Concentration of a random variable around its mean

AGT-MIRI QT 2020-2021

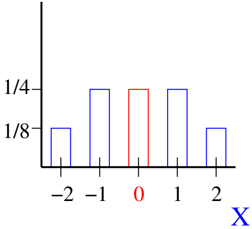
Expectation does not suffice

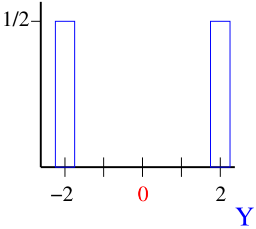
The expected value of a random variable is a nice single number to *tag* the random variable, but it leaves out most of the important properties of the r.v.

Consider r.v X with $X(\Omega) = \{-2, -1, 0, 1, 2\}$ with
$\mathbf{Pr}[X = -2] = \frac{1}{8}, \mathbf{Pr}[X = -1] = \frac{1}{4}, \mathbf{Pr}[X = 0] = \frac{1}{4}$,
$\mathbf{Pr}[X = 1] = \frac{1}{4}, \mathbf{Pr}[X = 2] = \frac{1}{8}$.

and consider r.v. Y with $Y(\Omega') = \{-2, 2\}$ and PMF:
$\mathbf{Pr}[Y = -2] = \frac{1}{2}, \mathbf{Pr}[Y = 2] = \frac{1}{2}$.

Note that $\mathbf{E}[X] = 0 = \mathbf{E}[Y]$, but p_X is totally different from p_Y.





Deviation of a r.v. from its mean

- Consider the deterministic Quicksort algorithm on n-size inputs. Let $T(n)$ be a r.v. counting the number of steps of Quicksort on a specific input with size n
- Its worst case complexity is $O(n^2)$, but its average complexity is $O(n \lg n)$.
- It does not give information about the behavior of the algorithm on a particular input.
- Given an algorithm, for any input x of size $|x| = n$, how close is $T(x)$ to $\mathbf{E}[T(n)]$.

Deviation of a r.v. and Concentration

- For ex.: If $\mathbf{E}[T(n)] = 10$, then 10 is an average running time on "most inputs" to the algorithm. We want to assure, that for most inputs, $T(n)$ is concentrated around 10.
- That is, to make sure that the probability of having instances for which $|\mathbf{E}[T(n)] - T(n)|$ is large, is very small.
- Intuitively, it seems clear from the definition of $\mathbf{E}[]$, if for the above running time, we get an instance e for which $T(e) = 10^9$, and $\mathbf{E}[T(n)] = 10$, the probability of selecting that specific e is going to be quite small, so that its contribution to the average, $10^9 \mathbf{Pr}\left[T(n) = 10^9\right]$, is small.

Markov's inequality

Lemma If $X \geq 0$ is a r.v, for any constant $a > 0$,

$$\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

Proof Given the r.v. $X \geq 0$ define the indicator r.v.

$$Y = \begin{cases} 1 & \text{if } X \geq a \text{ true} \\ 0 & \text{otherwise} \end{cases}$$

Notice if $Y = 1$ then $Y \leq X/a$, and if $Y = 0$ also $Y \leq X/a$, so
$\mathbf{E}[Y] = \mathbf{Pr}[Y = 1] = \mathbf{Pr}[X \geq a]$ and
$\mathbf{E}[Y] = \mathbf{Pr}[Y = 1] \leq \mathbf{E}\left[\frac{X}{a}\right] = \frac{\mathbf{E}[X]}{a}$. $\square$

Alternative expression for Markov: Taking $a = b\mathbf{E}[X]$

Corollary If $X \geq 0$ is a r.v, for any constant $b > 0$,

$$\mathbf{Pr}[X \geq b\mathbf{E}[X]] \leq \frac{1}{b}.$$

Markov could be too weak

Consider the randomized hiring algorithm. We computed that the expected number of pre-selected students is $\mathbf{E}[X] = \lg n$. We also know there are instances for which $X = n$.

