb{E}[X]$

$X = \sum_{i=1}^{n} X_i$

Chernoff Lower Tail: $\mathbb{P}(X < (1-\delta)\mu) < e^{-\delta^2\mu/2}$

$$\mathbb{P}(Z > t) < \frac{\mathbb{E}[Z]}{t}$$

$$\mathbb{P}(X < (1-\delta)\mu) = \mathbb{P}(-X > -(1-\delta)\mu)$$

$$= \mathbb{P}(\underbrace{e^{-sX}}_{Z} > e^{-s(1-\delta)\mu})$$

$$\overset{\text{(Markov)}}{<} \frac{\boxed{\mathbb{E}[e^{-sX}]}}{e^{-s(1-\delta)\mu}}$$

$$\mathbb{E}[e^{-sX}] = \mathbb{E}[e^{-s\sum X_i}] = \mathbb{E}\Big[\prod_{i=1}^{n} e^{-sX_i}\Big] \overset{\text{indep. of } X_i}{=} \prod_{i=1}^{n} \mathbb{E}[e^{-sX_i}]$$

$$\mathbb{E}[e^{-sX_i}] = \mathbb{P}(X_i = 1)\cdot e^{-s\cdot 1} + \mathbb{P}(X_i = 0)\cdot e^{-s\cdot 0} = \mathbb{P}(X_i = 1)(\underbrace{e^{-s} - 1}_{<0}) + 1$$

$$\leq 1 + p_i(e^{-s} - 1)$$

Proof of Chernoff Bound $\underline{s > 0}$

- $\mathbb{P}(X_i = 1) \geq p_i$, $X \coloneqq \sum_{i=1}^{n} X_i$, $\mu \coloneqq \sum_{i=1}^{n} p_i \geq \mathbb{E}[X]$

Chernoff Lower Tail: $\mathbb{P}(\boldsymbol{X} < (\mathbf{1} - \boldsymbol{\delta})\boldsymbol{\mu}) < \boldsymbol{e}^{-\boldsymbol{\delta}^2\boldsymbol{\mu}/2}$

$$\mathbb{P}(X < (1-\delta)\mu) < \frac{\mathbb{E}[e^{-sX}]}{e^{-s(1-\delta)\mu}} \leq e^{\mu(e^{-s}-1+s(1-\delta))}$$

$$\mathbb{E}[e^{-sX}] = \prod_{i=1}^{n} \mathbb{E}[e^{-sX_i}] \leq \prod_{i=1}^{n} (1 + p_i(e^{-s}-1))$$

$$\leq \prod_{i=1}^{n} e^{p_i(e^{-s}-1)}$$

$1 + x \leq e^x$

$$= e^{\sum_{i=1}^{n} p_i(e^{-s}-1)} = e^{\mu(e^{-s}-1)}$$

set $s = \delta$

$$\mathbb{P}(X < (1-\delta)\mu) < e^{\mu(e^{-\delta}-1+\delta(1-\delta))} = e^{\mu(e^{-\delta}-1+\delta-\delta^2)}$$

Proof of Chernoff Bound

- $\mathbb{P}(X_i = 1) \geq p_i$, $X \coloneqq \sum_{i=1}^{n} X_i$, $\mu \coloneqq \sum_{i=1}^{n} p_i \geq \mathbb{E}[X]$

Chernoff Lower Tail: $\mathbb{P}(\boldsymbol{X} < (\boldsymbol{1}-\boldsymbol{\delta})\boldsymbol{\mu}) < \boldsymbol{e^{-\delta^2\mu/2}}$

$$\mathbb{P}(X < (1-\delta)\mu) < e^{\mu(e^{-\delta}-1+\delta-\delta^2)}$$

$$= e^{\mu(\cancel{1}\cancel{-\delta}+\frac{\delta^2}{2}\cancel{-1}\cancel{+\delta}-\delta^2)}$$

$$= e^{\mu(-\frac{\delta^2}{2})}$$

$\square$

$$e^{-\delta} = 1-\delta+\frac{\delta^2}{2}\underbrace{- \ldots}_{<0}$$

$$\overset{(\delta\leq 1)}{\leq} 1-\delta+\frac{\delta^2}{2}$$

Number of Comparisons for x

Lemma: For every array element x, with high probability, as a non-pivot, x is compared to a pivot at most $O(\log n)$ times.

Proof:

k $\quad \frac{k}{3}$ $\quad \frac{k}{6}$

$$e^{-k/8} = \frac{1}{n^c}$$

$k = 8c \ln n$ $\quad \frac{1}{n^{c-1}}$