Moments

Name: Chong-kwon Kim

Key Definitions

- Definition: **k-th moment** of X ≡ E[X^k]

- Definition: **Variance**

  Var[X] = E[$(X - E[X])^2$]

  = E[X^2] - $E[X]^2$

- Definition: **Standard deviation**

  σ[X] = $\sqrt{\text{Var[X]}}$

Variance: Binomial

KENTECH Korea Institute of Energy Technology AI

- X: Binomial r v with n, p
- E$[X^2]$ = $\sum_{j=0}^{n} \binom{n}{j}\, p^j \cdot (1-p)^{n-j} \cdot\ j^2$

$= \sum_{j=0}^{n} \frac{n!}{(n-j)!j!}\, p^j \cdot (1-p)^{n-j} \cdot ((j^2 - j) + j)$

$\vdots$

$= n(n-1)\ p^2 + np$

- Var[X] =$E[X^2] - (E[X])^2$

$= np(1-p)$

- More simply, X is sum of n independent Bernoulli r v
- Var[X] = Var[$\sum_i$ Xi]

$= \sum_i Var[X_i]$) X_i, X_j are mutually independent

$= \underline{n}\underline{p(1-p)}$ $p \cdot (1-p)$

Var[$\sum_i$Xi] ?= $\sum_i Var[X_i]$
True only when X_i are mutually independent
Will prove soon

Variance: Geometric R V

KENTECH Korea Institute of Energy Technology AI

- Y: Geometric random variable
  - We know $\boxed{E[Y] = 1/p}$
  - From $\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$, we obtain

$(\frac{1}{1-x})^2 = \sum_{i=1}^{\infty} i \cdot x^{i-1}$ Taking the derivative on both sides

$2 \cdot (\frac{1}{1-x})^3 = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot x^{i-2}$

  - $E[Y^2] = \sum_{i=1}^{\infty} p(1-p)^{i-1} \cdot i^2$

$= \frac{2-p}{p^2}$

  - $Var[Y] = E[Y^2] - (E[Y])^2$

$= \frac{1-p}{p^2}$

Independence

- Note that E[X+Y] = E[X] + E[Y] holds even if X and Y are dependent

- How about **E[X·Y] ≡ E[X]·E[Y] ?**
  - True only if X and Y are independent
  - Counter example:
    - Flip two coins
    - X: Indicator function of first coin = heads
    - Y: Sum of heads in two coin flips
    - E[X] = ½
    - E[Y] = 1
    - E[X · Y] = $\sum_i \sum_j$ i · j Pr((X=i) ∩ (Y=j))
  - Independent
    - X: Indicator function of first coin = heads
    - Y: Indicator function of second coin = heads

Independence → E[X·Y] = E[X]·E[Y]

KENTECH Korea Institute of Energy Technology AI

- Theorem: If X and Y are independent,

  then E[X·Y] = E[X]·E[Y]

- Proof
  - E[X·Y] = $\sum_i \sum_j$ i · j Pr((X=i) ∩ (Y=j))

    = $\sum_i \sum_j$ i · j Pr(X=i) · Pr(Y=j)

    =

Covariance

KENTECH Korea Institute of Energy Technology AI

- **Covariance** of two r v X and Y

  Cov(X, Y) = E[(X-E[X])·(Y-E[Y])]

- Theorem: Var[X+Y] = Var[X] + Var[Y] + 2 · Cov(X,Y)
- Proof:
  - Var[X+Y] = E[$((X+Y)-(E[X]+E[Y]))^2$]

    = E[$(X-E[X])+(Y-E[Y])^2$]

- If X and Y are independent,

  then Cov(X, Y) = 0 and

  Var[X+Y] = Var[X] + Var[Y]

- Proof
  - Cov(X,Y) = E[$(X-E[X])\cdot(Y-E[Y])$]

    = E[$X\cdot Y-X\cdot E[Y]-Y\cdot E[X]+E[X]\cdot E[Y]$]

    = 0

Moment Generating Function

KENTECH Korea Institute of Energy Technology AI

- **Function that can generate moments**

  M_X(t) = E[e^{tX}] = $\sum_i e^{tx_i}\cdot \Pr(X=x_i)$ $\frac{d M_X(t)}{dt} = \Sigma_i x_i e^{tx_i} P_r(X=x_i)$

  $\frac{d^2 M_X(t)}{dt^2} = \Sigma_i x_i^2 \overset{1}{\boxed{e^{tx_i}}} P_r(X=x_i)$