We would like to show that the probability of selecting a "bad instance" is very small.

Using Markov, for any constant b, $\mathbf{Pr}[X \geq b \lg n] \leq 1/b$. (for ex. $b = 100$)

The problem with Markov is that it does not bound away the probability of *bad cases* as a function of the input size.

With High Probability

In the randomized algorithms, we aim to obtain results that hold with high probability: the probability that the complexity of the algorithm for any input is "near" the expected value, i.e., it tends to 1 as the size n grows.

An event that occurs with high probability (whp) is one that happens with probability $\geq 1 - \frac{1}{f(n)}$, so that it goes to 1 as $n \to \infty$.

The parameter n is usually the size of the inputs, or the size of the combinatorial structure, ...

Variance

Given a r.v. X, its variance measures the spread of its distribution.

Given X, with $\mu = \mathbf{E}[X]$, the variance of X is:

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

Usually it is more easy to use the expression:
$\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2$
Proof

$$\begin{aligned}\mathbf{Var}[X] &= \mathbf{E}\left[(X-\mu)^2\right] = \mathbf{E}\left[X^2 - 2\mu\mathbf{E}[X] + \mu^2\right] \\ &= \mathbf{E}\left[X^2\right] - 2\mu\underbrace{\mathbf{E}[X]}_{\mu} + \mu^2 = \mathbf{E}\left[X^2\right] - \mu^2 \qquad \square\end{aligned}$$

Further properties of the Variance

- $\mathbf{Var}[X] \geq 0$ as by Jensen's inequality, for any r.v X, $\mathbf{E}\left[X^2\right] \geq \mathbf{E}[X]^2$.
- $\mathbf{Var}[X] = 0$ iff $X =$ constant.
  Proof $(\Leftarrow)$ If $X = c$ then $\mathbf{E}[X] = c \Rightarrow \mathbf{Var}[X] = 0$.
  $(\Rightarrow)$ If $\mathbf{Var}[X] = 0 \Rightarrow \mathbf{E}\left[X^2\right] = \mathbf{E}[X]^2 \Rightarrow \mathbf{E}[X] = c$.
- $\mathbf{Var}[cX] = c^2\mathbf{Var}[X]$.
  Proof
  $\mathbf{Var}[cX] = \mathbf{E}\left[(cX)^2\right] - \mathbf{E}[cX]^2 = c^2\mathbf{E}\left[X^2\right] - (c\mathbf{E}[X])^2$

Computing $\mathbf{Var}[X]$

Given a r.v. X on Ω, such that $X(\Omega) = \{x_1, x_2, \dots, x_n\}$, we first compute
$\mu = \mathbf{E}[X] = \sum_{i=1}^{n} x_i \mathbf{Pr}[X = x_i]$. Then, use one of the following methods:

1. Use $\mathbf{Var}[X] = \mathbf{E}\left[(X-\mu)^2\right]$: For each x_i compute $(x_i - \mu)^2$, and then $\mathbf{Var}[X] = \sum_{i=1}^{n} (x_i - \mu)^2 \mathbf{Pr}[X = x_i]$
2. Use $\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2$: For each x_i compute x_i^2, then $\mathbf{E}\left[X^2\right] = \sum_{i=1}^{n} x_i^2 \mathbf{Pr}[X = x_i]$.

From now on, we use the probability mass function of X, $p_X : \Omega \to [0,1]$, defined as $p_X(\omega) = \mathbf{Pr}[X = \omega]$.

Computing $\mathbf{Var}[X]$: Examples

EX.: Consider r.v. X with $X(\Omega) = \{1, 3, 5\}$ and PMF:
$p_X(1) = \frac{1}{4}, p_X(3) = \frac{1}{4}, P_X(5) = \frac{1}{2}$. Then $\mu = 7/2$.

