Quiz grading:
↳ currently being graded

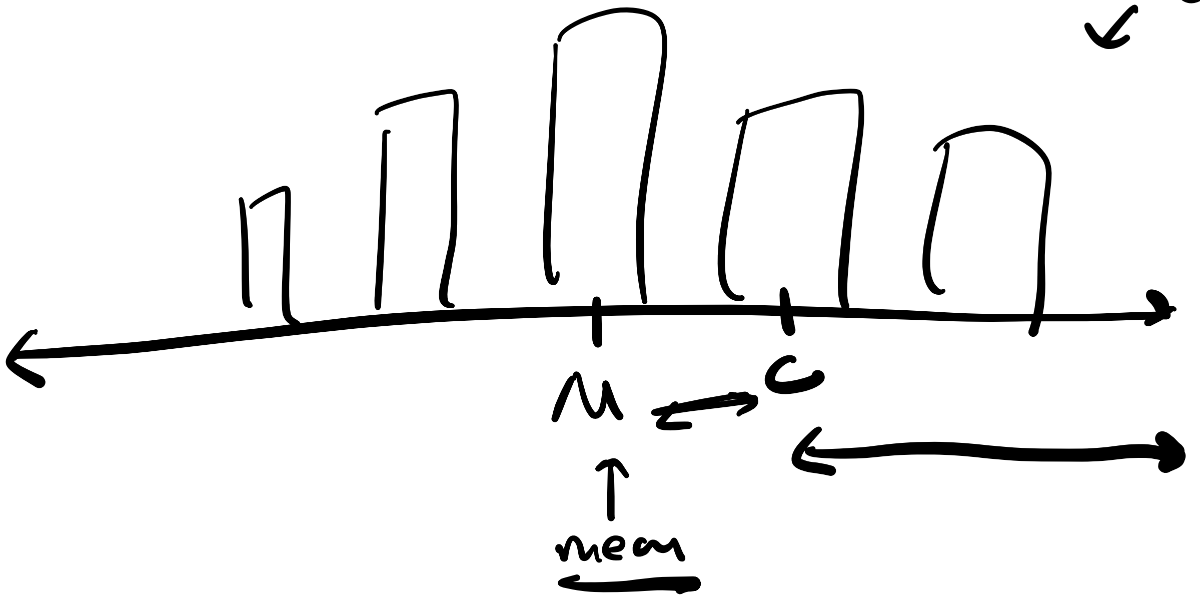
Taylor Swift tomorrow

Lecture 5D: Concentration Inequalities

UC Berkeley CS70
Summer 2023
Nikki Suzani

How do we bound probabilities?

don't see dist.



Option 1: Markov's Inequality

always have to check RV is non-negative

Dealing with the definition of expectation – looking at **non-negative random variables.**

Option 1: Markov's Inequality

Theorem 17.1 (Markov's Inequality). *For a* nonnegative *random variable* X *(i.e.,* $X(\omega) \geq 0$ *for all* $\omega \in \Omega$*) with finite mean,*

$$\mathbb{P}[X \geq c] \leq \frac{\mathbb{E}[X]}{c},$$

arbitrary

for any positive constant c.

Emphasis: **Non-Negative Random Variable**

Markov's Inequality Proof

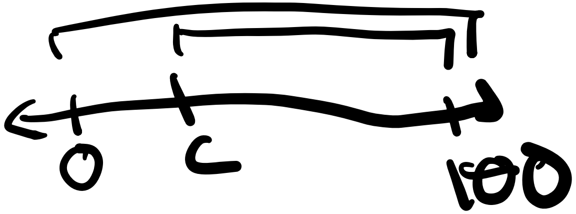
$$P[X \ge c] \le \frac{E(X)}{c}$$

$$E[X] = \sum_k k P(X=k)$$

entire sum

$$\ge \sum_{k \ge c} k P(X=k)$$

subset



$$\ge \sum_{k \ge c} c P(X=k)$$

$$= c \sum_{k \ge c} P(X=k)$$

$$= c P(X \ge c)$$

$c \to \infty$

Markov's Inequality Proof

$$E[X] \geq cP(X \geq c)$$

$$cP(X \geq c) \leq E(X)$$

$$P(X \geq c) \leq \frac{E(X)}{c}$$

Markov's Inequality

$c = 50$

$k = 50, 75, 100$

$50P(X=50)$

$+75$

$c = 50$

$k = 50, 75, 100$

$50P(X=50)$

$+75P(X=75) + 100P(X=100)$

$\underline{50}P(X=50) + \underline{50}P(X=75)$

$+\underline{50}P(X=100)$

$\overset{1/4}{-100}, \overset{1/4}{-50}, \overset{1/4}{0}, \overset{1/4}{50}$

smaller number

$-100 \cdot \frac{1}{4} + -50 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 50 \cdot \frac{1}{4}$

