May 7

Friday, May 7, 2021 2:22 PM

# Comp Sci 212

1. Tail Bounds
2. Markov's Inequality
3. Chebyshev's Inequality

---

Tail bound

$\Pr[\underset{\text{random variable}}{R} \geq \underset{\text{number}}{x}] \leq C$ → upper-bound on probability

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

Markov's Inequality - what tail bounds can we get just by knowing the expected value?

Ex. In a room with 100 people, average age is 40.

$S$ = set of people

$\Pr[w] = \frac{1}{|S|}$ for each person $w \in S$

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

$\mathbb{E}[R] = \sum_{w \in S} R(w)\Pr[w] = 40$

What is the largest $\Pr[R \geq 100]$ can be? (proportion of people at least 100)

- Not everyone can be older than 100
- Even $\Pr[R \geq 100] = \frac{1}{2}$, not possible
  ↳ half people 100, other half 0 → $\mathbb{E}[R] = 50$
- $\Pr[R \geq 100] = \frac{1}{4}$, is possible
  25 people 100 years
  75 people 20 years → average 40
- largest is $\Pr[R \geq 100] = \frac{4}{10}$ $\quad x = 100$, $\mathbb{E}[R] = 40$, $\frac{40}{100}$
  40 people 100, 60 people 0 years old

Thm. Markov's Inequality

If $R$ is a non-negative random variable, $x \geq 0$, then

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

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

$\sum_{w: R(w) \geq x} \Pr[w] \cdot x \qquad \sum_{w \in S} R(w)\Pr[w]$

$$\sum_{w \in S} R(w)\Pr[w] = \sum_{w: R(w) \geq x} R(w)\Pr[w] + \underbrace{\sum_{w: R(w) < x} R(w)\Pr[w]}_{\text{at least } 0}$$

$$\geq \sum_{w: R(w) \geq x} R(w)\Pr[w]$$

$$\geq \sum_{w: R(w) \geq x} x \cdot \Pr[w] \quad (R(w) \geq x)$$

□

Birthday Paradox - $D$ set of days

$S = D^n$, uniform distribution

$R: S \to \mathbb{R}$, $R(w)$ = # of pairs w/ same birthday

$E_{i,j}$ → event where person $i$ & person $j$ have the same birthday

$R = \sum_{\{i,j\} \subseteq [n]} \mathbb{1}_{E_{i,j}}$, $\mathbb{E}[R] = \sum_{\{i,j\} \subseteq [n]} \Pr[E_{i,j}]$

indicator r.v.

$\mathbb{1}_{E_{i,j}}: S \to \mathbb{R}$

$\mathbb{1}_{E_{i,j}}(w) = \begin{cases} 1 & \text{if } w \in E_{i,j} \\ 0 & \text{o.w.} \end{cases}$

$= \sum_{\{i,j\} \subseteq [n]} \frac{1}{|D|}$ (skipped step)

$= \frac{n(n-1)}{2|D|}$, if $n \approx \sqrt{2|D|}$, $n \approx 27$

$\approx 1$

Markov's Inequality

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

$R: S \to \mathbb{R}$, $S = \{H, T\}^{100}$, $R(s)$ = # of heads

uniform dist, sticky dist

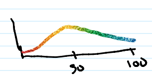
$\mathbb{E}[R] = 50$ for both

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

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

↳ $\Pr[\text{all heads}] = \Pr[\text{all tails}] = \frac{1}{2}$

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




Ex. Average age 40

Assume everyone in the room is at least 20 years old

- 40 people at least 100 years old is impossible
  40 people 100 years
  60 people 20 years old $\quad$ avg 52
  $b = 20$, $x = 100$, $\mathbb{E}[R] = 40$, $\frac{40-20}{100-20} = \frac{20}{80} = \frac{1}{4}$

largest is $\Pr[R \geq 100] = \frac{25}{100}$ → r.v. representing age

  25 people 100 years
  75 people 20 years

Corollary. If $R(w) \geq b$ for every $w \in S$, $x > b$

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

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

$Y(w)$ is non-negative, $\mathbb{E}[Y] = \mathbb{E}[R - b] = \mathbb{E}[R] - b$ (linearity of expectation)

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

$(R(w) \geq x$ iff $Y(w) \geq x - b)$ $\quad$ (Markov's inequality)

□