1. $\mathbf{Var}[X] = \frac{1}{4}(3 - \frac{7}{2})^2 + \frac{1}{4}(5 - \frac{7}{2})^2 + \frac{1}{2}(1 - \frac{7}{2})^2 = \frac{11}{4}$
2. $X^2(\Omega) = \{1, 9, 25\}$, so $\mathbf{E}\left[X^2\right] = \frac{1}{4} + \frac{9}{4} + \frac{25}{2} = 15$
   $\mathbf{Var}[X] = 15 - (\frac{7}{2})^2 = \frac{11}{4}$

Consider r.v. Y with $X(\Omega) = \{-2, 2\}$ and PMF:
$p_Y(-2) = \frac{1}{2}, p_Y(2) = \frac{1}{2}$.
Therefore, the values $(X - \mu)^2$ are $(-2 - 0)^2$ and $(2 - 0)^2$
$\Rightarrow \mathbf{Var}[X] = \frac{1}{2}4 + \frac{1}{2}4 = 4$
Notice in this case $\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] = 4$

You win 100€ with probability $= 1/10$, otherwise you win 0€. Let X be a r.v. counting your earnings. What is $\mathbf{Var}[X]$?
$\mu = 100/10 = 10$. Therefore, $\mathbf{E}\left[X^2\right] = \frac{1}{10}(100^2) = 1000$, and as $\mu^2 = 100$, so $\mathbf{Var}[X] = 900$.

Var [] is not necessarily linear

Let $X_1, \ldots, X_n$ be independent r.v., then

$$\mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}\left[X_i\right].$$

We prove the particular case that if X and Y are independent $\mathbf{Var}\left[X+Y\right] = \mathbf{Var}\left[X\right] + \mathbf{Var}\left[Y\right]$

$$\begin{aligned}
\mathbf{Var}\left[X+Y\right] &= \mathbf{E}\left[(X+Y)^2\right] - (\mathbf{E}\left[X+Y\right])^2 \\
&= \mathbf{E}\left[X^2\right] + \mathbf{E}\left[Y^2\right] + 2\mathbf{E}\left[XY\right] - (\mathbf{E}\left[X\right])^2 - (\mathbf{E}\left[Y\right])^2 - 2\mathbf{E}\left[X\right]\mathbf{E}\left[Y\right] \\
&= \mathbf{E}\left[X^2\right] - (\mathbf{E}\left[X\right])^2 + \mathbf{E}\left[Y^2\right] - (\mathbf{E}\left[Y\right])^2 + 2\underbrace{(\mathbf{E}\left[XY\right] - \mathbf{E}\left[X\right]\mathbf{E}\left[Y\right])}_{\mathbf{E}[XY]=\mathbf{E}[X]\mathbf{E}[Y]}
\end{aligned}$$

Variance of some basic distributions

1. If $X \in B(p, n)$ then $\mathbf{Var}[X] = pqn$, where $q = (1 - p)$.
2. If $X \in P(\lambda)$ then $\mathbf{Var}[X] = \lambda$.
3. If $X \in G(p)$ then $\mathbf{Var}[X] = \frac{q}{p^2}$.

Proof

(1.-) Let $X = \sum_{i=1}^{n} X_i$, where X_i is an indicator r.v s.t. $X_i = 1$ with probability p

Then, $\mathbf{Var}[X_i] = \mathbf{E}\left[X_i^2\right] - \mathbf{E}[X]^2 = (p \cdot 1^2 + q \cdot 0 - p^2 = p(1 - p)$,
as all X_i are independent, $\mathbf{Var}[X] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = np(1 - p)$.

