CSTO - Spring 2024
Lecture 21 - April 4
Review of Previous Lecture

- **Variance**: For a random variable with $E[X]=\mu$
  \[
  \text{Var}(X) = E[(X-\mu)^2] = E[X^2] - \mu^2
  \]

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

  Measures “spread” of the distribution

- To compute $E[X^2]$
  \[
  E[X^2] = \sum_a a^2 \times \Pr[X=a]
  \]
Review (cont.)

- \( X \sim \text{Bin}(n,p) \):
  \[ E[X] = np \quad \text{Var}[X] = np(1-p) \]

- \( X \sim \text{Geom}(p) \):
  \[ E[X] = \frac{1}{p} \quad \text{Var}[X] = \frac{1-p}{p^2} \]

- \( X \sim \text{Poisson}(\lambda) \):
  \[ E[X] = \lambda \quad \text{Var}[X] = \lambda \]

- For any r.v. \( X \) and constant \( c \):
  \[ \text{Var}(cX) = c^2 \text{Var}(X) \]

- If \( X, Y \) are independent, then
  \[ \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \]
Review (cont.)

- For any two r.v.'s \( X, Y \):
  \[
  \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
  \]

- **Covariance**
  \[
  \text{Cov}(X, Y) = E[XY] - E(X)E(Y)
  \]
  \[
  = E[(X-\mu_X)(Y-\mu_Y)]
  \]

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

- \( \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \) (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 $c$:

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

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

By definition of $E[X]$:

$$E[X] = \sum_{a} a \cdot \Pr(X = a)$$

$$\geq \sum_{a \geq c} a \cdot \Pr(X = a)$$

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

Hence $\Pr(X \geq c) \leq \frac{1}{c} E[X]$ \qed
**Example:** \( X \sim \text{Binomial}(n, \frac{1}{2}) \)

Recall: \( E[X] = np = n/2 \)

**Markov:** \( \Pr [X \geq c] \leq \frac{E[X]}{c} \)

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

**Note:** This upper bound is correct but far from the best bound we can get — see later!
Q: Suppose we also know $\text{Var}(X)$ - does this help?

A: Yes! Recall that $\text{Var}(X)$ measures expected (squared) distance of $X$ from $E[X]$. If $\text{Var}(X)$ 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 $c$:

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

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 \( c \):

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

Proof: Define the r.v. \( Y = (X - \mathbb{E}[X])^2 \).

Note that \( Y \) is non-negative so we can apply Markov’s inequality to it:

\[
\Pr \left[ Y \geq c^2 \right] \leq \frac{\mathbb{E}[Y]}{c^2}
\]

i.e.

\[
\Pr \left[ (X - \mathbb{E}[X])^2 \geq c^2 \right] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{c^2}
\]

i.e.

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

\( \square \)
Example: \( X \sim \text{Binomial} (n, \frac{1}{2}) \)

Recall: \( E[X] = np = n/2 \quad \text{Var}(X) = np(1-p) = n/4 \)

Chebyshev: \( \Pr \left[ |X - E[X]| \geq C \right] \leq \frac{\text{Var}(X)}{C^2} \)

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

\[ \leq \frac{\text{Var}(X)}{(\frac{n}{4})^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n} \]

This is much better 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. $X$:

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

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

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

Example: For any r.v. $X$, the probability of being more than 2 s.d.'s from mean is $\leq \frac{1}{4}$
Example: \( X \sim \text{Poisson}(\lambda) \)

Recall \( \mathbb{E}[X] = \lambda \), \( \operatorname{Var}(X) = \lambda \), \( \sigma(X) = \sqrt{\lambda} \)

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

E.g. \( \lambda = 100 \rightarrow \Pr[|X - 100| > 20] \leq \frac{1}{4} \)
Application: Statistical Estimation

Goal: Estimate the proportion of smokers in the population within ±1% with confidence > 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: How large does $N$ have to be to ensure accuracy ±1% with confidence 95%?