$c = 25$

$\rightarrow 50 \cdot \frac{1}{4}$

$k \geq c$

What happens if we **know** the (negative) lower bound?

Can we still use Markov's?

X: smallest number X can take on is -100

$$\boxed{Y = X + 100}$$

non-negative

Coin Tosses Example

We toss a coin with probability ⅓ of being heads n times. What is the probability of getting more than 50% heads?

toss it n times

$\geq \frac{n}{2}$ heads

$\frac{1}{3}$ of heads, n times

X: # of heads

$X \sim \text{Binomial}(n, \frac{1}{3})$

$E[X] = \frac{n}{3}$

$$P(X \geq c) \leq \frac{E(X)}{c}$$

$$P\left(X \geq \frac{n}{2}\right) \leq \frac{E(X)}{\left(\frac{n}{2}\right)} = \frac{\left(\frac{n}{3}\right)}{\left(\frac{n}{2}\right)} = 2/3$$

Cheating Chess Players

Chess Professionals take an average of 5 trips away from the board during a standard chess game. What is an upper bound on the probability that Hans Niemann takes 25 trips away?

$E[X]=5$

Markov's:

$$P(X \geq c) \leq \frac{E(X)}{c}$$

X is non-negative

$$P(X \geq 25) \leq \frac{E(X)}{25} = \frac{5}{25} = \frac{1}{5}$$

smallest at the equality stage

$$P(X < c) = 1 - \underbrace{\boxed{P(X \geq c)}}_{\text{upper bounded}}$$

biggest at the equality stage

More Chess Players

Chess Professionals take an average of 5 trips away from the board during a standard chess game. What is a bound on the probability that Magnus Carlsen takes less than 2 trips away?

$$P(X \geq c) \leq \frac{E(X)}{c}$$

manipulate this inequality

$$-P(X \geq c) \geq -\frac{E(X)}{c}$$

$$1 - P(X \geq c) \geq 1 - \frac{E(X)}{c}$$

$$P(X < c) \geq 1 - \frac{E(X)}{c}$$

More Chess Players

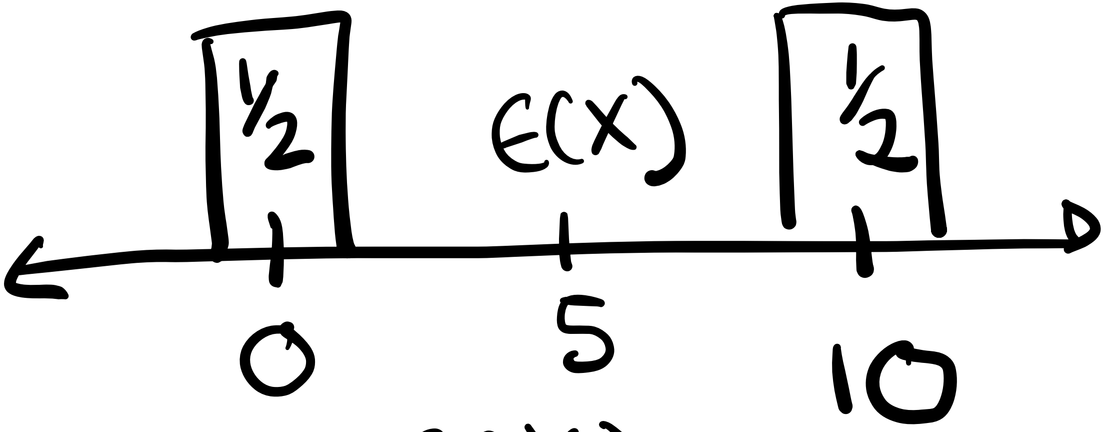
Chess Professionals take an average of 5 trips away from the board during a standard chess game. What is a bound on the probability that Magnus Carlsen takes less than 2 trips away?