Proof of 2

$$\begin{aligned}\mathbf{Var}\left[X\right] &= \mathbf{E}\left[X^2\right] - (\mathbf{E}\left[X\right])^2 = \mathbf{E}\left[(X)(X-1)+X\right] - (\mathbf{E}\left[X\right])^2 \\ &= \mathbf{E}\left[(X)(X-1)\right] + \mathbf{E}\left[X\right] - (\mathbf{E}\left[X\right])^2 = \mathbf{E}\left[(X)(X-1)\right] + \lambda - \lambda^2.\end{aligned}$$

$$\begin{aligned}\mathbf{E}\left[(X)(X-1)\right] &= \sum_{x=0}^{\infty}(x)(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \\ &= \sum_{x=2}^{\infty}(x)(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \quad \text{terms } x=0 \text{ and } x=1 \text{ are } 0 \\ \sum_{x=2}^{\infty}\frac{e^{-\lambda}\lambda^x}{(x-2)!} &= \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} \\ &= \lambda^2 e^{-\lambda}\left(\frac{\lambda^0}{0!}+\frac{\lambda^1}{1!}+\frac{\lambda^2}{2!}+\dots\right) \\ &= \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2.\end{aligned}$$

Proof of 3

If $X \in G(p)$ want to compute $\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \mathbf{E}[X^2] - \frac{1}{p^2}$.
Need to compute $\mathbf{E}[X^2]$.

$$\begin{aligned}\mathbf{E}[X^2] &= \sum_{k=1}^{\infty} k^2 \mathbf{Pr}[X = k] \\ &= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = p \underbrace{\sum_{k=1}^{\infty} k^2 (1-p)^{k-1}}_{*}\end{aligned}$$

Recall Taylor: $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$. Differentiating $\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1}$.
Multiplying by x and differentiating $\frac{1-x}{(1-x)^3} = \sum_{k=1}^{\infty} k^2 x^{k-1}$
Making $x = 1-p$ then $\frac{2-p}{p^3} = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1}$.
By $(*)$ $\mathbf{E}[X^2] = \frac{2-p}{p^2}$
Therefore: $\mathbf{Var}[X] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$

A more natural measure of spread: Standard Deviation

Why we did not define $\mathbf{Var}[X] = \mathbf{E}[|X - \mu|]$?
To be sure we are averaging only non-negative values.

But as we defined the variance, we are using squared units!

Recall the example X a r.v. counting the wins, when you win 100€ with probability $= 1/10$, otherwise you win 0€. We got $\mathbf{Var}[X] = 900€^2$.
To convert the numbers back to re-scale, we take the square root.

The Standard Deviation of a r.v. X is defined as

$$\sigma[X] = \sqrt{\mathbf{Var}[X]}.$$

In the previous example, to convert the spread from €^{2} to €, $\sigma[X] = \sqrt{900} = 30$ €.

Chebyshev's Inequality

Pafnuty Chebyshev (XIXc)

If you can compute the $\mathbf{Var}[]$ then you can compute σ and get better bounds for concentration of any r.v. (positive or negative).

Theorem Let X be a r.v. with expectation μ and standard deviation $\sigma > 0$, then for any $a > 0$

$$\mathbf{Pr}[|X-\mu| \geq a\sigma] \leq \frac{1}{a^2}.$$

Note that $|X-\mu| \geq a\sigma \Leftrightarrow (X \geq a\sigma + \mu) \cup (X \geq \mu - a\sigma)$.

Proof As the r.v. $|X-\mu| \geq 0$, we can apply Markov to it:

$$\begin{aligned}\mathbf{Pr}[|X-\mu| \geq a\sigma] &= \mathbf{Pr}\left[(X-\mu)^2 \geq a^2\sigma^2\right] \qquad \text{(by Markov)}\\ &\leq \frac{\mathbf{E}\left[(X-\mu)^2\right]}{a^2\sigma^2} = \frac{\mathbf{Var}[X]}{a^2\mathbf{Var}[X]} = \frac{1}{a^2} \qquad \square\end{aligned}$$

More on Chebyshev's Inequality

We had: $\mathbf{Pr}\left[|X-\mu| \geq a\sigma\right] \leq \frac{1}{a^2}$.

