Week 12

↳ Research talk tomorrow 4:15 in Kravis 62

↳ Office Hours Monday + Thursday 12:30-2:00

↳ Midterm on 11/25, practice exams next week :)

↳ Project Proposal due 12/2

Reinforcement Learning

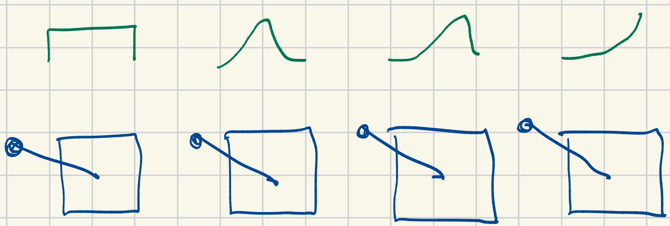
State, action, reward, ...

Key: Exploration vs. Exploitation

This Week: theoretical understanding

Simple model: multi-armed bandits

Question:
What can we say about draws from these distributions?



Hidden reward distribution

"slot machines"

Goal: Pull arms to maximize reward

↳ clinical trials

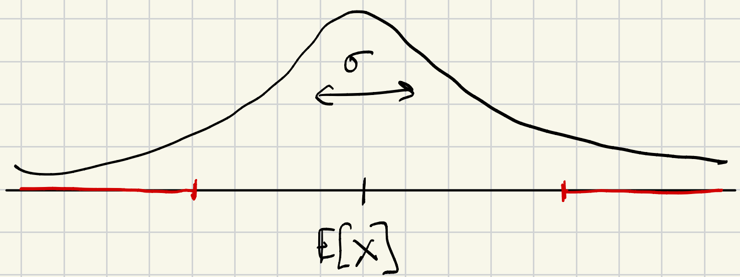
↳ recommendation systems

Random Variables

X r.v. $\qquad Pr(X=x)$ = prob X takes value x

$$\mathbb{E}[X] = \sum_x x \, Pr(X=x)$$

$$Var(X) = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2$$

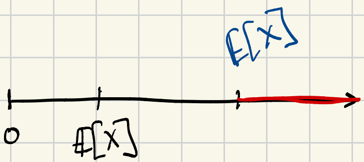


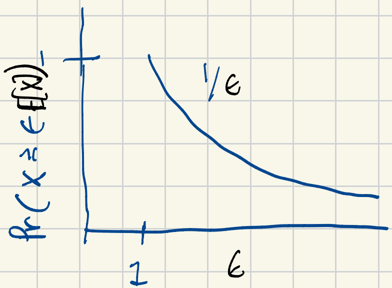
Goal: Bound probability of <u>extreme</u> values

Markov's Inequality

Consider non-negative X

For $\epsilon > 0$, $\quad Pr(X \geq \epsilon\, \mathbb{E}[X]) \leq \frac{1}{\epsilon}$





Proof:

$$\mathbb{E}[X] = \sum_{x} Pr(X=x)\, x$$

$$= \sum_{x:\, x \geq \epsilon} Pr(X=x)\, x + \sum_{x:\, x < \epsilon} Pr(X=x)\, x$$

$$\geq \sum_{x:\, x \geq \epsilon} Pr(X=x)\, \epsilon + \sum_{x:\, x < \epsilon} Pr(X=x) \cdot 0$$

$$= \epsilon \sum_{x:\, x \geq \epsilon} Pr(X=x)$$

$$= \epsilon\, Pr(X \geq \epsilon)$$

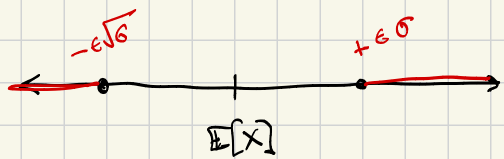
Q: Fix ϵ. Build a rv where $Pr(X \geq \epsilon) = \frac{\mathbb{E}[X]}{\epsilon}$

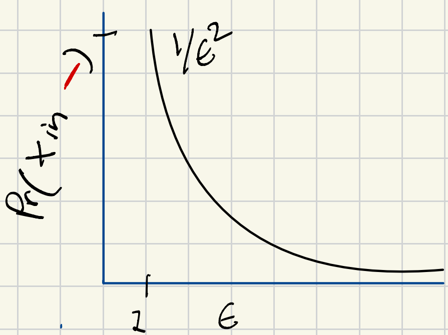
Chebyshev's Inequality

Idea: Use extra information for tighter bound

$\sigma^2 = Var(X)$. For $\epsilon > 0$,

$$Pr(|X - \mathbb{E}[X]| \geq \epsilon\sigma) \leq \frac{1}{\epsilon^2}$$





Proof: $Z = (X - \mathbb{E}[X])^2$

By Markov's,

$$Pr(Z \geq \epsilon\mathbb{E}[Z]) \leq \frac{1}{\epsilon}$$

$\Longleftrightarrow$

$$Pr\left((X - \mathbb{E}[X])^2 \geq \epsilon \overbrace{\mathbb{E}[(X - \mathbb{E}[X])^2]}^{\sigma^2}\right) \leq \frac{1}{\epsilon}$$