Note: Assume for simplicity we choose people with replacement so that samples are all independent
Define r.v. $S_N$ by $S_N = X_1 + X_2 + \ldots + X_N$ where $X_i = \begin{cases} 1 & \text{if person } i \text{ says "yes"} \\ 0 & \text{otherwise} \end{cases}$

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

Define r.v. $S_N$ by

$$S_N = X_1 + X_2 + \ldots + X_N$$

where

$$X_i = \begin{cases} 
1 & \text{if person } i \text{ says "yes"} \\
0 & \text{otherwise}
\end{cases}$$

Output

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

Our estimate of the true unknown proportion $p$

Expected Value

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

“unbiased estimator”

Variance

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

decreases with $N$!
The Math: Define r.v. $S_N$ by

$$S_N = X_1 + X_2 + \ldots + X_N$$

where $X_i = \begin{cases} 1 & \text{if person } i \text{ says "Yes"} \\ 0 & \text{otherwise} \end{cases}$

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

\[
E[\hat{p}] = \frac{1}{N} \sum_{i=1}^{N} E[X_i] = \frac{1}{N} \times N \cdot p = p
\]

\[
\text{Var}(\hat{p}) = \frac{1}{N^2} \sum_{i=1}^{N} \text{Var}(X_i) = \frac{1}{N^2} \times N \cdot p(1-p) = \frac{p(1-p)}{N}
\]

Chebyshev: $P \left[ |\hat{p} - p| \geq \varepsilon \right] \leq \frac{\text{Var}(\hat{p})}{\varepsilon^2} = \frac{p(1-p)}{\varepsilon^2 N} \leq \frac{1}{4\varepsilon^2 N}$
Chebyshev: $P[|\hat{p} - p| \geq \varepsilon] \leq \frac{1}{4\varepsilon^2 N}$

Recall: we want $P[|\hat{p} - p| \geq 0.01] \leq 0.05$

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

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

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

- 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 $i$th person

Output $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i$ \hspace{1cm} \text{estimate of true mean $\mu = E[X_i]$}

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

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

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

\[ \text{Chebyshev: } \Pr[|\hat{\mu} - \mu| \geq \varepsilon \mu] \leq \frac{\text{Var}(\hat{\mu})}{\varepsilon^2 \mu^2} = \frac{\sigma^2}{N \varepsilon^2 \mu^2} \]

So to ensure confidence \( 1-\delta \) we need

\[ \frac{\sigma^2}{N \varepsilon^2 \mu^2} \leq \delta \quad \Rightarrow \quad N \geq \frac{\sigma^2}{\mu^2 \varepsilon^2 \delta} \]

\[ N \geq \frac{\sigma^2}{\mu^2 \varepsilon^2 \delta} \]
\[ E[\hat{\mu}] = \mu \quad \text{Var}(\hat{\mu}) = \frac{\sigma^2}{N} \]

Suppose we want accuracy \( \pm \epsilon \mu \), confidence \( 1 - \delta \)

**Chebyshev:** \( \Pr[|\hat{\mu} - \mu| \geq \epsilon \mu] \leq \frac{\text{Var}(\hat{\mu})}{\epsilon^2 \mu^2} = \frac{\sigma^2}{N \epsilon^2 \mu^2} \)

So to ensure confidence \( 1 - \delta \) we need

\[ \frac{\sigma^2}{N \epsilon^2 \mu^2} < \delta \quad \Rightarrow \quad N > \frac{\sigma^2}{\epsilon^2 \mu^2} \times \frac{1}{\delta} \]

What values should we plug in for \( \sigma, \mu \)?

We can use any upper bound on \( \sigma \) and any lower bound on \( \mu \)

E.g. for US wealth, could use \( \mu \gtrsim 50,000 \)

But \( \sigma \) 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 \( \mu \). Then \( X_i \leq k \mu \) and so

\[ \text{Var}(X_i) = E[(X_i - \mu)^2] \leq (k-1)^2 \mu^2 \]

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

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

E.g. for \( k = 3 \), \( \varepsilon = 0.1 \), \( \delta = 0.05 \) \( \rightarrow N = 8000 \) suffices
Law of Large Numbers

Theorem: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common expectation $M = E[X_i]$. Then $S_n = \frac{1}{N} \sum_{i=1}^{N} X_i$ satisfies

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

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

Theorem: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common expectation $M = E[X_i]$. Then $S_N = \frac{1}{N} \sum_{i=1}^{N} X_i$ satisfies

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

for any $\varepsilon > 0$.

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

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

where $\sigma^2 = \text{Var}(X_i)$

Chebyshev: $\Pr \left[ \left| Y - M \right| \geq \varepsilon \right] \leq \frac{\text{Var}(Y)}{\varepsilon^2} = \frac{\sigma^2}{N \varepsilon^2} \xrightarrow{N \rightarrow \infty} 0$