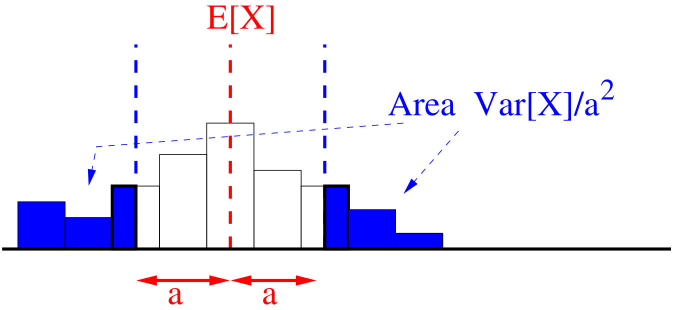
Alternative equivalent statement:

$$\forall b > 0, \mathbf{Pr}\left[|X-\mu| \geq b\right] \leq \frac{\mathbf{Var}\left[X\right]}{b^2}.$$

Proof As before: $\mathbf{Pr}\left[(X-\mu)^2 \geq b^2\right] \leq \frac{\mathbf{E}\left[(X-\mu)^2\right]}{b^2}$.

Chebyshev's Inequality: Picture

$$\mathbf{Pr}\left[|X-\mu| \geq a\right] \leq \frac{\mathbf{Var}\left[X\right]}{a^2}.$$



An easy application

Let flip n-times a fair coin, give an upper bound on the probability of having at least $\frac{3n}{4}$ heads.

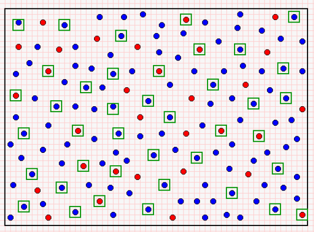
Let $X \in B(n, 1/2)$, then, $\mu = n/2, \mathbf{Var}[X] = n/4$.

We want to bound $\mathbf{Pr}\left[X \geq \frac{3n}{4}\right]$.

- Markov: $\mathbf{Pr}\left[X \geq \frac{3n}{4}\right] \leq \frac{\mu}{3n/4} = 2/3$.
- Chebyshev's: We need the value of a s.t. $\mathbf{Pr}\left[X \geq \frac{3n}{4}\right] \leq \mathbf{Pr}\left[|X - \frac{n}{2}| \geq a\right] \Rightarrow a = \frac{3n}{4} - \frac{n}{2} = \frac{n}{4}$.

  $\mathbf{Pr}\left[X \geq \frac{3n}{4}\right] \leq \mathbf{Pr}\left[|X - \frac{n}{2}| \geq \frac{n}{4}\right] \leq \frac{\mathbf{Var}[X]}{(n/4)^2} = \frac{4}{n}$.

Sampling



- Given a large population Σ, $|\Sigma| = n$, we wish to estimate the proportion p of elements in Σ, with a given property.
- Sampling: Take a random sample S with size $m << n$ and observe p^- in S.
- Sometimes, if n is large, the obvious estimator $m \times p^-$ is sufficiently good, i.e. it is sharply concentrated.
- Many times getting the random sample S is non-trivial.

Finding the median of n elements

From MU 3.4

- Recall that, given a set S with n distinct elements, the median of S is the $\lceil n/2 \rceil$ larger element in S.
- We can use Quickselect to find the median with expected time $O(n)$. Even there is a linear time deterministic algorithm, which in practice for large instance works worst than Quick-select.
- We present another randomized algorithm to find the median m in S, which is based in sampling.
- The purpose of this algorithm is to introduce the technique of filtering large data by sampling small amount of the data.

Finding the median of n elements: A Filtering Data algorithm

INPUT: An unordered set $S = \{x_1, x_2, \dots x_n\}$, with $n = 2k + 1$ elements.
OUTPUT: The median, which is the $k + 1$ largest element in S.
For any element y define the $\text{rank}(y) = |\{x \in S | x \leq y\}|$.