  $= \Sigma_i x_i^2 P_r(X=x_i) = E[X^2]$

- E[X^n] = $M_X{}^{(n)}(0)$
  - where $M_X{}^{(n)}$(t) is the *n* th derivative of M_X(t)

- Proof

  True! Accept w/o proof
  - If we (can) exchange expectation and differentiation operands
  - Then, $M_X{}^{(n)}$(t) = E[$X^n\cdot e^{tX}$]
  - At t=0, $M_X{}^{(n)}$(0) = E[X^n]

MGF - Example

- Geometric Distribution, Pr(X=k) = $(1-p)^{k-1} \cdot p$
  - $M_X(\mathrm{t}) = \mathrm{E}[e^{tX}]$ $\quad (1+\sum_{k=1}^{\infty}\boxed{}-1)\frac{p}{1-p}$

    $= \sum_{k=1}^{\infty}(1-p)^{k-1} \cdot p \cdot e^{tk}$ $\quad = \sum_{k=1}^{\infty}[(1-p)\cdot e^{t}]^{k} \cdot p \frac{1}{1-p}$ $\quad +1$

    $= \frac{p}{1-p}((1-(1-p)e^{t})^{-1}-1)$

  - ${M_X}^{(1)}(t) = p(1-(1-p)e^{t})^{-2}e^{t}$ $= \frac{1}{p}$
  - ${M_X}^{(2)}(t) = 2p(1-p)(1-(1-p)e^{t})^{-3}e^{2t} + p(1-(1-p)e^{t})^{-2}e^{t}$

Properties

- If two random variables X and Y have the same MGF, then X ≡ Y

Uniqueness

- If X and Y are independent r.v., then

  $M_{X+Y}(\mathrm{t}) = M_X(\mathrm{t})\, M_Y(\mathrm{t})$

- Proof
  - $M_{X+Y}(\mathrm{t}) = \mathrm{E}\left[e^{t(X+Y)}\right]$

    $= E[e^{tX} \cdot e^{tY}]$

    $= E[e^{tX}] \cdot E[e^{tY}]$ **X, Y independent → e^{tX}, e^{tX} independnet**

    $= M_X(\mathrm{t}) \cdot M_Y(\mathrm{t})$

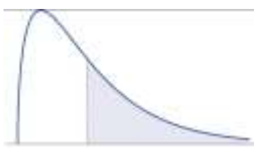
Bounds

Name: Chong-kwon Kim

Bounds

KENTECH Korea Institute of Energy Technology AI

- We are interested in "**Tail Bound**", like Pr(X $\geq$ a)



- Markov
  - Only E[X] is given

a

- Chebyshev
  - E[X] and Var[X] are known
- Chernoff
  - MGF based

Markov's Inequality

KENTECH Korea Institute of Energy Technology AI

- Let X assumes only non-negative values.

  For any a > 0, **Pr (X ≥ a) ≤ $\frac{E[X]}{a}$**

Markov (1856-1922) was a Russian Mathematician known for Markov chain/process
Student of Chebyshev at St. Petersburg Univ.

- Proof

  $\Sigma_{x_i<a} x_i p_i + \Sigma_{x_i \ge a} \underbrace{\boxed{x_i}}_{a} p_i \ge$

  - E[X] = $\sum_i x_i \cdot \Pr(x_i)$

    $\ge \sum_{i:\, x_i<a} x_i \cdot \Pr(x_i) + \sum_{i: x_i \ge a} a \cdot \Pr(x_i)$

    $\ge \sum_{i: x_i \ge a} a \cdot \Pr(x_i) = a \cdot \Pr(X \ge a\,)$

- Example

  Bin $(n, \frac{1}{2})$ $E[X] = \frac{n}{2}$

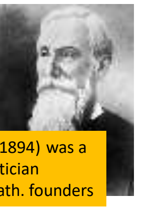
  - X: # heads in n fair coin flips (note X ≥ 0)
  - Probability of obtaining ≥3n/4 heads from n coin flips
  - E[X] = n/2
  - Pr(X ≥ 3n/4) ≤ (n/2) ÷ (3n/4) = 2/3

  ↳ $\Pr(X \ge \boxed{\frac{3n}{4}}) \le \frac{E[X]}{a} = \frac{\frac{n}{2}}{\frac{3n}{4}} = \frac{2}{3}$