$\Longleftrightarrow$

$$Pr(|X - \mathbb{E}[X]| \geq \sqrt{\epsilon}\,\sigma) \leq \frac{1}{\epsilon}$$

$\Longleftrightarrow$

$$Pr(|X - \mathbb{E}[X]| \geq \epsilon\sigma) \leq \frac{1}{\epsilon^2}$$

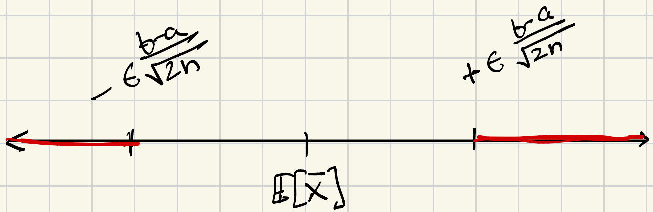
Hoeffding's Inequality

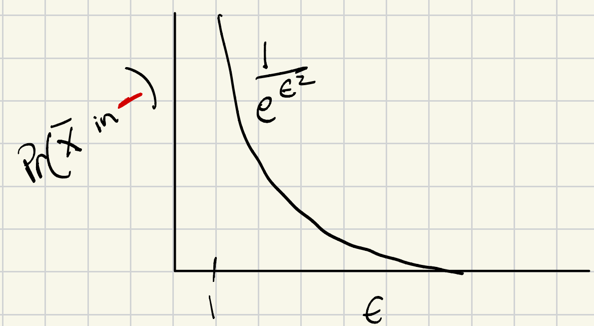
Idea: Formalize central limit theorem

Consider $X_1, \ldots, X_n$ where $a \leq X_i \leq b$

Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. For $\epsilon > 0$,

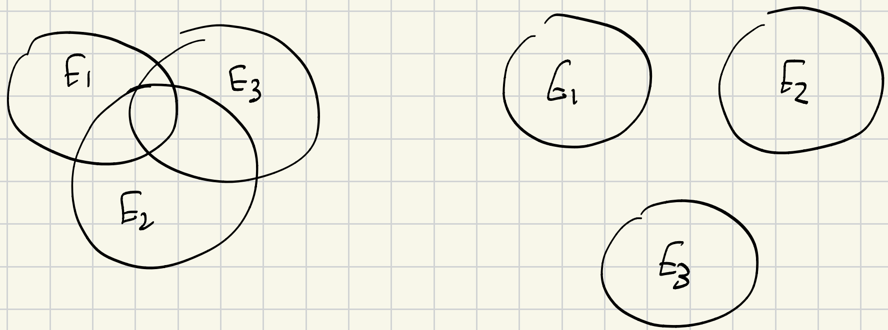
$$\Pr\left(|\bar{X} - \mathbb{E}[\bar{X}]| \geq \epsilon \right) \leq 2 \exp\left(- \frac{2n\epsilon^2}{(b-a)^2} \right)$$





Union Bound

$$\Pr(E_1 \cup E_2 \cup \dots \cup E_m) \leq \Pr(E_1) + \Pr(E_2) + \dots + \Pr(E_m)$$



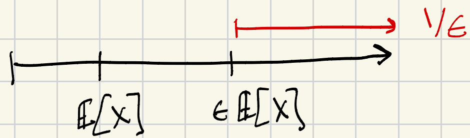
Logistics

↳ Midterm

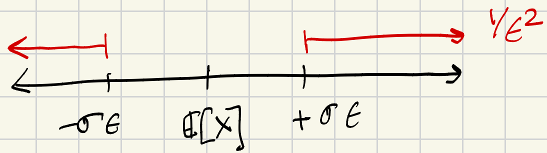
↳ Project

Concentration Inequalities

Markov's

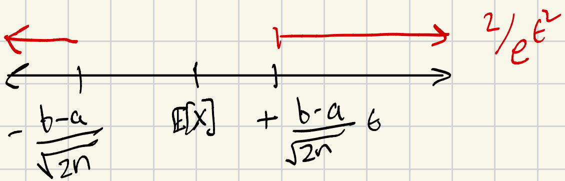


Chebyshev's



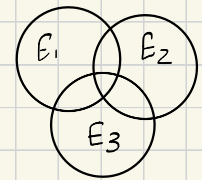
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad a \le X_i \le b$

