# CMPT 210: Probability and Computing

Lecture 21

Sharan Vaswani

March 28, 2024

# Markov's Theorem

Markov's theorem formalizes the intuition on the last slide of the previous class, and can be stated as follows.

**Markov's Theorem**: If $X$ is a non-negative random variable, then for all $x > 0$,

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

# Markov's Theorem

Markov's theorem formalizes the intuition on the last slide of the previous class, and can be stated as follows.

**Markov's Theorem**: If $X$ is a non-negative random variable, then for all $x > 0$,

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

*Proof*: Define $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of $X$, $x\mathcal{I}\{X \geq x\} \leq X$.

# Markov's Theorem

Markov's theorem formalizes the intuition on the last slide of the previous class, and can be stated as follows.

**Markov's Theorem**: If $X$ is a non-negative random variable, then for all $x > 0$,

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

*Proof*: Define $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of $X$, $x\,\mathcal{I}\{X \geq x\} \leq X$.

$$\mathbb{E}[x\,\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\,\mathbb{E}[\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x \Pr[X \geq x] \leq \mathbb{E}[X]$$
$$\implies \Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

# Markov's Theorem

Markov's theorem formalizes the intuition on the last slide of the previous class, and can be stated as follows.

**Markov's Theorem**: If $X$ is a non-negative random variable, then for all $x > 0$,

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

*Proof*: Define $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of $X$, $x\mathcal{I}\{X \geq x\} \leq X$.

$$\mathbb{E}[x\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\,\mathbb{E}[\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\Pr[X \geq x] \leq \mathbb{E}[X]$$
$$\implies \Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

Since the above theorem holds for all $x > 0$, we can set $x = c\mathbb{E}[X]$ for $c \geq 1$. In this case, $\Pr[X \geq c\mathbb{E}[X]] \leq \frac{1}{c}$. Hence, the probability that $X$ is "far" from the mean in terms of the multiplicative factor $c$ is upper-bounded by $\frac{1}{c}$.

# Markov's Theorem – Example

**Q**: Suppose there is a dinner party where $n$ people check in their coats. The coats are mixed up during dinner, so that afterward each person receives a random coat. In particular, a person gets their own coat with probability $\frac{1}{n}$.

# Markov's Theorem – Example

**Q**: Suppose there is a dinner party where $n$ people check in their coats. The coats are mixed up during dinner, so that afterward each person receives a random coat. In particular, a person gets their own coat with probability $\frac{1}{n}$.

Recall that if $G$ is the r.v. corresponding to the number of people that receive their own coat, then we used the linearity of expectation to derive that $\mathbb{E}[G] = 1$. Using Markov's Theorem,

$$\Pr[G \geq x] \leq \frac{\mathbb{E}[G]}{x} = \frac{1}{x}.$$

Hence, we can bound the probability that $x$ people receive their own coat. For example, there is no better than 20% chance that more than 5 people get their own coat.

# Markov's Theorem – Example

Q: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

# Markov's Theorem – Example

Q: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

# Markov's Theorem – Example

Q: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

**Q**: If we are provided additional information that $X$ can not take values less than 100 and $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

# Markov's Theorem – Example

Q: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

**Q**: If we are provided additional information that $X$ can not take values less than 100 and $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

Define $Y := X - 100$. $\mathbb{E}[Y] = \mathbb{E}[X] - 100 = 50$ and $Y$ is non-negative.

$$\Pr[X \geq 200] = \Pr[Y + 100 \geq 200] = \Pr[Y \geq 100] \leq \frac{\mathbb{E}[Y]}{100} = \frac{50}{100} = \frac{1}{2}$$

# Markov's Theorem – Example

Q: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

**Q**: If we are provided additional information that $X$ can not take values less than 100 and $\mathbb{E}[X] = 150$, compute the probability that $X$ is at least 200.