Is Markov bound tight? ➔ YES
Ex. 3.16

Chebyshev's Inequality

KENTECH Korea Institute of Energy Technology AI



- Also known as **Weak Law of Large Number**
- For any a > 0,

  Note: Non-negativity restriction on X is removed

  $\Pr(|X - E[X]| \ge a) \le \frac{Var[X]}{a^2}$

  $\Pr(X \ge a) \le \frac{E[X]}{a}$

- Proof

  Random variable

  - $\Pr(|X - E[X]| \ge a) = \Pr((X - E[X])^2 \ge a^2)$
  - Applying the Markov's Inequality, we obtain
  - $\Pr((X - E[X])^2 \ge a^2) \le \frac{E[(X - E[X])^2]}{a^2} = \frac{Var[X]}{a^2}$

Chebyshev (1821-1894) was a Russian Mathematician
One of Russian math. founders

Weak Law of Large Number
Ex. 3.25

- **Corollary**: For any $t > 1$

  $\Pr(|X - E[X]| \ge \underbrace{t \cdot \sigma[X]}_{a}) \le \frac{1}{t^2}$

  $\Pr(|X - E[X]| \ge t \cdot E[X]) \le \frac{Var[X]}{t^2 (E[X])^2}$

Corollary

KENTECH Korea Institute of Energy Technology AI

- **Example**
  - X: # heads in n coin flips
  - Probability of obtaining 3n/4 heads from n coin flips
  - E[X] = n/2, Var[X] = n/4
  - Pr(X ≥ 3n/4) = Pr(X-n/2 ≥ n/4)

$$\le \Pr(|X-n/2| \ge n/4)$$
$$\le \frac{\mathrm{Var}[X]}{(n/4)^2} = 4/n$$

$X \sim Bin(n, \frac{1}{2})$ $E[X] = \frac{n}{2}$ (np) $Var[X] = \frac{n}{4}$ $(np(1-p))$

$\Pr(X \ge \frac{3n}{4})$ ① Markov: $\frac{E[X] = \frac{n}{2}}{\frac{3n}{4}} = \frac{2}{3}$

$\boxed{|X - E[X]| \ge a}$ 이 형식이어야 함

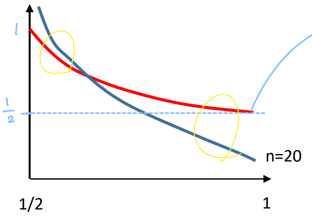
$-\frac{n}{2}$ $-\frac{n}{2}$

$|X - \frac{n}{2}| \ge \frac{n}{4}$
$(x - \frac{n}{2})^2 \ge \frac{n}{4}$
$(x - \frac{n}{2}) \le$

$= 4\% = 0.04$ (if $n = 100$)

- **Compare to the Markov bound (2/3)**

As a function of k and n where Pr(X ≥ kn)



① Markov $\le \frac{\frac{n}{2}}{kn} = \frac{1}{2k}$

② $\Pr(X \ge kn) \le \Pr(|X - \frac{n}{2}| \ge (k - \frac{1}{2})n)$

$\le \frac{n/4}{(k-\frac{1}{2})^2 n^2} = \frac{1}{4n(k-\frac{1}{2})}$ → $k = \frac{1}{2}$ → infinite

Chernoff Bounds

KENTECH Korea Institute of Energy Technology AI

- **Apply Markov inequality to** e^{tX} $\boxed{\Pr(X \ge a) \le \frac{E[X]}{a}}$
- **From Markov inequality, for any t >0**
  - Pr (X ≥ a) = Pr($e^{tX} \ge e^{ta}$) $\le \frac{E[e^{tX}]}{e^{ta}}$
    $(t>0) \to tX \ge ta \to e^{tX} \ge e^{ta}$
  - In particular, Pr (X ≥ a) ≤ $\min_{t>0} \frac{E[e^{tX}]}{e^{ta}}$

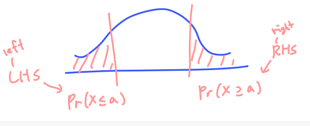
MGF

Find appropriate t that minimizes the bound

Chernoff (1923~) is an American mathematician Professor at MIT & Harvard

- **Similarly, for t < 0**
  - Pr (X ≤ a) = Pr($e^{tX} \ge e^{ta}$) $\le \frac{E[e^{tX}]}{e^{ta}}$
  - Hence, Pr (X ≤ a) ≤ $\min_{t<0} \frac{E[e^{tX}]}{e^{ta}}$