$$P(X < 2) \geq 1 - \frac{E(X)}{2}$$

$$\geq 1 - \frac{5}{2}$$

$$\geq -3/2$$

When is Markov's a **tight** bound?



$$P(X \geq 10) \leq \frac{E(X)}{10} = \frac{5}{10} = \frac{1}{2}$$

Generalized Markov's Inequality

Theorem 17.2 (Generalized Markov's Inequality). *Let Y be an arbitrary random variable with finite mean. Then, for any positive constants c and r,*

$$\mathbb{P}[|Y| \geq c] \leq \frac{\mathbb{E}[|Y|^r]}{c^r}.$$

Generalized Markov's Inequality Proof

Let X be an indicator variable for |Y| ~~meaning~~ being greater than or equal to c.

$X: \begin{cases} 1 & \text{if } |Y| \geq c \\ 0 & \text{otherwise} \end{cases}$

$X \sim \text{Bernoulli}(P(|Y| \geq c))$

$E[X] = \underbrace{P(|Y| \geq c)}$

$|Y|^r \geq |Y|^r X \geq c^r X$

$E(|Y|^r) \geq E[c^r X] = c^r E[X] = c^r P(|Y| \geq c)$

$$E(|Y|^r) \geq c^r P(|Y| \geq c)$$

$$c^r P(|Y| \geq c) \leq E(|Y|^r)$$

$$P(|Y| \geq c) \leq \frac{E(|Y|^r)}{c^r}$$ ✓

Option 2: Chebyshev's Inequality

Variance measures the deviation from the mean – can we use that information?

↳ create a better bound

Option 2: Chebyshev's Inequality (Formally)

Theorem 17.3 (Chebyshev's Inequality). *For a random variable X with finite expectation $\mathbb{E}[X] = \mu$,*

$$\mathbb{P}[|X-\mu| \geq c] \leq \frac{\operatorname{Var}(X)}{c^2}, \tag{2}$$

and for any positive constant c.

probability that X deviates from its mean by more than c, $\leq \frac{\operatorname{Var}(X)}{c^2}$

μ X

$> c$

Proof of Chebyshev's Inequality

$Y = (X - \mu)^2$ non-negative R.V.

$P(Y \geq c^2) \leq \frac{E(Y)}{c^2}$

$E(Y) = E((X-\mu)^2)$
$= Var(X)$

$P(Y \geq c^2) \leq \frac{Var(X)}{c^2}$

$P(\underbrace{(X-\mu)^2 \geq c^2}) \leq \frac{Var(X)}{c^2} \rightarrow \boxed{P(|X-\mu| \geq c) \leq \frac{Var(X)}{c^2}}$

any X

Generalized Chebyshev's Inequality

Corollary 17.1. *For any random variable X with finite expectation* $\mathbb{E}[X] = \mu$ *and finite standard deviation* $\sigma = \sqrt{\mathrm{Var}(X)}$,

$\sigma^2 = \mathrm{Var}(X)$

$$\mathbb{P}[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2},$$

for any constant $k > 0$.

$$P(|X-\mu| \geq k\sigma) \leq \frac{\mathrm{Var}(X)}{(k\sigma^2)} = \frac{\sigma^2}{(k\sigma)^2}$$

$$= \frac{1}{k^2}$$

Coin Tosses Again

We toss a coin with probability ⅓ of being heads n times. What is the probability of getting more than 50% heads? $\longrightarrow P = \frac{n}{2}$

X: # of heads $\quad X \sim \text{Binomial}(n, \frac{1}{3})$

$$Var(X) = np(1-p) = n \cdot \tfrac{1}{3} \cdot \tfrac{2}{3} = \frac{2n}{9}$$

$$E(X) = np = \frac{n}{3}$$

Chebyshev's

$P(|X-\mu| \geq c)$

$P(X \geq \frac{n}{2})$

$P(X-\frac{n}{3} \geq \frac{n}{2}-\frac{n}{3})$

$P(|X-\frac{n}{3}| \geq \frac{n}{6}) \leq \frac{Var(X)}{(\frac{n}{6})^2} = \frac{(\frac{2n}{9})}{(\frac{n}{6})^2}$

