CS70 – Spring 2024

Lecture 21 – April 4

Review of Previous Lecture

- **Variance**: For a random variable with $E[X]=\mu$

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

Standard deviation: $\sigma(X) = \sqrt{Var(X)}$

Measures "spread" of the distribution

- To compute $E[X^2]$:

$$E[X^2] = \sum_a a^2 \times Pr[X=a]$$

Review (cont.)

- $X \sim Bin(n,p)$: $E[X] = np$ $Var[X] = np(1-p)$
- $X \sim Geom(p)$: $E[X] = \frac{1}{p}$ $Var[X] = \frac{1-p}{p^2}$
- $X \sim Poisson(\lambda)$: $E[X] = \lambda$ $Var[X] = \lambda$
- For any r.v. X and constant c

$$Var(cX) = c^2 \, Var(X)$$

- If X, Y are independent, then

$$Var(X+Y) = Var(X) + Var(Y)$$

Review (cont.)

- For <u>any</u> two r.v.'s X, Y:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

- <u>Covariance</u> $Cov(X,Y) = E[XY] - E[X]E[Y]$

$$= E[(X-\mu_x)(Y-\mu_y)]$$

- $Cov(X,Y) \begin{cases} >0 & : \text{ pos. correlation} \\ <0 & : \text{ neg. correlation} \end{cases}$

- $Corr(X,Y) = \dfrac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$ (lies in $[-1,+1]$)

Plan for Today

- Concentration inequalities : "how far is a r.v. away from its expectation?"
- Markov's Inequality
- Chebyshev's Inequality (based on Variance)
- Applications to Estimation
- Law of Large Numbers

Concentration Inequalities

Q: What are they?

A: Inequalities that tell us how far a r.v. X is likely to be from its expectation $E[X]$?

Q: Why is this useful?

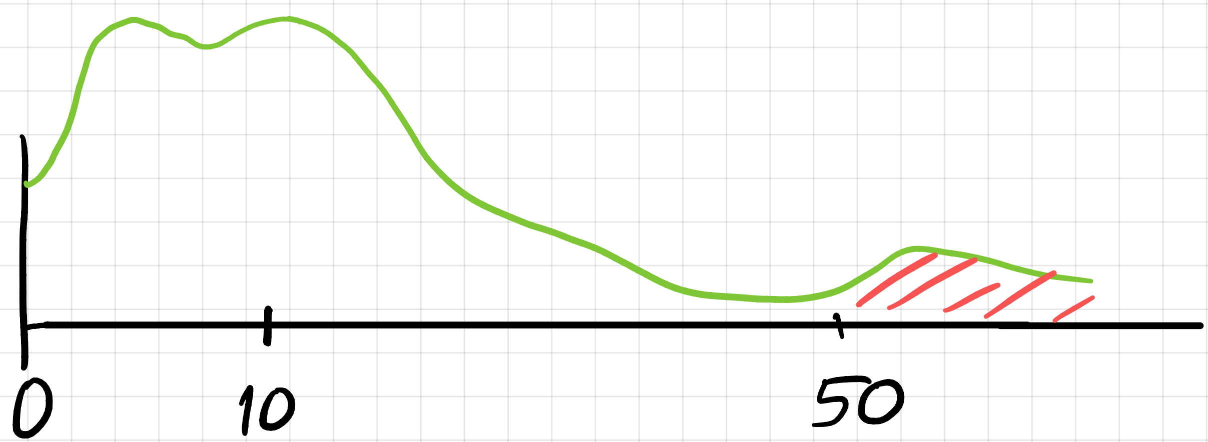
A: Expectations are easy to compute — so if X is close to $E[X]$, we have a lot of info. about X

Markov's Inequality

Example: Suppose I tell you:

1. Random variable X is non-negative (i.e., $X \geq 0$ always — w.prob. 1)
2. $E[X] = 10$

What can you tell me about $\Pr[X \geq 50]$?