Bound for L tail as well as R tail



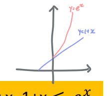
Chernoff Bound for Poisson Trials

KENTECH Korea Institute of Energy Technology AI

Bernoulli trial: Each experiment has the same distribution

- **Poisson trial** $M_{X+Y}(t) = M_X(t)\cdot M_Y(t)$
  - A sequence of experiments(trials) each of which has different distribution
  - Let $X_1, X_2,.., X_n$ be a sequence of **independent** Poisson trials with $\Pr(X_i$=1) = p_i
  - X = $X_1 + X_2 +.. + X_n$
  - Let μ = E[X] =E[$\sum_{i=1}^{n} X_i$] = $\sum_{i=1}^{n} p_i$
- Find the bounds of Pr(X ≥ (1 + δ)μ) and Pr(X ≤ (1−δ)μ)
- First derive M_X(t)
  - MGF of X_i
  - M_{X_i}(t) = E[e^{tX_i}] = $p_i \cdot e^t + (1 - p_i)$ = 1+$p_i \cdot (e^t - 1)$ $1+x \le e^x$

    $\le e^{p_i \cdot (e^t - 1)}$



For any x, 1+x ≤ e^x → 항상

Think it as x

  - M_X(t) = $\prod_{i=1}^{n} M_{X_i}$(t)

    $\le \prod_{i=1}^{n} e^{p_i \cdot (e^t - 1)}$ = exp{$\sum_{i=1}^{n} p_i \cdot (e^t - 1)$}

    = exp{μ · $(e^t - 1)$}

Chernoff Bound for Poisson Trials

KENTECH Korea Institute of Energy Technology AI

- Now prove
  1. For any δ > 0, Pr(X ≥ (1 + δ)μ) < $\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$
  2. For 0 < δ ≤ 1, Pr(X ≥ (1 + δ)μ) ≤ $e^{-\frac{\mu\delta^2}{3}}$
  3. For R ≥ 6μ, Pr(X > R) ≤ 2^{-R}
- Proof
  - From Markov's Inequality,

    $\Pr(X \ge (1+\delta)\mu) = \Pr(\underbrace{e^{tX}}_{X} \ge \underbrace{e^{t(1+\delta)\mu}}_{a}) \le \frac{E[X]}{a}$

    $\le \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}}$

    $\le \frac{\exp\{(e^t - 1)\cdot\mu\}}{e^{t(1+\delta)\mu}}$
  - For any $\delta > 0$, find t that minimizes $(e^t - 1)\cdot\mu - t(1+\delta)\mu$
  - Min. at $t = \ln(1+\delta) > 0$

    ➔ $\Pr(X \ge (1+\delta)\mu) \le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

Chernoff Bound for Poisson Trials

- **Proof of 2** **(For $0 < \delta \le 1$, $\Pr(X \ge (1+\delta)\mu) \le e^{-\frac{\mu\delta^2}{3}}$)**
  - For $0 < \delta \le 1$, show that $\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right) \le e^{-\frac{\delta^2}{3}}$
  - Taking the logarithm to both sides and define $f(\delta)$ as

$\Pr(X \ge (1+\delta)\mu)$
$\le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

$$f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3}$$

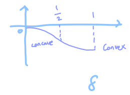
$f(0) = 0$
$\rightarrow f(\delta) \le 0$ for $0 \le \delta \le 1$

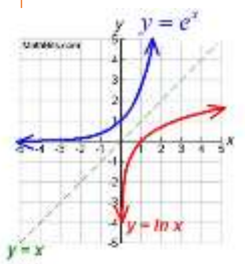
$$f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta$$

$f'(0) = 0, f'(1) < 0$
$f'(\delta) \le 0$ for $0 \le \delta \le 1$

$$f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$$

$\boxed{f''(\delta) < 0}$ for $0 \le \delta < 1/2$,
$f''(\delta) > 0$ for $\delta > 1/2$





Chernoff Bound for Poisson Trials

- **Proof of 3** (**For $R \ge 6\mu$, $\Pr(X > R) \le 2^{-R}$**)
  - $R = (1+\delta)\mu$
  - $R \ge 6\mu \rightarrow \delta \ge 5$
  - $\Pr(X \ge (1+\delta)\mu) \le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

$$\le \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}$$

$$\le \left(\frac{e}{6}\right)^{R}$$

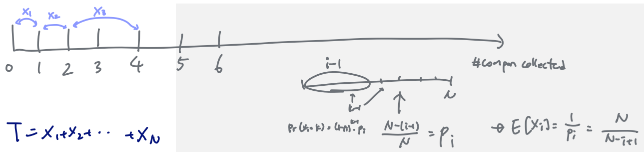
$$\le 2^{-R}$$

Coupon Collector's Problem T: # coupons

모든 쿠폰 종류를 모을 때 까지 모음. → E[T]??

