Data 102 Spring 2022 Lecture 23

Bandits I

These slides are linked to from the course website, data102.org/sp22

Weekly Outline

- Last lecture: Concentration inequality.
    - How close are expectation to reality most of the time?
    - Markov, Chebyshev, Chernoff, Hoeffding, …

- This and next lecture: Multi-armed bandits.
    - Application of concentration inequalities to decision making

- Next up: Robustness

Announcements

- Homework 5 is due this Friday
- Vitamin will be released after class
- Midterm 2 is next Thursday
  - Includes today’s material.
- More info about extra credit and expectations for passing grades will be posted on Ed.

Markov Inequality

Let X be a **non-negative** random variable. For any $t > 0$,

$Z = X$

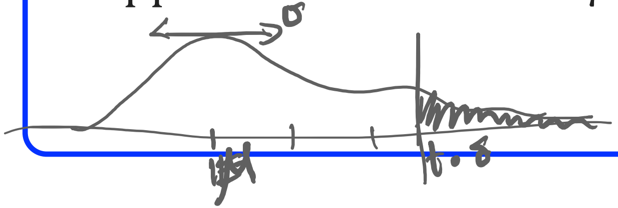
$$\Pr[X \geq t \cdot \mathbb{E}[X]] \leq \frac{1}{t}.$$

Chebyshev's Inequality

Suppose X has a mean of μ and standard deviation σ.

$$\Pr[|X - \mu| \geq t \cdot \sigma] \leq \frac{1}{t^2}$$

$\boxed{Z = |X - \mu|^2}$ in Markov ✓



$\exp(Z) = 1 + z + \frac{z^2}{2} + \cdots \frac{z^k}{k!} + \cdots$

Chernoff Bound

For any random variable X and t

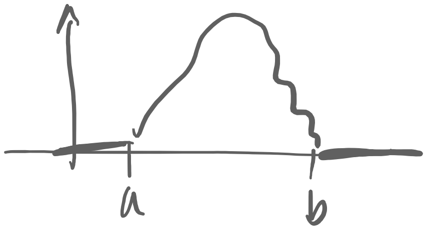
MGF

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[\exp(\lambda X)]}{\exp(\lambda t)} = \frac{M_X(\lambda)}{\exp(\lambda t)}$$

$Z = \exp(\lambda X)$ in Markov

(For all $\lambda \geq 0$)

MGFs for bounded random variables



Hoeffding's Lemma

Consider any random variable X whose mean is 0 and is bounded $i.e.,\underline{X \in [a,b]}$

$$M_X(\lambda) := \mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{(b-a)^2}{8}\lambda^2\right)$$

Implications of this when applied to Chernoff bound

$$\underbrace{\Pr[X \geq t] \leq \frac{M_X(\lambda)}{\exp(\lambda t)}}_{\text{Chernoff}} \overbrace{= \frac{\exp\left(\frac{(b-a)^2}{8}\lambda^2\right)}{\exp(\lambda t)}}^{\text{Hoeffding}} = \boxed{\exp\left(\frac{(b-a)^2}{8}\lambda^2 - \lambda t\right)}$$

$\mathbb{E}[\exp(\lambda S)] = \mathbb{E}[\exp(\frac{\lambda}{n}\sum x_i - \mu)] \rightarrow$ Multiplication Product of expo

Hoeffding's Inequality

Consider random variable $X_1, \dots X_n$ be i.i.d ~~independent~~ random variables with mean μ and bounded between a and b. Then

looks like a Gaussian tail

$\bar{\mu} - \mu > t$

$$\Pr\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)}_{S} \geq t\right] \leq \underline{\exp\left(-\frac{2nt^2}{(b-a)^2}\right)}$$

$n \approx \frac{1}{t^2} \times C$

$1/\sqrt{n}$

and

$\mu - \bar{\mu} > t$

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \leq -t\right] \leq \exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$

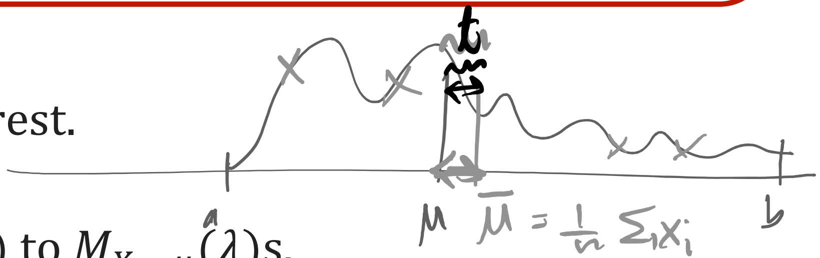
Proof idea:

1. Let $S = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$ be the variable of interest.
2. Compute MGF, $M_S(\lambda)$.

→ Independence should help decompose $M_S(\lambda)$ to $M_{X_i-\mu}(\lambda)$s.

