Data 102 Spring 2022
Lecture 22

Concentration Inequalities

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

Weekly Outline

- Last week: Repeated decision making with feedback
    - Reinforcement learning
- **This lecture: Concentration inequality.**
    - **How close are expectation to reality most of the time?**
- Next up: Application of concentration inequalities to decision making.

Announcements

- Homework 5 is due this Friday
- Project proposals instructions and rubrics will go out today

- Midterm 2 is next Thursday
  - More info on Ed soon,
  - Including material covered this week.
  - Start studying now and come to OH.
- More info coming on Thursday about extra credit and expectations for passing grades.

Expectation versus Reality

Question: There is a slot machine with a payout, whose expected value is \$5. Let's say p is the probability that the payout is \$100. What are the plausible values for p?

A) 2% ✓ B) 4% ✓ C) 5% ✓ D) 7% ✗

C) plausible 95% payout 0, 5% payout 100

$\mathbb{E}[\text{payout}] = 0.95 \times 0 + 0.05 \times 100 = 5$ ✓

B) 4% was payout 100, 1% on payout 99, 1% on payout 1, 94% on payout 0

$0.04 \times 100 + 0.01\,(99+1) + 0 = 5$ ✓

D not possible: $\mathbb{E}[\text{payout}] = 0.07 \times 100 + \underbrace{\cdots\cdots}_{\text{nonnegative}} = \int \Pr(\text{payout} = t) \cdot t \, dt$

$= 7 + \cdots\cdots > 5$

Fact: Can upper bound $\Pr(X \geq t)$ in terms of its expectation.

15

Markov Inequality

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

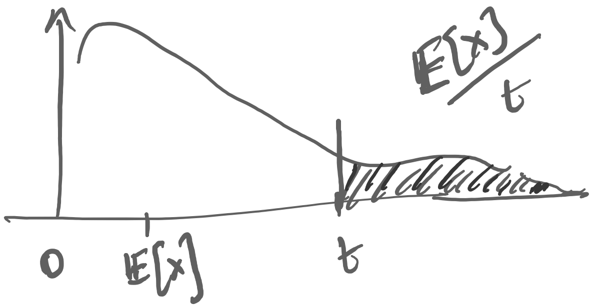
$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t} \checkmark$$

Alternatively,

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

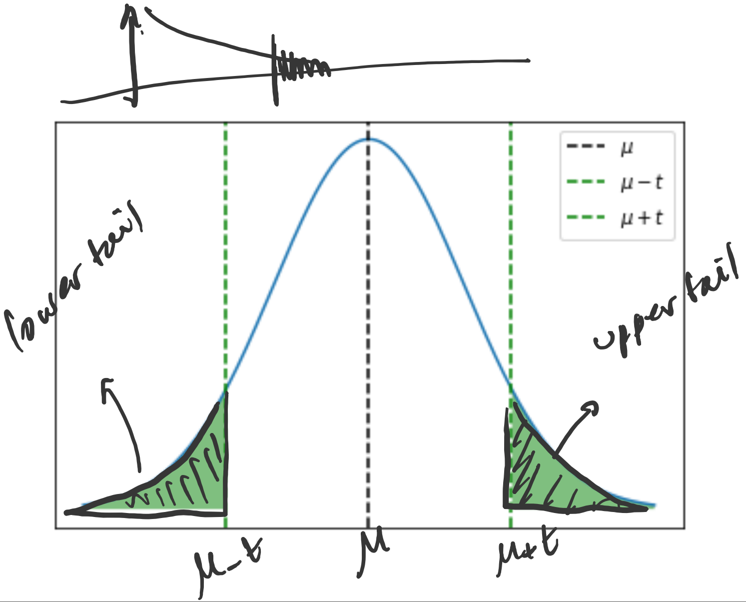
Proof of Markov's inequality:

$$\mathbb{E}[X] = \int \Pr(X = t) \cdot t \; dt \geq t \cdot \Pr[X \geq t]$$



Concentration Inequalities Generally

Concentration inequalities provide bounds on how a random variable deviates from its mean.



Benefits:

1. When X's distribution is unknown
2. No closed-form for X's tail
3. Result of complex combination of other random variables.

Interpretation of Markov’s Inequality

Sum of Random Variables

Question: You have 10 coins with probabilities of coin flipping to head being $p_1, p_2, \ldots, p_{10}$. Let's say $p_1 + p_2 + \cdots + p_{10} = 1$. What's the maximum probability that all 10 coins come up heads, using Markov's inequality? What is the random variable you are using?