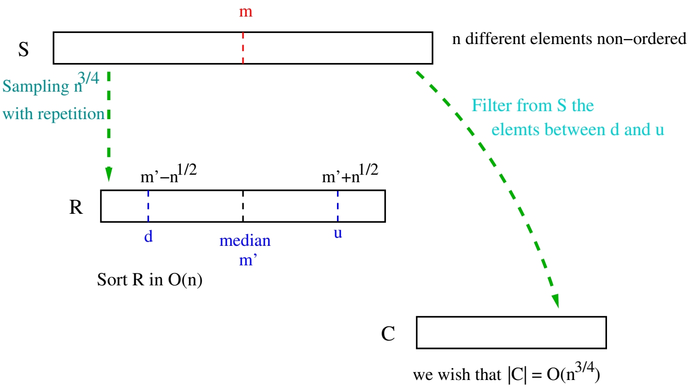
The idea of the filtering Algorithm is to sample with replacement a "small" subset of C elements from S, so we can order C in $O(n)$ time (linear with respect to the size of S).
Then the algorithm find the median of the elements in C and either return it as the median in S or return failure

We will prove that whp the algorithm finds the median m of S, in linear time.

Outline of the algorithm

1. Let $\tilde{S}$ be the ordered set S (we do not know $\tilde{S}$). Let m be its median.
2. Find elements $d, u \in S$ s.t. $d < m < u$ and distance between d and u in $\tilde{S}$ is $< n/\lg n$.
3. To find d and u sample with replacement S to get a multi-set R, with $|R| = O(\lceil n^{3/4} \rceil)$. Notice $\lceil n^{3/4} \rceil < n/\lg n$. Find $u, d \in R$ s.t. m will be close to median in R).
4. Filter-out the elements $x \in S$, which are $< d$ or $> u$ to form a set $C = \{x \in S | d \leq x \leq u\}$.
5. Sort elements in C in $O(n)$, and find its median. This will be the algorithm's output
6. Prove that w.h.p. the algorithm succeeds.

Outline of the algorithm



Things that can be wrong:
C too large,
$m \notin C$,
$m \in C$ but no the median in C.

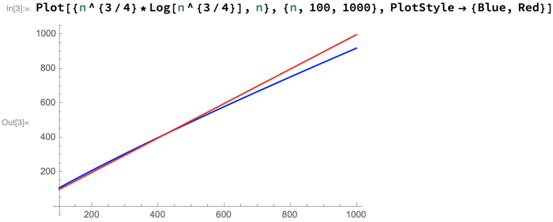
Randomized Median algorithm

1. Sample $\lceil n^{3/4} \rceil$ elements from S, u.a.r., independently, and with replacement.
2. Sort R in $O(n)$
3. Set $d = \lfloor (\frac{n^{3/4}}{2} - \sqrt{n}) \rfloor$-smallest element in R
4. Set $d = \lfloor (\frac{n^{3/4}}{2} + \sqrt{n}) \rfloor$-greatest element in R
5. Compute $C = \{x \in S | d \leq x \leq u\}$, $l_d = |\{x \in S | x < d\}|$ and $l_u = |\{x \in S | x > u\}|$ (cost $= \Theta(n)$).
6. If $l_d > \frac{n}{2}$ or $l_u < \frac{n}{2}$ OUTPUT FAIL ($m \notin C$)
7. If $|C| \leq 4n^{3/4}$, sort C, otherwise OUTPUT FAIL.
8. OUTPUT the $(\lfloor \frac{n}{2} \rfloor - l_d + 1)$-smallest element in sorted C, that should be m.

Complexity and correctness of the Randomized Median algorithm

Theorem: The Randomized Median algorithm terminates in $O(n)$ steps. If the algorithm does not output FAIL, then it outputs the median m of S.