→ $M_{X_i-\mu}(\lambda)$s are bounded by Hoeffing's Lemma

3. Put this in Chernoff inequality and optimize for λ.



$\frac{1}{n}\sum_i X_i$

$\mu = 0$ $\quad \frac{(b-a)^2}{8n} \cdot \frac{16 n^2 t^2}{(b-a)^4} = \frac{2nt^2}{(b-a)^2}$

$\lambda \cdot t = \frac{4nt^2}{(b-a)^2}$

1. Let S = $\boxed{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)}$ be the variable of interest.
2. Compute MFG, $M_S(\lambda)$.

→ Independence should help decompose $M_S(\lambda)$ to $M_{X_i-\mu}(\lambda)$s.

→ $M_{X_i-\mu}(\lambda)$s are bounded by Hoeffing's Lemma

3. Put this in Chernoff inequality and optimize for λ.

$\mathbb{E}[\prod] = \prod \mathbb{E}$

2

$$M_S(\lambda) = \mathbb{E}[\exp(\lambda S)] = \mathbb{E}\left[\exp\left(\frac{\lambda}{n}\sum_i X_i\right)\right] = \mathbb{E}\left\{\prod_{i=1}^{n} \exp\left(\frac{\lambda}{n} X_i\right)\right\}$$

(independence)

$$= \prod_{i=1}^{n} \mathbb{E}\left[\exp\left(\frac{\lambda}{n} X_i\right)\right] = \underline{\underline{\prod_{i=1}^{n} M_{X_i}\left(\frac{\lambda}{n}\right)}}$$

→ dependent on μ

↳ Hoeffding lemma: $M_{X_i}\left(\frac{\lambda}{n}\right) \le \exp\left(\frac{(b-a)^2}{8} \cdot \frac{\lambda^2}{n^2}\right)$

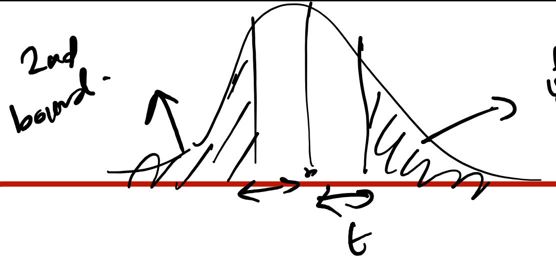
for any λ

$$\le \prod_{i=1}^{n} \exp\left(\frac{(b-a)^2}{8} \cdot \frac{\lambda^2}{n^2}\right) = \exp\left(\frac{(b-a)^2}{8} \cdot \frac{\lambda^2}{n}\right)$$

3 $\quad \Pr[S \geq t] \le \frac{M_S(\lambda)}{\exp(\lambda t)} = \min_{\lambda > 0} \frac{M_S(\lambda)}{\exp(\lambda t)} = \min_{\lambda > 0} \exp\left(\underline{\frac{(b-a)^2 \lambda^2}{8n} - \lambda t}\right)$

what's the best $\lambda = \frac{4nt}{(b-a)^2}$

$$= \exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$



first bound.

Hoeffding's Inequality

Consider random variable $X_1, \ldots X_n$ be i.i.d independent random variables with mean μ and bounded between a and b. Then

$S \geq t$

$\frac{1}{n}(X_i) = \bar{\mu}$

μ

$$\checkmark \Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \geq t\right] \leq \exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

and

$$\checkmark \Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \leq -t\right] \leq \exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$

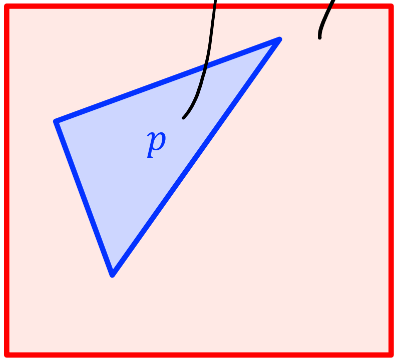
$-S \geq -t$

$$\Pr\left[|\bar{\mu} - \mu| \geq t\right] \leq 2 \times \exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

Applying Hoeffding's Inequality

m, ε, failure prob.?

$X_i = 1$ $X_i = 0$



A region of area p in a square of area 1. Throw m rocks uniformly in the square. How many rocks (k) fall in the triangle?

$$Pr\left[\left|\frac{k}{m} - p\right| > \epsilon\right] \leq 2\,exp(-2m\epsilon^2)$$

$= 0.28$ when $m = 100$ $\varepsilon = 0.1$

If $\boxed{m = 100}$ and $\boxed{\epsilon = 0.1}$

→ $\frac{k}{m}$ is within 0.1 of p, with high probability of 72%.

$0.72 = 1 - 2exp(-2 \times 100 \times (0.1)^2)$

Success with 72% to be close in 0.1 ⟹ you need $m \geq 100$.