$X_i = 1$ if coin i comes up heads

$Y = X_1 + X_2 + \cdots + X_{10}$

All of the coins are heads, $\Pr[Y \geq 10] \leq \frac{\mathbb{E}[Y]}{10}$ $\boxed{\leq \frac{1}{10}}$

$\mathbb{E}[Y] = \mathbb{E}[X_1] + \cdots \mathbb{E}[X_{10}] = p_1 + p_2 + \cdots + p_{10} = 1$

Better bound: Prob of all heads $p_1 \times p_2 \times \cdots \times p_n$ $\boxed{= \left(\frac{1}{10}\right)^{10}}$

$p_1 = p_2 = \cdots = p_n = \frac{1}{10}$

Can we do any better?

Better Concentration Inequalities for sum of variables

Issues with Markov's inequality:

Doesn't improve with summation independent values. independence.

- → Law of large numbers implies that sum of random variables tends to a Gaussian distribution. Gaussians have small tail probabilities.
- → We need to consider and leverage variance of random variables.
- → Application of Markov with Linearity of Expectation in last slide doesn't leverage variance.

Leveraging Variance in Application of Markov Inequality

Chebychev's Inequality

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

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

Apply Markov's inequality to the non-negative variable $Z = (X - \mu)^2$

$\mathbb{E}[Z] = \mathbb{E}[(X-\mu)^2] = \sigma^2$ variance.

$\Pr[|X-\mu| \geq t \cdot \sigma] = \Pr[(X-\mu)^2 \geq t^2\sigma^2] = \Pr[Z \geq t^2\sigma^2] = \Pr[Z \geq t^2\, \mathbb{E}[Z]]$

$\leq \frac{1}{t^2}$

$|X - \mu| \geq t \cdot \sigma$

Revisiting Coin Flips: Applying Chebychev's Inequality

You have 10 coins with $p_1, p_2, \dots, p_{10}$ of each coming up heads and $p_1 + p_2 + \dots + p_{10} = 1$. What's the maximum probability that all 10 coins come up heads?

X_i = outcome of whether coin i is heads.

$Y = X_1 + \dots + X_{10}$

$Var(X_i) = p_i(1-p_i) \leq p_i$

$\mathbb{E}[Y] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_{10}] = p_1 + \dots + p_{10} = 1$

$Var(Y) = Var(X_1) + \dots + Var(X_{10}) \leq p_1 + p_2 + \dots p_{10} \leq 1$

using independence.

Chebyshev: $\Pr[Y \geq 10] = \Pr[Y - 1 \geq 9] \leq \Pr[|Y-1| \geq 9] \leq \Pr[|Y-1| \geq 9 \cdot \sigma]$

$\boxed{\text{Ideal bound } \left(\frac{1}{10}\right)^{10}}$

$\boxed{\text{Markov } \frac{1}{10}}$

$\boxed{\leq \frac{1}{9^2} = \frac{1}{81} \approx \frac{1}{10^2}}$ Chebyshev

Markov VS. Chebychev

$$\Pr[X \geq t \cdot \mu] \leq \frac{1}{t}$$

Slower decay

Only Mean

$Z = (X - \mu)^2$ in Markov gives Chebychev.

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

faster decay $\frac{1}{t^2}$

Depends on Mean & Variance.

↳ Good for Sum/Averages/etc. independent random variable.

Var(Sum) ~~&~~ average

independence

Idea: Taking Markov and Chebychev to the next level!

$\mathbb{E}[X]$

$Z = (X - \mu)^2$

Chebychev's inequality $\Pr[|X - \mu| \geq t] \leq \frac{\mathbb{E}[|X-\mu|^2]}{t^2}$

$$\Pr[|X - \mu| \geq t] \leq \frac{\mathbb{E}[|X-\mu|^3]}{t^3} \qquad \boldsymbol{Z = |X - \mu|^3} \text{ in Markov}$$

$$\Pr[|X - \mu| \geq t] \leq \frac{\mathbb{E}[|X-\mu|^4]}{t^4} \qquad \boldsymbol{Z = |X - \mu|^4} \text{ in Markov}$$

Alternatively, what if we consider $Z = \exp(X)$ or even $\exp(\lambda X)$?