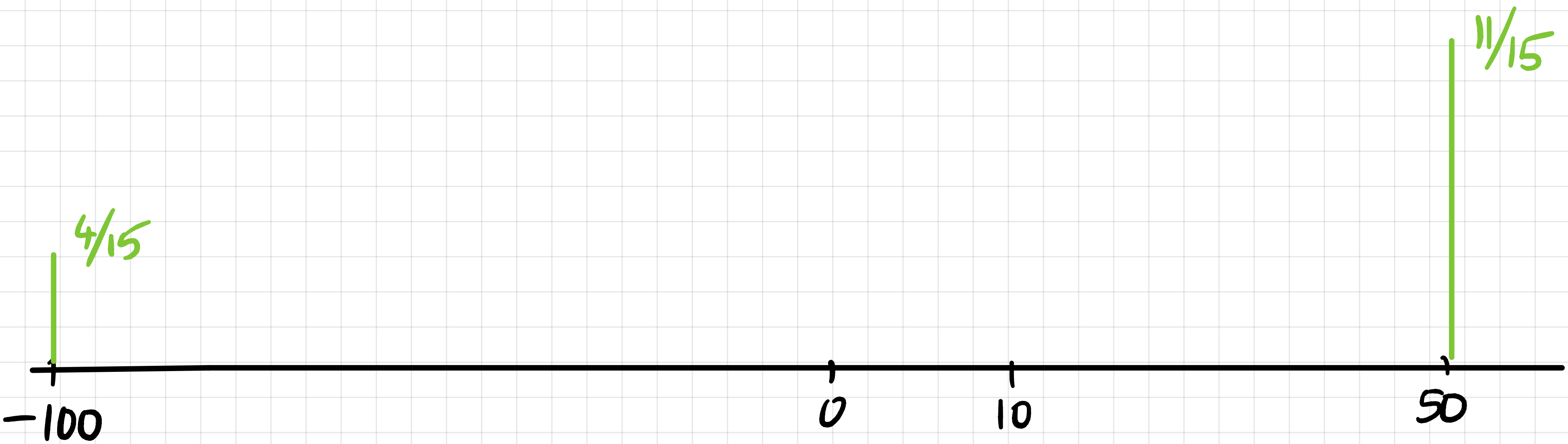


Markov's Inequality

Example: Suppose I tell you:

1. ~~Random variable X is non-negative (i.e., $X \geq 0$ always — w.prob. 1)~~
2. $E[X] = 10$

What can you tell me about $\Pr[X \geq 50]$?



$$E[X] = \frac{4}{15} \times (-100) + \frac{11}{15} \times 50 = 10$$

Theorem [Markov's Inequality]

For any non-negative random variable X and any c:

$$\boxed{\Pr[X \geq c] \leq \frac{1}{c} \times E[X]}$$

Proof: Suppose for contradiction that $\Pr[X \geq c] > \frac{1}{c} E[X]$.

By definition of $E[X]$:

$$E[X] = \sum_{a} a \times \Pr[X = a]$$

$$\geq \sum_{a \geq c} a \times \Pr[X = a] \qquad \text{because } X \geq 0!$$

$$\geq c \times \Pr[X \geq c]$$

Hence $\Pr[X \geq c] \leq \frac{1}{c} E[X]$ $\square$

Example: $X \sim \text{Binomial}(n, 1/2)$

Recall: $E[X] = np = n/2$

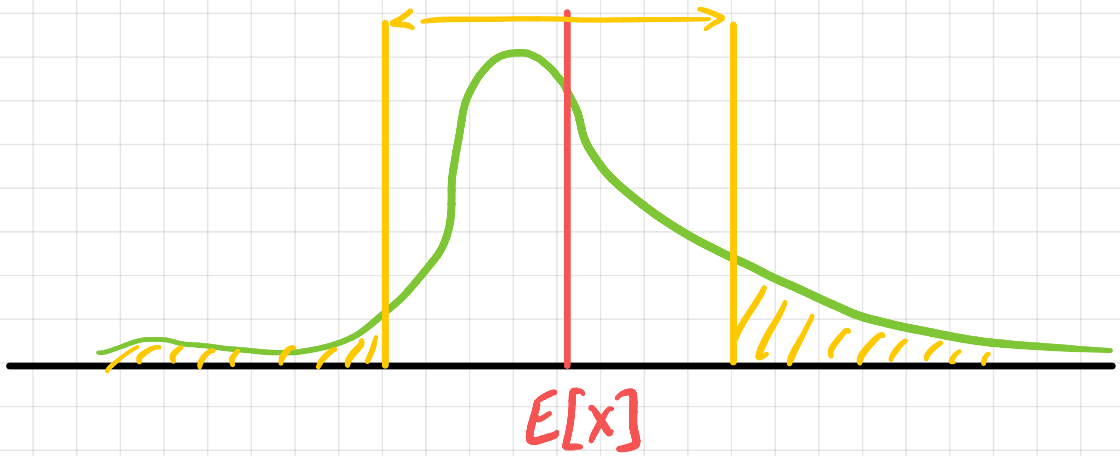
Markov: $\Pr[X \geq c] \leq \dfrac{E[X]}{c}$