Applying Hoeffding's Inequality

m , ε , failure

✓ ? ✓

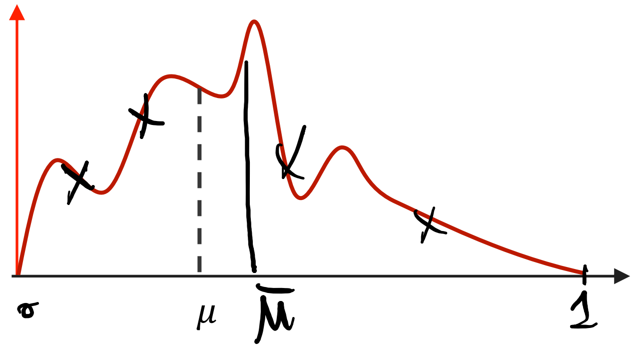
Confidence Interval: Sample m times from a distribution over [0,1]. And take their average $\bar{\mu}$. How close is $\bar{\mu}$ to the true mean μ?

$$\Pr[|\mu - \bar{\mu}| > \epsilon] \leq 2\exp(-2m\epsilon^2)$$

90**% confidence interval:**

- If $m = 100$, then $[\bar{\mu} - 0.12\,, \bar{\mu} + 0.12]$ is a 90% confidence interval.
- i.e, $\mu \in [\bar{\mu} - 0.12\,, \bar{\mu} + 0.12]$ with 90% probability.

$$\boxed{\begin{array}{c} 2\exp(-2m\varepsilon^2) = 0.1 \\ m = 100 \end{array}} \Longrightarrow \varepsilon = 0.12$$



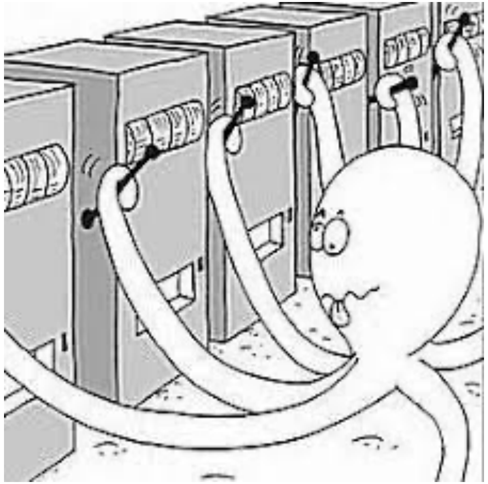
Multi-armed bandits

You step into a casino, and you see k slot machines.

For each machine i, every time you use the machine you get a random payoff, whose expectation is μ_i.

You don't know μ_i's, so you don't know the best machine.

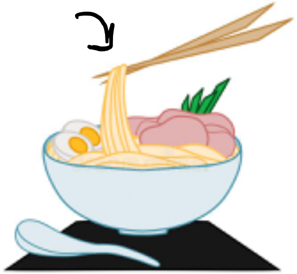
How should you use these machines to get the most payoff?



Multi-armed bandits, the non-gamblers' edition

Your new go-to restaurant has k dishes.

1. Each dish has an unknown μ_i: deliciousness
2. When you order a dish, you experience $\mu_i + noise$.
3. How do you select order?



3 , 2

4

Challenges:

1. No try, no information
2. Tradeoff: Exploration versus Exploitation
3. There is noise (or stochasticity) in the outcomes.

Other Examples

Advertising.

Oil drilling.

A/B testing: Market researching two options.

What are some examples you can think about?

Mathematical Setup:

There are k arms:

- Each arm i has a "payoff distribution" P_i, with mean μ_i.
- μ_is are unknown , P_i are unknown .

At every round $t = 1, 2, \dots, T$

- You pull one arm i_t.
- You observed $X_t \sim P_{i_t}$.

Hypothetically, if you knew $i^* = argmax\, \mu_i$ you should keep pulling that.

$\mathbb{E}[\text{hypothetical reward}] = \mu_{i^*} := \mu^*$

Goal: Collect as much expected reward as possible compared to the best arm:

$\frac{\hat{R}_T}{T} \to 0$ as $T \to \infty$ $\Rightarrow$ no-regret alg

(psuedo)Regret: $\hat{R}_T = T\mu_{i^*} - \sum_{t=1}^{T} \mathbb{E}[X_t] = T\mu^* - \sum_{t=1}^{T} \mu_{i_t}$ $\in o(T)$

$\hat{R}_T = 5$ ✓ , $\hat{R}_T = \sqrt{T}$ ✓

$\hat{R} = T$ ✗

Comparison to Regression

Logistic Regression

All data is given ahead of time
→ Called **Batch / Offline**

Has features x and values y.

Multi-armed bandits

Data is collected as we go
→ Called **Sequential/ Online**

No features, just values
→ There is a version of bandits with features
→ Called **contextual bandits**

Demo

Optimistic view.

Upper Confidence Bound (UCB) Algorithm

Idea: Pull the arm that has the highest "upper" confidence bound.

$UCB_i(t)$ = upper confidence bound for arm i in round t

UCB Alg: At time t play $i^t = \arg\max_i UCB_i(t)$.

How do we compute the upper confidence bound?

- Recall Hoeffding!
- Let $T_i(t)$: # of times arm i has been pulled up to time t. n_i
- Let $\hat{\mu}_i(t)$: average of the observed rewards on arm i in those $T_i(t)$ pulls

Constant

$$UCB_i(t) = \hat{\mu}_i(t) + \sqrt{\frac{C_t}{T_i(t)}} = \infty \text{ if } T_i(t) = 0.$$

By Hoeffding

$\frac{1}{\sqrt{m}} \times C_t$