DS-563/CS-543 2025-01-22

Reminders / announcements:

- course webpage: https://onak.pl/ds563 (or .../cs543)
- make sure you are on Piazza / Gradescope
- non-mandatory HW0: upload an audio file with the pronounciation of your name (or the name you want to use)
- electronic device policy: no, unless solely for taking notes

Today:

- review of some probabilistic inequalities and facts
- will apply tomorrow to our first big data problem (frequency estimation)

Union Bound

$\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$ - probabilistic events

$$\Pr(\text{at least one of events } \mathcal{E}_i \text{ ocurrs}) \leq \sum_{i=1}^{k} \underbrace{\Pr(\mathcal{E}_i)}_{\text{probability of } \mathcal{E}_i}$$

Typical application in this class:
$\mathcal{E}_1, \dots, \mathcal{E}_k$ are bad events for our randomized algorithm and we want to avoid them. If we make the probability of each of them small, then the union bound can be used to ensure that none of them occur most of the time.

Note: $\mathcal{E}_1, \dots, \mathcal{E}_k$ can be arbitrarily related so we do not have to analyze their relationship. In particular, they don't have to be independent. So it is easy to use this bound.

Markov's inequality

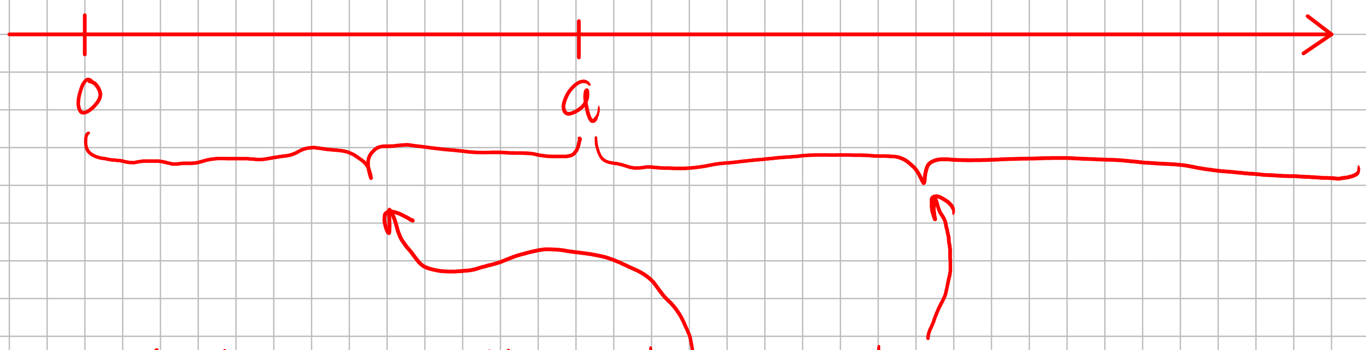
X – non-negative random variable (s.t. $\mathbb{E}[X] < \infty$)

For any $a > 0$,

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

Intuition: X cannot be too big too often or this would contradict the expectation bound.

Proof:



X always either here or here

$$\mathbb{E}[X] \geq \underbrace{\Pr[X < a] \cdot 0}_{=0} + \Pr[X \geq a] \cdot a$$

$$\mathbb{E}[X] \geq \Pr[X \geq a] \cdot a$$

$$\frac{\mathbb{E}[X]}{a} \geq \Pr[X \geq a] \qquad \blacksquare$$

Question: Is the non-negativity of X important?

Note: If one of $\mathbb{E}[X]$ or a were negative, we would get a negative upper bound on probability of an event, which clearly would be false, So suppose that both a and $\mathbb{E}[X]$ are still positive.

What about our proof? It breaks because we assume that if $X < a$, then it is at least 0. But maybe a different proof would work? In this case, it's best to show a counterexample.

Counterexample:

large positive number

$$X = \begin{cases} 2 + \Delta & \text{w.p. } 1/2 \\ -\Delta & \text{w.p. } 1/2 \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{2}(2+\Delta) + \frac{1}{2}(-\Delta) = 1$$

Set $a = 1000$ and $\Delta = 2000$. Markov's would

give $\Pr[X \geq 1000] \leq \frac{1}{1000}$ but $\Pr[X \geq 1000] =$
$= \Pr[X = 2002] = \frac{1}{2}$.

Comment/intuition: In this case, negative values hide how large X can become.
Δ could be arbitrarily large, and X would be arbitrarily large w.p. $1/2$.
So there is no "weaker" version of Markov's inequality in this case. No information about how big X can get survives in $\mathbb{E}[X]$.

Additional Challenge (useful tomorrow):
Suppose you play some game k times. Each instance is fully independent and each time you win or lose.
You know you win each time with probability at least $1/2$.
What value of k guarantees that you win at least once with

probability $\geq 1-\delta$ (where $\delta \in (0, 1/2)$).
(In other words, how many times do you have to play to win at least once w.p. $\geq 1-\delta$.)

(Almost the same challenge: You are tossing an unbiased coin k times. What should k be so you see tails at least once w.p. $\geq 1-\delta$?)

Solution:

Due to the independence of games,

$$\Pr[\text{all games lost}] = \prod_{i=1}^{k} \Pr[\text{game } i \text{ lost}]$$
$$\leq \frac{1}{2^k}.$$

$$\Pr[\text{at least one game won}] =$$
$$= 1 - \Pr[\text{all games lost}] \geq 1 - \frac{1}{2^k}.$$

So it suffices to set k so that

$$1 - \frac{1}{2^k} \geq 1-\delta$$

$$\Updownarrow$$

$$\delta \geq \frac{1}{2^k}$$
$$\Updownarrow$$
$$2^k \geq \frac{1}{\delta}$$
$$\Updownarrow$$
$$k \geq \log_2(1/\delta)$$

This can be achieved by setting

$$k = \lceil \log_2(1/\delta) \rceil$$

always a valid positive integer because $\delta < 1/2$

the ceiling operator

|||

rounding up to an integer

(see the handout)