X_i : Given that (i-1) types of coupons are collected, how many more to collect to obtain the i-th type.

i번째 새로운 type을 얻을 확률



$\to T = X_1 + X_2 + \cdots + X_N$

X_i = Geometric RV with $p_i = (1 - \frac{i-1}{N}) = \frac{N-i+1}{N}$

$E[X_i] = \frac{1}{p_i} = \frac{1}{N-i+1}$

$E[T] = \sum_i E[X_i] = \sum_i \frac{1}{N-i+1} = N \cdot \sum_i \frac{1}{i}$ → 30개 정도 걸림

하범 1이네

Bounds

Name: Chong-kwon Kim

Example – Coupon Collection

- Let X: Time to collect all n types of coupon
- X = $X_1 + X_2 + \ldots + X_n$ (Xi is time to collect i-th coupon types after (i-1) coupon types are collected)

- Xi: Geometric r. v. with pi = (1 - (i-1)/n)
  - ➔ E[X] = $n \cdot H_n$
  - ➔ Var[X] = $\sum_{i=1}^{n} Var[Xi]$

$$\leq \sum_{i=1}^{n} (n/(n-i+1))^2$$

$$\leq \frac{\pi^2 \cdot n^2}{6}$$

E[Xi] = n/(n-i+1)

Var[Xi] = $\frac{(1-p_i)}{p_i^2} \leq \frac{1}{p_i^2}$

$\sum_{i=1}^{n} (1/i)^2 = \frac{\pi^2}{6}$

- Find Markov and Chebyshev bounds of Pr(X ≥ $2n \cdot H_n$)

$\Pr(X \geq \underbrace{2n \cdot H_n}_{a}) \leq \frac{E[X]}{2n \cdot H_n} = \frac{1}{2}$ ← $n \cdot H_n$

$\Pr(X \geq 2n \cdot H_n)$

$= \Pr(X - nH_n \geq nH_n) \leq \frac{\frac{\pi^2}{6}}{(nH_n)^2} = \frac{\pi^2}{6n^2H_n^2}$ ← $\frac{\pi^2}{6} = Var(X)$

Example – Coin Flips Revisited

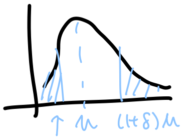
- We proved that, for $0 < \delta \leq 1$, $\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\mu \cdot \delta^2}{3}}$

X should be sum of Poisson trials

- Also, it can be shown that $\Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu \cdot \delta^2}{3}}$

➔ $\boxed{\Pr(|X-\mu| \geq \delta \cdot \mu) \leq 2e^{-\frac{\mu \cdot \delta^2}{3}}}$

Refer to Theorem 4.5 & Corollary 4.6



↳ 더 tight 함.

- X: # heads in n coin flip

$E[X] = \frac{n}{2}$

$Var[X] = \frac{n}{4}$

- Find bounds of $\Pr(|X - n/2| \geq n/4)$
  - Markov: $\Pr((X - n/2) \geq n/4) =$ → $\Pr(X \geq \frac{3}{4}n) \leq \frac{E[X]}{\frac{3}{4}n} = \frac{\frac{n}{2}}{\frac{3}{4}n} = \frac{2}{3}$
  - Chebyshev: $\Pr(|X - n/2| \geq n/4) =$ → $\leq \frac{Var(X)}{(\frac{n}{4})^2} = \frac{\frac{n}{4}}{(\frac{n}{4})^2} = \frac{4}{n}$
  - Chernoff: $\Pr(|X - n/2| \geq n/4) =$

$\delta = \frac{1}{2}, \mu = \frac{n}{2} \Rightarrow 2 \cdot e^{-\frac{\frac{n}{2} \cdot \frac{1}{4}}{3}} = 2 \cdot e^{-\frac{n}{24}}$

⇒ n이 큰 경우, chernoff 성능이 가장 좋다.

Selection Problem

- Problem: Given an input of N distinct numbers, find **i-th largest** number
- **Median**: $\lceil N/2 \rceil$ - th or $\lceil (N+1)/2 \rceil$ - th largest number
- Complexity of find minimum (or maximum) number

➔ O(N)

- What is the complexity of finding the median?
  - Obviously, we can do in O(N lnN)
- Any selection algorithm with **O(N)?**

Randomized Selection

- Similar to **randomized QuickSort**

- Pick a pivot number randomly
- Partition the input into two subsets,S1 and S2, such that all in S1 are smaller than the pivot and all in S2 are larger than the pivot
- Pick S1 or S2 and repeat the procedure recursively
  - ➔ O(N)?