$$\Rightarrow \Pr[X \geq 3n/4] \leq \frac{4}{3n} \times E[X] = \boxed{\frac{2}{3}}$$

Note: This upper bound is correct but far from the best bound we can get — see later!

Q: Suppose we also know Var(X) – does this help?

A: Yes! Recall that Var(X) measures expected (squared) distance of X from E[X]



If Var(X) is small, then the prob. that X is far from E[X] should be small

Chebyshev's Inequality

Theorem: For any r.v. X and any c:

$$\boxed{\Pr[|X - E[X]| \geq c] \leq \frac{Var(X)}{c^2}}$$

Compare with Markov:

- Doesn't require X to be non-negative
- Gives a two-sided bound (above and below $E[X]$)
- c is replaced by c^2

Chebyshev's Inequality

Theorem: For any r.v. X and any c:

$$\boxed{\Pr\left[|X - E[X]| \geq c\right] \leq \frac{Var(X)}{c^2}}$$

Proof: Define the r.v. $Y = (X - E[X])^2$

Note that Y is non-negative so we can apply Markov's inequality to it:

$$\Pr[Y \geq c^2] \leq \frac{E[Y]}{c^2}$$

i.e. $$\Pr\left[(X - E[X])^2 \geq c^2\right] \leq \frac{E\left[(X - E[X])^2\right]}{c^2}$$

i.e. $$\Pr\left[|X - E[X]| \geq c\right] \leq \frac{Var(X)}{c^2} \qquad \square$$

Example: $X \sim \text{Binomial}(n, 1/2)$

Recall: $E[X] = np = n/2$ $\qquad Var(X) = np(1-p) = n/4$

Chebyshev: $\Pr[|X - E[X]| \geq c] \leq \dfrac{Var(X)}{c^2}$

$$\Rightarrow \quad \Pr\left[X \geq \frac{3n}{4}\right] \leq \Pr\left[|X - E[X]| \geq n/4\right]$$

$$\leq \frac{Var(X)}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \boxed{\frac{4}{n}}$$

! This is <u>much better</u> than Markov (which gave us $\Pr\left[X \geq \frac{3n}{4}\right] \leq \frac{2}{3}$)

Equivalent Statement of Chebyshev

For any r.v. X:

$$\boxed{\Pr\left[|X - E[X]| \geq k\,\sigma(X)\right] \leq \frac{1}{k^2}}$$

Proof: Plug in $c = k\,\sigma(X)$ to Chebyshev:

$$\Pr\left[|X - E[X]| \geq k\sigma(X)\right] \leq \frac{Var(X)}{(k\,\sigma(X))^2}$$

$$= \frac{Var(X)}{k^2\,Var(X)} = \frac{1}{k^2}$$

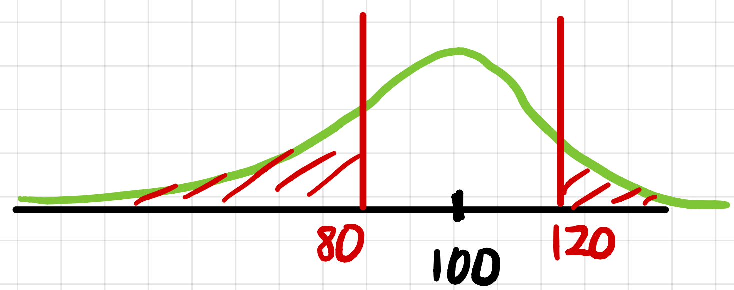
Example: For any r.v. X, the probability of being more than 2 s.d.'s from mean is $\leq 1/4$