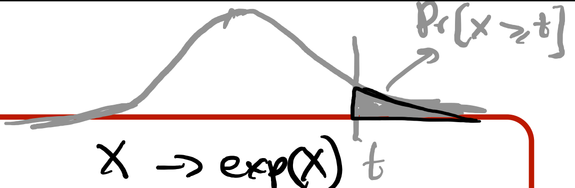
→At a high level, $\exp(\lambda X)$ includes all the higher powers (called moments)

$$\mathbb{E}\left[\exp(z)\right] = 1 + \mathbb{E}\left[z\right] + \mathbb{E}\left[\frac{1}{2}z^2\right] + \mathbb{E}\left[\frac{1}{6}z^3\right] + \cdots + \frac{1}{k!}z^k + \cdots$$

Mean Var 3rd

→ $\mathbb{E}[\exp(\lambda X)]$ is called the ***Moment Generating Function*** $\boldsymbol{M_X}(\lambda)$. (MGF)

→ Markov together with MGF is called the Chernoff bound.



Chernoff Bound

For any random variable X and t

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

(For all $\lambda \geq 0$)

$$\Pr[X \geq t] \leq \min_{\lambda \geq 0} \frac{M_X(\lambda)}{\exp(\lambda t)}$$

$X \to \exp(X)$

$\mathbb{E}[\exp(X)]$ small

$\hookrightarrow$ tail is light

This gives us improved composition for sum of random variables

Let $X_1 \cdots X_n$ be independent $\quad Y = X_1 + \cdots + X_n$, $\lambda = 1$

$$\underbrace{\mathbb{E}[\exp(Y)]}_{M_Y(1)} = \mathbb{E}[\exp(X_1 + \cdots X_n)] = \mathbb{E}[\exp(X_1) \cdot \exp(X_2) \cdot \cdots \cdot \exp(X_n)]$$

$$\underset{(\text{independt})}{=} \prod_{i=1}^{n} \mathbb{E}[\exp(X_i)] = \prod_{i=1}^{n} M_{X_i}(1)$$

shrink the tail faster

Get better decay bound.

Revisiting Coin Flips: Applying Chernoff

You have 10 coins with $p_1, p_2, \ldots, p_{10}$ of each coming up heads and $p_1 + p_2 + \cdots + p_{10} = 1$. What's the maximum probability that all 10 coins come up heads?

$X_i = \begin{cases} 1 & \text{heads} \quad p_i \\ 0 & \text{tail} \quad 1-p_i \end{cases} \rightarrow \underline{\mathbb{E}[\exp(\lambda X_i)]} = \Pr(X_i = 1)\cdot \exp(\lambda) + \Pr(X_i = 0)\cdot \exp(0)$

$= p_i \cdot e^{\lambda} + (1-p_i) = 1 + p_i(e^{\lambda}-1) \leq \exp(p_i(e^{\lambda}-1))$

$\boxed{1 + x \leq \exp(x)}$

$Y = X_1 + \cdots X_{10}$

$\mathbb{E}[\exp(\lambda Y)] = \prod_{i=1}^{10} \mathbb{E}[\exp(\lambda X_i)] \leq \prod_{i=1}^{10} \exp(p_i(e^{\lambda}-1))$

$= \exp\left(\sum_{i=1}^{10} p_i(e^{\lambda}-1)\right) = \exp(e^{\lambda}-1)$

Chernoff: $\Pr[Y \geq 10] \leq \dfrac{\mathbb{E}[\exp(\lambda Y)]}{\exp(\lambda 10)} \leq \dfrac{\exp(e^{\lambda}-1)}{\exp(\lambda 10)} \approx \boxed{10^{-3.6}}$ Chernoff

min $\lambda > 0$

$\lambda = \ln(10)$ $\quad \approx 10^{-6.1}$ Careful Chernoff $\quad (10)^{-10} =$ Best

10^{-2}: Chebyshev

$1/10$ Markov

Beyond Coin Flip: Dealing with all bounded random variables

Hoeffding's Lemma

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

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

$\Pr[X \in [a,b]] = 1$

Implications of this when applied to Chernoff bound

$$\Pr[X \geq t] \leq \frac{M_X(\lambda)}{\exp(\lambda t)} = \frac{\exp\left(\frac{(b-a)^2}{8}\lambda^2\right)}{\exp(\lambda t)} = \exp\left(\frac{(b-a)^2}{8}\lambda^2 - \lambda t\right)$$

$\Downarrow$

$X_1 + \cdots + X_n$

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

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

and

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

Proof idea