Proof: As asymptotically $n^{3/4}\lg(n^{3/4}) \leq n$, using Mergesort on R takes $O(\frac{n}{\lg n}\lg(\frac{n}{\lg n})) = O(n)$.
The only incorrect answer is that it outputs an item, but $m \notin C$, but if so, it would fail in step 6, as either $l_d > n/2$ or $l_u < n/2$. $\square$



?

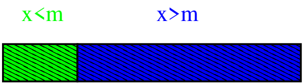
Figure: $n^{3/4}\lg(n^{3/4})$ versus n

Bounding the probability of output FAIL

Theorem: The Randomized Median algorithm finds m with probability $\geq 1 - \frac{1}{n^{1/4}}$, i.e., whp.
Proof (Highlights): Consider the following 3 events:
E_1: $d > m$,
E_2: $u < m$,
E_3: $|C| > 4n^{3/4}$.
Then, the algorithm outputs FAIL iff one of the three events occurs, i.e.
$\mathbf{Pr}[\text{FAILS}] = \mathbf{Pr}[E_1 \cup E_2 \cup E_3] \leq \mathbf{Pr}[E_1] + \mathbf{Pr}[E_2] + \mathbf{Pr}[E_3]$

Bounding $\mathbf{Pr}[E_1]$

Consider R ordered, where R is obtained by sampling $n^{3/4}$ elements from S



Recall: $d = \lfloor(\frac{n^{3/4}}{2} - \sqrt{n})\rfloor$-th element

- $d > m$, when the green block has size $< \lfloor n^{3/4}/2 - \sqrt{n} \rfloor$.
- Let $Y = |\{x \in R \,|\, x \leq m\}|$, then $\mathbf{Pr}[E_1] = \mathbf{Pr}\left[Y < n^{3/4}/2 - \sqrt{n}\right]$.
- For $1 \leq j \leq n^{3/4}$, define $Y_j = 1$ iff the value in the j-th. position in R is $\leq m$.
- Then $Y = \sum_{j=1}^{n^{3/4}} Y_j$, moreover as the sampling is with replacement, then each Y_j is independent.

As $m =$ median of S ($|S| = n$), then we have $\frac{(n-1)}{2} + 1$ elements in S that are $\leq m$.

Bounding $\mathbf{Pr}[E_1]$

- $\mathbf{Pr}[Y_j = 1] = \frac{(n-2)/2+1}{n} = \frac{1}{2} + \frac{1}{2n}$, as there are $(n-1)/2+1$ elements $\leq m$.
- so $Y \in B(n^{3/4}, \frac{1}{2} + \frac{1}{2n})$.
- Then $\mathbf{E}[Y_i] \geq 1/2 \Rightarrow \mathbf{E}[Y] \geq \frac{n^{3/4}}{2}$,
- Y is $B(n^{3/4}, 1/2 + 1/2n$, so $\mathbf{Var}[Y] = n^{3/4}(\frac{1}{2} + \frac{1}{2n})(\frac{1}{2} - \frac{1}{2n}) \leq \frac{n^{3/4}}{4}$.

Using Chebyshev:

$$\mathbf{Pr}[E_1] = \mathbf{Pr}\left[Y < \frac{n^{3/4}}{4} - \sqrt{n}\right]$$
$$\leq \mathbf{Pr}\left[|Y - \mathbf{E}[Y]| \geq \sqrt{n}\right] \leq \frac{\mathbf{Var}[Y]}{(\sqrt{n})^2} = \frac{1}{4n^{1/4}} \quad \square$$

Bounding $\mathbf{Pr}[E_2]$

In the same way as for E_1, it holds $\mathbf{Pr}[E_2] \leq \frac{1}{4n^{1/4}}$

Bounding $\mathbf{Pr}[E_3]$

E_3: $|C| > 4n^{3/4}$.