Example: $\quad X \sim \text{Poisson}(\lambda)$

Recall $\quad E[X] = \lambda \qquad Var(X) = \lambda \qquad \sigma(X) = \sqrt{\lambda}$

Chebyshev: $\quad \Pr[|X-\lambda| \geq k\sqrt{\lambda}] \leq \frac{1}{k^2}$

E.g. $\lambda = 100 \longrightarrow \quad \Pr[|X-100| \geq 20] \leq 1/4$



Application: Statistical Estimation

Goal: Estimate the proportion of smokers in the population within $\pm 1\%$ with confidence $\geq 95\%$

"Opinion Poll": Take a random sample of N people

Ask each person if they're a smoker

Output the fraction of the sample that says "Yes"

Key Question: $\boxed{\text{How large does } N \text{ have to be to ensure accuracy } \pm 1\% \text{ \& confidence } 95\%?}$

Note: Assume for simplicity we choose people <u>with</u> replacement so that samples are all <u>independent</u>

The Math

Define r.v. S_N by

$$S_N = X_1 + X_2 + \cdots + X_N \quad \text{where} \quad X_i = \begin{cases} 1 & \text{if person } i \text{ says "yes"} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Output } \hat{p} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

← our estimate of the true unknown proportion p

The Math

Define r.v. S_N by

$$S_N = X_1 + X_2 + \cdots x X_N \quad \text{where} \quad X_i = \begin{cases} 1 & \text{if person } i \text{ says "yes"} \\ 0 & \text{otherwise} \end{cases}$$

Output $$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

our estimate of the true unknown proportion p

$$E[\hat{p}] = \frac{1}{N} \sum_{i=1} E[X_i] = \frac{1}{N} \times Np = \boxed{p}$$

"unbiased estimator"

$$Var(\hat{p}) = \frac{1}{N^2} \sum_{i=1} Var(X_i) = \frac{1}{N^2} \times Np(1-p) = \boxed{\frac{p(1-p)}{N}}$$

decreases with N!

The Math

Define r.v. S_N by

$$S_N = X_1 + X_2 + \dots x X_N \quad \text{where} \quad X_i = \begin{cases} 1 & \text{if person } i \text{ says "Yes"} \\ 0 & \text{otherwise} \end{cases}$$

Output $$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

our estimate of the true unknown proportion p

$$E[\hat{p}] = \frac{1}{N} \sum_{i=1} E[X_i] = \frac{1}{N} \times Np = \boxed{p}$$

"unbiased estimator"

$$Var(\hat{p}) = \frac{1}{N^2} \sum_{i=1} Var(X_i) = \frac{1}{N^2} \times Np(1-p) = \boxed{\frac{p(1-p)}{N}}$$

decreases with N !

Chebyshev: $$\Pr[|\hat{p} - p| \geq \varepsilon] \leq \frac{Var(\hat{p})}{\varepsilon^2} = \frac{p(1-p)}{\varepsilon^2 N}$$

$$\leq \boxed{\frac{1}{4\varepsilon^2 N}}$$

Chebyshev: $\Pr[|\hat{p}-p| \geqslant \varepsilon] \leq \boxed{\frac{1}{4\varepsilon^2 N}}$

Recall: we want $\Pr[|\hat{p}-p| \geqslant \underset{\pm 1\%}{0.01}] \leq \underset{95\%\text{ conf.}}{0.05}$

So we set $\varepsilon = 0.01$ and $\frac{1}{4\varepsilon^2 N} \leq 0.05$:

$$4\varepsilon^2 N \geqslant \frac{1}{0.05}$$

$$N \geqslant \frac{5}{\varepsilon^2} = \boxed{50000}$$

- Same calculation works for any desired accuracy & confidence
- Actual sample size required is a lot smaller (using stronger concentration bounds instead of Chebyshev)

Generalization: Estimating $E[X]$ for any r.v. X

E.g. estimate average wealth of US population

Strategy: Sample N people randomly & indep.