Let's talk about **estimation**

Say we have a coin – but don't know the probability it flips heads. How can we guess? Flip a coin a bunch of times → n times

X: # of heads after I flip my coin n times

$\frac{X}{N}$: guess for the probability it ~~flips heads~~

$X_i \sim \begin{cases} 1 & \text{if heads} \\ 0 & \text{otherwise} \end{cases}$

$$\boxed{|\hat{p} - p| \le \varepsilon}$$

$$\hat{p} = \frac{X_1 + X_2 + \dots X_n}{n}$$

Let's talk about **confidence**

We can't be <u>sure</u> that our estimate is good, though. So we want to bound the probability that we're wrong.

level of confidence I want to have

95% confident my guess is right

$$\boxed{\mathbb{P}[|\hat{p} - p| \leq \varepsilon] \geq \underbrace{1 - \delta}}$$

$$-P[|\hat{p}-p| \leq \varepsilon] \leq \delta - 1$$

$$1 - P[|\hat{p}-p| \leq \varepsilon] \leq \delta$$

$$\boxed{P[|\hat{p}-p| > \varepsilon] \leq \delta}$$

Estimating Probabilities

Let's put it into practice. How many samples do we need to be confident?

$X_i \sim \text{Bernoulli}(p)$

$$P[|\hat{p} - p| \geq \varepsilon] \leq \frac{Var(\hat{p})}{\varepsilon^2}$$

$$E[\hat{p}] = E\left[\frac{X_1 + X_2 + \dots X_n}{n}\right] = \frac{1}{n} E[X_1 + \dots X_n]$$

$$= \frac{1}{n} \cdot np = p$$

$$\rightarrow \underline{Var(\hat{p})} = Var\left(\frac{X_1+X_2+\dots X_n}{n}\right) = \frac{1}{n^2} Var\left(X_1 + X_2 + \dots X_n\right)$$

$$\frac{1}{n^2} Var(X_1+\dots X_n) = \frac{1}{n^2} \cdot n \, Var(X_1) = \frac{1}{n^2} \cdot n \cdot p(1-p)$$

$X_1 \sim Bernoulli(p) \quad = p(1-p)$

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

$$P[|\hat{p}-p| \geq \varepsilon] \leq \frac{p(1-p)}{n\varepsilon^2}$$

$$\frac{p(1-p)}{n\varepsilon^2} = \delta$$

max. at $p = 1/2$

$$n = \boxed{\frac{p(1-p)}{\delta\varepsilon^2}}$$

most it can be at $\frac{1}{4\delta\varepsilon^2}$

$$\boxed{n \geq \frac{1}{4\delta\varepsilon^2}}$$

$$n = \frac{\frac{1}{3}(1-\frac{1}{3})}{\delta\varepsilon^2}$$

2

$n \geq \frac{\quad}{9\delta\varepsilon^2}$

Estimating Probabilities

Let's put it into practice. How many samples do we need to be confident?

Estimating Probabilities

$$n \geq \max_p \frac{p(1-p)}{\varepsilon^2 \delta} = \boxed{\frac{1}{4\varepsilon^2 \delta}},$$

Example: How many people play chess?

We want to estimate the percent of people in CS70 who play chess. Let's say we want an error of 0.05, and confidence of 90%. How many people do I need to sample?

$$n \geq \frac{1}{4\varepsilon^2\delta}$$

$$1-\delta = 0.9$$

$$\delta = 0.1$$

$$\varepsilon = 0.05$$

$$n \geq \frac{1}{4(0.05)^2 0.1} = 1000$$

What about estimating a **value**?

$$\mathbb{P}[|\hat{\mu}-\mu| \geq \varepsilon\mu] \leq \delta.$$

$$P[|\hat{\mu}-\mu| \geq \varepsilon\mu] \leq \frac{Var(\hat{\mu})}{\varepsilon^2\mu^2} = \frac{\sigma^2}{\varepsilon^2\mu^2 n}$$

$Var(\hat{\mu})$

$$\frac{\sigma^2}{\varepsilon^2\mu^2 n} = \delta \qquad n = \frac{\sigma^2}{\varepsilon^2\mu^2\delta}$$

What about estimating a value?

$\frac{1}{n^2} Var(X_1) \cdot n$

$\frac{1}{n} Var(X_1) = \frac{\sigma^2}{n}$

$$n \geq \frac{\sigma^2}{\mu^2} \times \frac{1}{\varepsilon^2 \delta}.$$

Upper or $\boxed{\text{lower}}$ bound on μ?