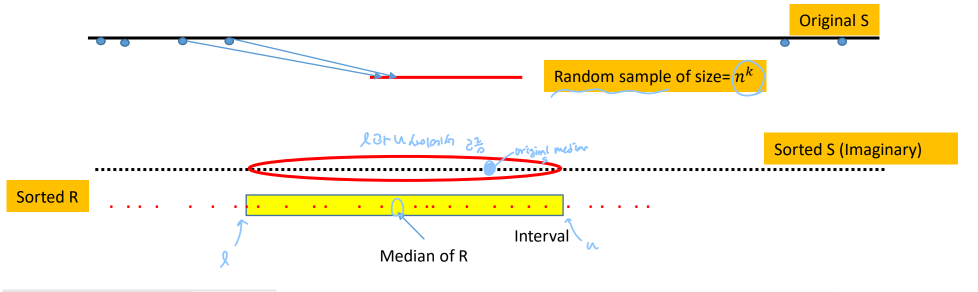
- Let T(N): # comparison to find the median
  - Then T(N) ≤ 1/N ·($\sum_{k=1}^{n-1} T(\max(k, N-k))$)
  - T(N) = O(N) **Refer to CLRS**

Quick Sort

$\Sigma\, T(k) + T(N-k)$

Randomized Median Algorithm

- Sketch of the algorithm
  - Given an original set S (size: n objects)
  - Generate a random sample (say R) of a properly small size, say $\sqrt{n}$, or n^k (k < 1)
    - Sort R (Complexity = O($n^k \cdot \log n^k$))
    - Fix an short interval (say I) that contains the median of R
  - Now, collects the objects that belong to the interval (Complexity??)
  - Sort the selected objects



Example

- S = {17, 7, 14, 6, 1,19, 3, 4, 7, 11, 18, 12, 21, 9, 5, 10, 2, 19, 8, 13, 16}
- Let R1 = {17, 7, 14, 6, 1,19, 3}, R2 = {17, 14, 19, 7, 18, 12, 21}

- Sorted S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21}
- Sorted R1 = {1, 3, 6, 7, 14, 17, 19}
- Sorted R2 = { 7, 12, 14, 17, 18 19, 21}

Randomized Median Algorithm

Input: A set S of n elements

Output: Median (m) of S

1. Construct a multi-set R of $\lceil n^{3/4} \rceil$ elements from S, each element chosen independently and uniformly at random with replacement
2. Sort R $O(n^{\frac{3}{4}} \log n^{\frac{3}{4}}) \le O(n)$
3. Let d and u be the $\lfloor \frac{1}{2} n^{3/4} - \sqrt{n} \rfloor$ and $\lceil \frac{1}{2} n^{3/4} + \sqrt{n} \rceil$-th elements, respectively, in sorted R
4. Compare every element in S to d and u. Construct a set C with elements in [d, u] and count l_d and l_u, the number of elements smaller than d and greater than u, respectively
5. If l_d > n/2 or l_u > n/2 ➔ FAIL
6. If $|C| \le 4\, n^{3/4}$, then sort C,

   OW FAIL

M is ($\lfloor \frac{n}{2} \rfloor - l_d$+1) –th element in sorted C

Analysis

Randomized Median Algorithm

$Pr(\underset{X \le 0}{\bigcirc}) \le \boxed{}$

high probability definition에 맞는가

- With **high probability**
  - Probability at least 1-O(1/ n^c) for some c > 0
  - 1. m is between d and u ← Condition of step 5 SUCCESS
  - 2. |C| is not greater than 4 $n^{3/4}$ ← Condition of step 6 SUCCESS

- Easy to prove that / Obviously
  - If the algorithm does not FAIL, then it finds the median of S

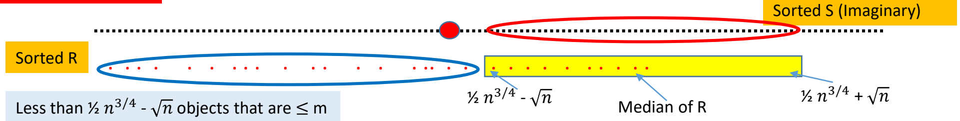
- Need to prove
  1. Randomized median algorithm terminates in **linear time O(N)**
  2. **SUCCESS with high probability**