Define $Y := X - 100$. $\mathbb{E}[Y] = \mathbb{E}[X] - 100 = 50$ and $Y$ is non-negative.

$$\Pr[X \geq 200] = \Pr[Y + 100 \geq 200] = \Pr[Y \geq 100] \leq \frac{\mathbb{E}[Y]}{100} = \frac{50}{100} = \frac{1}{2}$$

Hence, if we have additional information (in the form of a lower-bound that a r.v. can not be smaller than some constant $b > 0$), we can use Markov's Theorem on the shifted r.v. ($Y$ in our example) and obtain a tighter bound on the probability of deviation.

# Chebyshev's Theorem

**Chebyshev's Theorem**: For a r.v. $X$ and any constant $y > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathrm{Var}[X]}{y^2}.$$

# Chebyshev's Theorem

**Chebyshev's Theorem**: For a r.v. $X$ and any constant $y > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathrm{Var}[X]}{y^2}.$$

*Proof*: Use Markov's Theorem with some cleverly chosen function of $X$. Formally, for some function $f$ such that $Y := f(X)$ is non-negative. Using Markov's Theorem for $Y$,

$$\Pr[f(X) \geq x] \leq \frac{\mathbb{E}[f(X)]}{x}$$

# Chebyshev's Theorem

**Chebyshev's Theorem**: For a r.v. $X$ and any constant $y > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathrm{Var}[X]}{y^2}.$$

*Proof*: Use Markov's Theorem with some cleverly chosen function of $X$. Formally, for some function $f$ such that $Y := f(X)$ is non-negative. Using Markov's Theorem for $Y$,

$$\Pr[f(X) \geq x] \leq \frac{\mathbb{E}[f(X)]}{x}$$

Choosing $f(X) = |X - \mathbb{E}[X]|^2$ and $x = y^2$ implies that $f(X)$ is non-negative and $x > 0$. Using Markov's Theorem,

$$\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2}$$

# Chebyshev's Theorem

**Chebyshev's Theorem**: For a r.v. $X$ and any constant $y > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathrm{Var}[X]}{y^2}.$$

*Proof*: Use Markov's Theorem with some cleverly chosen function of $X$. Formally, for some function $f$ such that $Y := f(X)$ is non-negative. Using Markov's Theorem for $Y$,

$$\Pr[f(X) \geq x] \leq \frac{\mathbb{E}[f(X)]}{x}$$

Choosing $f(X) = |X - \mathbb{E}[X]|^2$ and $x = y^2$ implies that $f(X)$ is non-negative and $x > 0$. Using Markov's Theorem,

$$\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2}$$

Note that $\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] = \Pr[|X - \mathbb{E}[X]| \geq y]$, and hence,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2} = \frac{\mathrm{Var}[X]}{y^2}$$

# Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable $X$ is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of $x$.

# Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable $X$ is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of $x$.

If we set $x = c\sigma_X$ where $\sigma_X$ is the standard deviation of $X$, then by Chebyshev's Theorem,

$$\Pr[(X \geq \mathbb{E}[X] + c\,\sigma_X) \cup (X \leq \mathbb{E}[X] - c\,\sigma_X)] = \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \leq \frac{\mathrm{Var}[X]}{c^2\sigma_X^2} = \frac{1}{c^2}$$

# Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable $X$ is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of $x$.

If we set $x = c\sigma_X$ where $\sigma_X$ is the standard deviation of $X$, then by Chebyshev's Theorem,

$$\Pr[(X \geq \mathbb{E}[X] + c\,\sigma_X) \cup (X \leq \mathbb{E}[X] - c\,\sigma_X)] = \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \leq \frac{\mathrm{Var}[X]}{c^2\sigma_X^2} = \frac{1}{c^2}$$

$$\begin{aligned} &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = \Pr[|X - \mathbb{E}[X]| \leq c\sigma_X] \\ \implies &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = 1 - \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \geq 1 - \frac{1}{c^2}. \end{aligned}$$

# Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable $X$ is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of $x$.