$\boxed{\text{Upper}}$ or lower bound on σ?

Lower bound: A
Upper bound: B

Variance $= \sigma^2$ SD $= \sigma$

Average Rating of Chess Players on Course Staff

Many people on CS70 Course Staff play chess. Some are… better than others. Let's say our lower bound on the mean rating is 600, and our upper bound on variance is 100. We want to be 70% confident in our guess, with an error level of 0.05. How many people do we need to ask?

$\delta = 0.3$

$\varepsilon = 0.05$

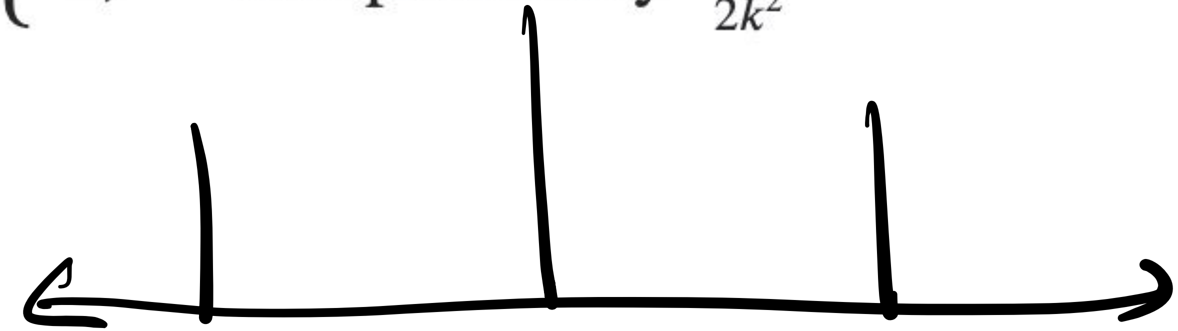
$$n \geq \frac{\sigma^2}{\mu^2} \times \frac{1}{\varepsilon^2 \delta}$$

$$n \geq \frac{100}{600^2} \times \frac{1}{0.3 \times (0.05)^2}$$

When is Chebyshev's a tight bound?

mean of 0 & SD of 1

$$Z = \begin{cases} -k, & \text{with probability } \frac{1}{2k^2} \\ 0, & \text{with probability } 1 - \frac{1}{k^2} \\ k, & \text{with probability } \frac{1}{2k^2} \end{cases}$$



The Law of Large Numbers

Intuition: We expect a random variable to **converge** to its average.

The Law of Large Numbers (Formally)

Theorem 17.4 (Law of Large Numbers). *Let $X_1, X_2, \ldots,$ be a sequence of i.i.d. (independent and identically distributed) random variables with common finite expectation $\mathbb{E}[X_i] = \mu$ for all i. Then, their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy*

sum all the times I flip a coin on
otherwise
simulate my
r.v.

$$\mathbb{P}\left[\left|\frac{1}{n}S_n - \mu\right| < \varepsilon\right] \to 1 \qquad \text{as } n \to \infty,$$

for every $\varepsilon > 0$, however small.

divide
it by the
total # of
times

Proof of LLN

$$P\left[\left|\frac{1}{n}S_n - \mu\right| \geq \varepsilon\right] \leq \frac{Var(X)}{\varepsilon^2}$$

$$Var\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}Var(S_n) = \frac{1}{n^2}Var(X_1 + X_2 + \dots X_n)$$

$$= \frac{1}{n^2}(Var(X_1) + Var(X_2) + \dots Var(X_n))$$

$$= \frac{1}{n^2} \cdot n\, Var(X_1)$$

$$= \frac{Var(X_1)}{n}$$

$$\leq \frac{Var(X_1)}{n\varepsilon^2} \Big]$$

$n \to \infty$

$\to 0$

(If time) Derivation of Variance of the Poisson

Let $X \sim \text{Poisson}(\lambda)$. What is $\text{Var}(X)$?

Recap

Learned about Concentration Inequalities:

- Markov's Inequality
  - Generalized Markov's Inequality
- Chebyshev's Inequality
  - Confidence Levels!!

Learned about the Law of Large Numbers:

- Proof using Chebyshev's!