- The algorithm FAILs if any one of following events occur
  - E1: Y1 = |{r ∈ R | r ≤ m}| < ½ $n^{3/4}$ - $\sqrt{n}$
  - E2: Y2 = |{r ∈ R | r ≥ m}| < ½ $n^{3/4}$ - $\sqrt{n}$
  - E3: |C| > 4 $n^{3/4}$

l_d > n/2 ➔ d is larger than m
➔ Less than (½ $n^{3/4}$ - $\sqrt{n}$) elements in R are smaller than or equal to m

Randomized Median Algorithm

Case: l_d > n/2



Randomized Median Algorithm

KENTECH Korea Institute of Energy Technology AI

- **Lemma**: Pr(E1) $\leq$ (1/4) $\cdot\ n^{-1/4}$

↳ 실패할 확률
(성공: 1-Pr(E1))

Probability at least 1-O(1/ n^c) for some c > 0

- **Proof**
  - Consider random sampling of i-th element and let Xi be a Bernoulli random variable such that
  - Xi = $\begin{cases} 1, & if\ the\ sample\ \leq m \\ 0, & o.w \end{cases}$

  Pr(Xi=1) = $\frac{(n-1)/2+1}{n}$ = ½ + 1/2n

  E[Y1]= ? → $n^{\frac{3}{4}} \cdot (\frac{1}{2}+\frac{1}{2n})$
  Var[Y1] = ? → $n^{\frac{3}{4}} (\frac{1}{2}+\frac{1}{2n})(\frac{1}{2}-\frac{1}{2n})$

  - Define Binomial random variable Y1 = $\sum_{i=1}^{n^{3/4}} X_i$

  ➔ B(n, p) where n= $n^{3/4}$ and p = ½ + 1/2n

  - Event E1 is equivalent to Y1 = $\sum_{i=1}^{n^{3/4}} X_i <$ ½ $n^{3/4}$ - $\sqrt{n}$

  Pr(Y1) = Pr(Y1 < ½ $n^{3/4}$ - $\sqrt{n}$)
  $\leq$ Pr(|Y1 – E[Y1]| > $\sqrt{n}$)
  $\leq \frac{Var[Y1]}{n}$

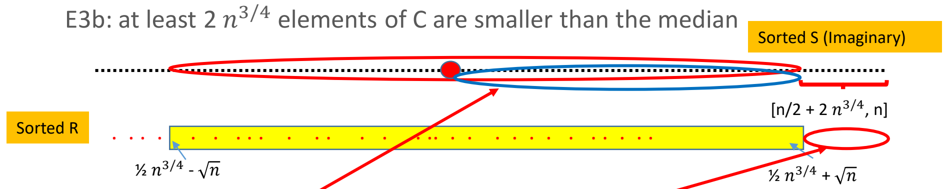
Randomized Median Algorithm

KENTECH Korea Institute of Energy Technology AI

- Now prove Pr(E3=|C| > 4 $n^{3/4}$) $\leq$ (1/2) $\cdot\ n^{-1/4}$
- **Proof**
  - Note that if E3 occur, then at least one of following two events occurs

  E3a: at least 2 $n^{3/4}$ elements of C are greater than the median

  E3b: at least 2 $n^{3/4}$ elements of C are smaller than the median



- Focus on E3a
- There are at least 2 $n^{3/4}$ elements in C that are greater than the median

➔ order of u in S is at least n/2 + 2 $n^{3/4}$

➔ R has at least (1/2) $\cdot\ n^{3/4}$ - $\sqrt{n}$ elements in [n/2 + 2 $n^{3/4}$, n]

Randomized Median Algorithm

KENTECH Korea Institute of Energy Technology AI

- Again define Bernoulli r. v. Xi such that
- Xi = $\begin{cases} 1, & \textit{if the sample is in } [n/2 + 2\, n^{3/4}, n] \\ 0, & o.w \end{cases}$

- Let Y3a = $\sum_{i=1}^{n^{3/4}} X_i$
- Pr(E3a) = Pr(Y3a $\geq$ $(1/2) \cdot n^{3/4}$ - $\sqrt{n}$

  $\leq$ Pr(|Y3a – E[Y3a]| $\geq \sqrt{n}$)

  $\leq \frac{Var[X]}{n} <$ ¼ $n^{-1/4}$

$$\Pr(X_i = 1) = \frac{n - \frac{n}{2} - 2n^{3/4}}{\mathrm{n}} = \frac{1}{2} - 2n^{-1/4}$$
$$\mathrm{E[Y3a]} = \frac{1}{2} n^{3/4} - 2\sqrt{n}$$
$$Var[Y3a] = n^{3/4}\left(\frac{1}{2} - 2n^{-1/4}\right)\left(\frac{1}{2} + 2n^{-1/4}\right)$$
$$= \frac{1}{4} n^{3/4} - 4n^{-1/4} < \frac{1}{4} n^{3/4}$$