Let X_i = wealth of ith person

Output $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i$

estimate of true mean $\mu = E[X_i]$

$$E[\hat{\mu}] = \frac{1}{N} \cdot N\mu = \mu$$

write $Var(X_i) = \sigma^2$

$$Var(\hat{\mu}) = \frac{1}{N^2} \sum_{i=1}^{N} Var(X_i) = \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}$$

$$E[\hat{\mu}] = \mu \qquad \mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{N}$$

Suppose we want accuracy $\pm \varepsilon\mu$, confidence $1-\delta$

<u>Chebyshev</u> : $\Pr\left[|\hat{\mu} - \mu| \geq \varepsilon\mu\right] \leq \frac{\mathrm{Var}(\hat{\mu})}{\varepsilon^2\mu^2} = \frac{\sigma^2}{N\varepsilon^2\mu^2}$

So to ensure confidence $1-\delta$ we need

$$\frac{\sigma^2}{N\varepsilon^2\mu^2} \leq \delta \quad \Longrightarrow \quad \boxed{N \geq \underbrace{\frac{\sigma^2}{\mu^2}}_{\text{problem-specific cost}} \times \underbrace{\frac{1}{\varepsilon^2\delta}}_{\text{inherent cost of estimation}}}$$

$$E[\hat{\mu}] = \mu \qquad \mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{N}$$

Suppose we want accuracy $\pm \varepsilon\mu$, confidence $1-\delta$

<u>Chebyshev</u> : $\Pr[|\hat{\mu} - \mu| \geq \varepsilon\mu] \leq \frac{\mathrm{Var}(\hat{\mu})}{\varepsilon^2\mu^2} = \frac{\sigma^2}{N\varepsilon^2\mu^2}$

So to ensure confidence $1-\delta$ we need

$$\frac{\sigma^2}{N\varepsilon^2\mu^2} \leq \delta \implies \boxed{N \geq \underbrace{\frac{\sigma^2}{\mu^2}}_{\text{problem-specific cost}} \times \underbrace{\frac{1}{\varepsilon^2\delta}}_{\text{inherent cost of estimation}}}$$

What values should we plug in for σ, μ?

We can use any <u>upper bound</u> on σ and any <u>lower bound</u> on μ

E.g. for US wealth, could use $\mu \geq 50{,}000$

But σ is a problem! Elon Musk (\$190B) $\Rightarrow \sigma^2 \gtrsim \frac{(190 \times 10^9)^2}{325 \times 10^6} \approx 10^{14}$!!!

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

However, suppose we know that nobody's wealth is more than k times the average wealth μ

Then $X_i \leq k\mu$ and so

$$\mathrm{Var}(X_i) = E\left[(X_i - \mu)^2\right] \leq (k-1)^2 \mu^2$$

And then $\frac{\sigma^2}{\mu^2} \leq (k-1)^2$, so it's enough to take

$$\boxed{N \geq (k-1)^2 \times \frac{1}{\varepsilon^2 \delta}}$$

E.g. for $k=3$, $\varepsilon = 0.1$, $\delta = 0.05$ $\longrightarrow$ $N = 8000$ suffices

Law of Large Numbers

independent, identically distributed

Theorem: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common expectation $\mu = E[X_i]$.

Then $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ satisfies

$$\Pr\left[\left|\frac{1}{N}S_N - \mu\right| \geq \varepsilon\right] \longrightarrow 0 \quad \text{as } N \to \infty$$

for any $\varepsilon > 0$.

English: We can achieve any desired accuracy $\varepsilon > 0$ and any desired confidence $1 - \delta < 1$ by taking the sample size N large enough

Law of Large Numbers

independent, identically distibuted

Theorem: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common expectation $\mu = E[X_i]$.

Then $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ satisfies

$$\Pr\left[\left|\frac{1}{N} S_N - \mu\right| \geq \varepsilon\right] \longrightarrow 0 \quad \text{as } N \to \infty$$

for any $\varepsilon > 0$.

Proof: Let $Y = \frac{1}{N} S_N$. Then $E[Y] = \frac{1}{N}\sum_{i=1}^{N} E[X_i] = \mu$

$$Var(Y) = \frac{1}{N^2}\sum_{i=1}^{N} Var(X_i) = \frac{\sigma^2}{N} \qquad \text{where } \sigma^2 \stackrel{\text{def}}{=} Var(X_i)$$

Chebyshev: $\Pr[|Y - \mu| \geq \varepsilon] \leq \frac{Var(Y)}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2} \xrightarrow[N \to \infty]{} 0$

$\square$