October 26

Monday, October 26, 2020 3:58 PM

Comp Sci 212

1. Markov's Inequality
2. Chebyshev's Inequality

Tails bounds

$\Pr[R \geq x] \leq ?$

R → random variable, x → number

- If R represents running time of a randomized algorithm, want $\Pr[R \geq x]$ to be small.

Markov's Inequality - what tail bounds can we get from just its expected value.

Ex. In a room w/ 100 people, avg. age 40 years old.

S = set of people, uniform dist.

$R: S \to \mathbb{R}$, $R(s)$ = age of person s.

$$\mathbb{E}[R] = \sum_{s \in S} \frac{R(s)}{|S|} = 40.$$

How large can $\Pr[R \geq 100]$ be?

→ proportion of people at least 100.

- Not everyone can be older than 100.
- Even $\Pr[R \geq 100] = \frac{1}{2}$ is not possible
  → 50 people 100 years, 50 0 years
- $\Pr[R \geq 100] = \frac{1}{4}$ is possible: 25 people 100, 75 people 20, avg 40

- largest possible value $\Pr[R \geq 100] = \frac{4}{10} = \frac{\mathbb{E}[R]}{100}$
  - 40 people 100 years old
  - 60 people 0

Thm. Markov's Inequality

Thm. If R is a non-negative r.v., $x > 0$,

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R]}{x}.$$

Proof. Enough to prove $\mathbb{E}[R] \geq x \Pr[R \geq x]$.

$$\underbrace{\mathbb{E}[R]}_{\sum_{s \in S} R(s)\Pr[s]} \geq \underbrace{x \Pr[R \geq x]}_{\sum_{s \in S,\, R(s) \geq x} x \cdot \Pr[s]}$$

$$\sum_{s \in S} R(s)\Pr[s] = \sum_{\substack{s \in S \\ R(s) \geq x}} R(s)\Pr[s] + \sum_{\substack{s \in S \\ R(s) < x}} R(s)\Pr[s]$$

← always bigger than 0 (assumption; prob is always ≥ 0.)

$$\geq \sum_{\substack{s \in S \\ R(s) \geq x}} R(s)\Pr[s] \geq \sum_{\substack{s \in S \\ R(s) \geq x}} x\Pr[s] \quad \square$$

Birthday Paradox - D set of days, D^n sample space

$R: D^n \to \mathbb{R}$ uniform dist.

$R(s)$ = # of pairs of people w/ same birthday.

$E_{i,j}$ = event that person i & person j have the same birthday.

$$R(s) = \sum_{\substack{i,j \in [n] \\ i<j}} \mathbb{1}_{E_{i,j}}(s) \quad \text{(over all pairs)}$$

$$\mathbb{E}[R] = \sum_{\substack{i,j \in [n] \\ i<j}} \Pr[E_{i,j}] = \underbrace{\frac{n(n-1)}{2}}_{\text{\# of pairs}} \cdot \underbrace{\frac{1}{101}}_{\text{prob.}}$$

$n \approx \sqrt{2101}$, $\mathbb{E}[R] \approx 1$

$$\Pr[R \geq 10] \leq \frac{\mathbb{E}[R]}{10} \approx \frac{1}{10}.$$

R → non-negative.

$S = \{H, T\}^{100}$ Uniform dist

Sticky coin dist - all heads / all tails w.p. $\frac{1}{2}$

$R: S \to \mathbb{R}$ $R(s) = \#$ of heads

$\mathbb{E}[R] = 50$

By Markov's inequality

$$\Pr[R \geq 75] \leq \frac{\mathbb{E}[R]}{75} = \frac{50}{75} = \frac{2}{3}$$

Sticky coin $\Pr[R \geq 75] = \frac{1}{2}$

Uniform dist $\Pr[R \geq 75] =$ really small

- Assume everyone in room is at least 20 years old $\mathbb{E}[R] = 40$ $|S| = 100$

How large can $\Pr[R \geq 100]$ be?

- 40 people over 100 is impossible

largest 25 people $\Pr[R \geq 100] \leq \frac{25}{100}$

25 people 100 years

75 people 20 years

$$\frac{40-20}{100-20} = \frac{1}{4}$$

Corollary. If $R(s) \geq b$ for all s, $x > b$

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R] - b}{x - b}$$

Proof. Let $Y: S \to \mathbb{R}$, $Y(s) = R(s) - b$.

$Y(s)$ always non-negative

$\mathbb{E}[Y] = \mathbb{E}[R] - b$

$$\Pr[R \geq x] = \Pr[Y \geq x - b] \leq \frac{\mathbb{E}[Y]}{x - b} = \frac{\mathbb{E}[R] - b}{x - b}$$

by Markov's inequality