➔ Pr(E1)+Pr(E2)+Pr(E3a)+Pr(E3b) $\leq n^{-1/4}$

Example – Parameter Estimation

KENTECH Korea Institute of Energy Technology AI

- We are trying to estimate the parameters of a certain distribution
- For example, judge if a coin is fair or biased
- Suspect that Pr(heads) = p
- Perform n coin flips and let X=n· $\tilde{p}$ be # heads
- Definition: 1-γ Confidence Interval (CI) for a parameter p is an interval [$\tilde{p}$-δ, $\tilde{p}$ +δ] such that

  Minimize both

  $$\Pr(p \in [\tilde{p} - \delta, \tilde{p} + \delta]) \geq 1 - \gamma$$

“전국 19세 이상 성인 남녀 1000명을 대상으로 한 설문조사 결과 X, Y 정당 지지율은 각각 40%, 30% 이다. 이번 조사는 신뢰수준 95%, 오차는 ±3.1%포인트다.”

Trade-off between n, δ, and γ

Example – Parameter Estimation

KENTECH Korea Institute of Energy Technology AI

- X=n· $\tilde{p}$ is a binomial distribution with n and p

- p $\notin[\tilde{p}$-δ, $\tilde{p}$ +δ] ←→ either
  - p < $\tilde{p}$-δ → $n\tilde{p} > n(p+\delta) = E[X]\left(1+\frac{\delta}{p}\right)$
  - p > $\tilde{p}$ +δ → $n\tilde{p} > n(p-\delta) = E[X](1-\frac{\delta}{p})$

- From Chernoff bound,
  - $\Pr(p \notin[\tilde{p}-\delta, \tilde{p}+\delta]) = \Pr\left(X < np\left(1-\frac{\delta}{p}\right)\right) + \Pr\left(X > np\left(1+\frac{\delta}{p}\right)\right)$

$$< e^{-np\left(\frac{\delta}{p}\right)^2/2} + e^{-np\left(\frac{\delta}{p}\right)^2/3}$$
$$< e^{-n\delta^2/2p} + e^{-n\delta^2/3p}$$

Tighter Bounds for Special Cases

KENTECH Korea Institute of Energy Technology AI

- Case 1: Each trial assumes value 1 or -1 with equal probability
- Theorem:

  Let $X_1, X_2, \cdots, X_n$ be independent r.v. such that Pr(Xi=1) = Pr(Xi=-1) = ½

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

  → For any a > 0, Pr(X ≥ a) ≤ $e^{-a^2/2n}$

- Proof:
  - MGF of Xi:
    - $\mathrm{E}[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \sum_{i\geq 0}\frac{t^{2i}}{(2i)!}$

$$\leq \sum_{i\geq 0}\frac{\left(t^2/2\right)^i}{i!} = e^{t^2/2}$$

$$e^t = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^i}{i!} + \cdots$$
$$e^{-t} = 1 - t + \frac{t^2}{2!} + \cdots + (-1)^i\frac{t^i}{i!} + \cdots$$

  - MGF of X: $E[e^{tX}] = \prod_{i=1}^{n}\mathrm{E}[e^{tX_i}] \leq e^{n\,t^2/2}$
  - Pr(X ≥ a) ≤ $e^{\frac{t^2 n}{2} - ta} = e^{-a^2/2n}$

Min. at t=a/n

Tighter Bounds for Special Cases

KENTECH Korea Institute of Energy Technology AI

- Case 2: Bernoulli trials with p = 1/2
- Corollary:
  - Let $Y_1, Y_2, \cdots, Y_n$ be independent r.v. such that Pr(Yi=1) = Pr(Yi=0) = ½. Let Y = $\sum_{i=1}^{n} Y_i$

  ➔ 1. For a > 0, $\Pr(Y \geq \mu + a) \leq e^{-2a^2/n}$

  ➔ 2. For δ > 0, $\Pr(Y \geq (1+\delta)\mu) \leq e^{-\delta^2\mu}$

- Proof:
  - Let $Y_i = (X_i + 1)/2$, Y=$\sum Y_i = \frac{X}{2} + n/2$
  - $\mu = E[Y] = \frac{n}{2}$
  - $\Pr(Y \geq \mu + a) = \Pr(X \geq 2a) \leq e^{-4a^2/2n})$