Lecture 5D: Concentration Inequalities

UC Berkeley CS70
Summer 2023
Nikki Suzani

How do we bound probabilities?

Option 1: Markov's Inequality

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,*

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

for any positive constant c.

Emphasis: **Non-Negative Random Variable**

Markov's Inequality Proof

Markov's Inequality Proof

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

Can we still use Markov's?

Coin Tosses Example

We toss a coin with probability $\frac{1}{3}$ of being heads n times. What is the probability of getting more than 50% heads?

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?

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?

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?

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

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 greater than or equal to c.

Option 2: Chebyshev's Inequality

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

Option 2: Chebyshev's Inequality (Formally)

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

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

and for any positive constant c.

Proof of Chebyshev's Inequality

Generalized Chebyshev's Inequality

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

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

for any constant $k > 0$.

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?

Let’s talk about **estimation**

Say we have a coin – but don’t know the probability it flips heads. How can we guess?

$$|\hat{p} - p| \leq \varepsilon$$

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.

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

Estimating Probabilities

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

Estimating Probabilities

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

Estimating Probabilities

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

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?

What about estimating a **value**?

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

What about estimating a value?

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

Upper or lower bound on μ?

Upper or lower bound on σ?

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?

When is Chebyshev's a tight bound?

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

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*

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

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

Proof of LLN

(If time) Derivation of Variance of the Poisson

Let X ~ Poisson(λ). What is 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!