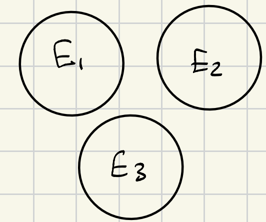
Hoeffding's



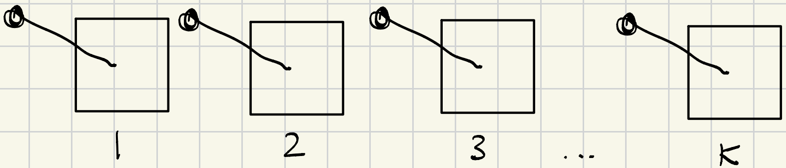
Union Bound

$Pr(E_1 + E_2 + \dots + E_m) \le Pr(E_1) + Pr(E_2) + \dots + Pr(E_m)$





Multi-armed Bandits



T time steps

Choose action (arm) $a^{(t)} \in \{1, \dots, k\}$ at time t

Receive reward $r_{a^{(t)}} \in [-1, 1]$

Expected reward of arm a: $\mu_a = \mathbb{E}[r_a]$

Goal: Minimize average regret

$$R_T = \max_{a \in \{1, \dots, k\}} \mu_a - \frac{1}{T} \sum_{t=1}^{T} \mu_{a^{(t)}}$$

Exploration vs. exploitation

Strategy: "optimism in the face of uncertainty"

Upper Confidence Bound (UCB) Algorithm

$t = \underbrace{1, 2, \ldots, K,}_{\text{Try each arm}} \; \underbrace{K+1, \ldots, T}_{\text{Choose optimistically best arm}}$

At time t,

- $\tilde{\mu}_a^{(t)}$ = empirical average of arm a so far $= \frac{1}{n_a^{(t)}} \sum_{\ell=1}^{t} \mathbb{1}[a^{(\ell)} = a] \, r_a(t)$
- $\epsilon_a^{(t)}$ = uncertainty of estimate so far

Informally, $|\mu_a - \tilde{\mu}_a^{(t)}| \leq \epsilon_a^{(t)}$ with high prob. e.g., wp $1-\delta$

Choose $\quad a^{(t)} = \underset{a}{\operatorname{argmax}} \; \tilde{\mu}_a^{(t)} + \epsilon_a^{(t)}$

Lemma: For all $t = K+1, \dots, T$ and $a = 1, \dots, K$ with probability $1-\delta$

$$|\mu_a - \tilde{\mu}_a^{(t)}| \le \epsilon_a^{(t)} = \sqrt{\frac{2\log(\frac{2TK}{\delta})}{n_a^{(t)}}}$$

Proof: Hoeffding's:

$$Pr\left(|\mathbb{E}[\bar{X}] - \bar{X}| \ge \epsilon \frac{b-a}{\sqrt{2n}}\right) \le 2e^{-\epsilon^2}$$

$b-a = 1-(-1) = 2$, $n = n_a^{(t)}$, $\tilde{\mu}_a^{(t)} = \bar{X}$

$$2e^{-\epsilon^2} \overset{\text{want}}{=} \frac{\delta}{TK} \iff \frac{2TK}{\delta} = e^{\epsilon^2}$$

$$\iff \sqrt{\log \frac{2TK}{\delta}} = \epsilon$$

$$\epsilon \frac{b-a}{\sqrt{2n}} = \sqrt{\frac{2\log \frac{2TK}{\delta}}{n_a^{(t)}}}$$

wp $\frac{\delta}{TK}$, $Pr\left(|\mu_a - \tilde{\mu}_a^{(t)}| \ge \sqrt{\frac{2\log\frac{2TK}{\delta}}{n_a^{(t)}}}\right) \le \frac{\delta}{TK}$

for one t and a

$$Pr\left(\bigcup_{t=1}^{T}\bigcup_{a=1}^{K} |\mu_a - \tilde{\mu}_a^{(t)}| \ge \epsilon_a^{(t)}\right) \overset{\text{union bound}}{\le} \sum_{t=1}^{T}\sum_{a=1}^{K} \frac{\delta}{TK} = \delta$$

Theorem: With prob $1-\delta$,

$$R_t = \max_a \mu_a - \frac{1}{T}\sum_{t=1}^{T} \mu_{a(t)} \leq O\left(\frac{K}{T} + \sqrt{\log \frac{TK}{\delta}}\ \frac{\sqrt{K}}{\sqrt{T}}\right)$$

$$\approx O\left(\sqrt{\frac{K}{T}}\right)$$

since 1) $K \ll T$

2) log (anything) = small

log (# atoms) $\approx$ 82

Intuition:

- Increase # arms by 100x, only 10x increase in regret
- Increase # steps by 100x, only 10x decrease in regret