C is obtained directly from S by filtering, using the values d and u obtained in R.

For C to have $> 4n^{3/4}$ keys either of the following events must happen:

1. A: At least $> 2n^{3/4}$ items in C are $> m$.
2. B: At least $> 2n^{3/4}$ items in C are $< m$.

Then,

$$\mathbf{Pr}[E_3] \leq \mathbf{Pr}[A \cup B] \leq \mathbf{Pr}[A] + \mathbf{Pr}[B].$$

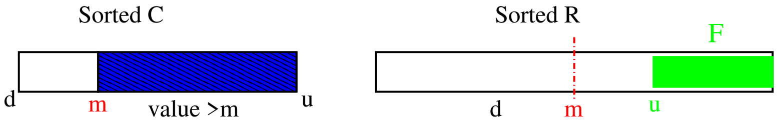
Bounding $\mathbf{Pr}[A]$

Event A happens when there are at least $2n^{3/4}$ element in C which are $> m$

If so, the rank(u) in $\tilde{S}$ is $\geq n/2 + 2n^{3/4}$.

Let $F = \{x \in R \mid x > u\}$, $|F| \geq n^{3/4}/2 - \sqrt{n}$

Any element in F has rank $\geq n/2 + 2n^{3/4}$



We will prove that $\mathbf{Pr}\left[\bar{A}\right] = 1 - O(1/n) \to 1$.

Bounding $\mathbf{Pr}[A]$

- Let $X = \#$ selected items in R that are in F (have rank $\geq n/2 + 2n^{3/4}$)
- Then $\mathbf{Pr}[A] \leq \mathbf{Pr}\left[X \geq \lfloor n^{3/2}/2 - \sqrt{n} \rfloor\right]$.
- For $1 \leq j \leq n^{3/4}$, define $X_j = 1$ iff the j-th item in R is in F.
- Note $X = \sum_{j=1}^{n^{3/4}} X_j$ and $\mathbf{Pr}[X_j = 1] = \frac{1}{2} - \frac{2}{n^{1/4}} + \frac{1}{n}$.
- So $\mathbf{E}[X] = \frac{n^{3/4}}{2} - 2n^{1/2} + n^{1/4}$ and $\mathbf{Var}[X] \leq n^{3/4}/4$

$$\begin{aligned}
\mathbf{Pr}[A] &\leq \mathbf{Pr}\left[X \geq \lfloor \frac{n^{3/2}}{2} - n^{1/2} \rfloor\right] \leq \mathbf{Pr}\left[X \geq \frac{n^{3/4}}{2} - 2n^{1/2} + n^{1/4}\right] \\
&\leq \mathbf{Pr}\left[X \geq \mathbf{E}[X] + n^{1/2} - 1 - n^{1/4}\right] \\
&\leq \mathbf{Pr}\left[|X - \mathbf{E}[X]| \geq n^{1/2} - 1 - n^{1/4}\right] = O(\frac{1}{n^{1/4}}). \quad \square
\end{aligned}$$

Bounding $\mathbf{Pr}[B]$ and finishing the proof

In the same way we can compute $\mathbf{Pr}[B] = O(\frac{1}{n^{1/4}})$

To end the whole proof, we also proved that
$\mathbf{Pr}[E_3] \leq \mathbf{Pr}[A] + \mathbf{Pr}[B] = O(\frac{1}{n^{1/4}})$

$\Rightarrow \mathbf{Pr}[\text{algorithm fails}] = \mathbf{Pr}[E_1 \cup E_2 \cup E_3] \leq^{\text{UB}} O(\frac{1}{n^{1/4}})$.

Therefore,
$\mathbf{Pr}[\text{algorithm succeeds}] = 1 - \mathbf{Pr}[\text{algorithm fails}] \geq 1 - \frac{1}{n^{1/4}}$
i.e. w.h.p. the Randomized Median algorithm finds the correct m
$\square$