If we set $x = c\sigma_X$ where $\sigma_X$ is the standard deviation of $X$, then by Chebyshev's Theorem,

$$\Pr[(X \geq \mathbb{E}[X] + c\,\sigma_X) \cup (X \leq \mathbb{E}[X] - c\,\sigma_X)] = \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \leq \frac{\mathrm{Var}[X]}{c^2\sigma_X^2} = \frac{1}{c^2}$$

$$\begin{aligned} &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = \Pr[|X - \mathbb{E}[X]| \leq c\sigma_X] \\ \implies\ &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = 1 - \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \geq 1 - \frac{1}{c^2}. \end{aligned}$$

Chebyshev's Theorem is used to bound the probability that $X$ is "concentrated" near its mean.

# Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable $X$ is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of $x$.

If we set $x = c\sigma_X$ where $\sigma_X$ is the standard deviation of $X$, then by Chebyshev's Theorem,

$$\Pr[(X \geq \mathbb{E}[X] + c\,\sigma_X) \cup (X \leq \mathbb{E}[X] - c\,\sigma_X)] = \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \leq \frac{\mathrm{Var}[X]}{c^2\sigma_X^2} = \frac{1}{c^2}$$

$$\begin{aligned} &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = \Pr[|X - \mathbb{E}[X]| \leq c\sigma_X] \\ \implies &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = 1 - \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \geq 1 - \frac{1}{c^2}. \end{aligned}$$

Chebyshev's Theorem is used to bound the probability that $X$ is "concentrated" near its mean.

Unlike Markov's Theorem, Chebyshev's Theorem does not require the r.v. to be non-negative, but requires knowledge of the variance.

# Chebyshev's Theorem - Example

**Q**: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 100$ and $\sigma_X = 15$, compute the probability that $X$ is at least 300.

# Chebyshev's Theorem - Example

**Q**: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 100$ and $\sigma_X = 15$, compute the probability that $X$ is at least 300.

If we use Markov's Theorem, $\Pr[X \geq 300] \leq \frac{\mathbb{E}[X]}{300} = \frac{1}{3}$.

# Chebyshev's Theorem - Example

**Q**: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 100$ and $\sigma_X = 15$, compute the probability that $X$ is at least 300.

If we use Markov's Theorem, $\Pr[X \geq 300] \leq \frac{\mathbb{E}[X]}{300} = \frac{1}{3}$.

Note that $\Pr[|X - 100| \geq 200] = \Pr[X \leq -100 \cup X \geq 300] = \Pr[X \geq 300]$. Using Chebyshev's Theorem,

$$\Pr[X \geq 300] = \Pr[|X - 100| \geq 200] \leq \frac{\mathrm{Var}[X]}{(200)^2} = \frac{15^2}{200^2} \approx \frac{1}{178}.$$

# Chebyshev's Theorem - Example

**Q**: If $X$ is a non-negative r.v. such that $\mathbb{E}[X] = 100$ and $\sigma_X = 15$, compute the probability that $X$ is at least 300.

If we use Markov's Theorem, $\Pr[X \geq 300] \leq \frac{\mathbb{E}[X]}{300} = \frac{1}{3}$.

Note that $\Pr[|X - 100| \geq 200] = \Pr[X \leq -100 \cup X \geq 300] = \Pr[X \geq 300]$. Using Chebyshev's Theorem,

$$\Pr[X \geq 300] = \Pr[|X - 100| \geq 200] \leq \frac{\mathrm{Var}[X]}{(200)^2} = \frac{15^2}{200^2} \approx \frac{1}{178}.$$

Hence, by exploiting the knowledge of the variance and using Chebyshev's inequality, we can obtain a tighter bound.

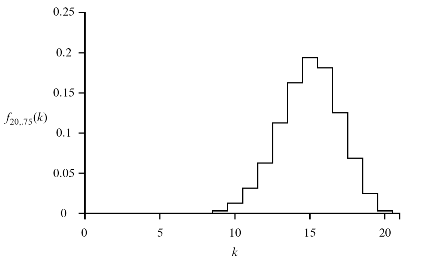
# Chebyshev's Theorem - Example

**Q**: Consider a r.v. $X \sim \mathrm{Bin}(20, 0.75)$. Plot the $\mathrm{PDF}_X$, compute its mean and standard deviation and bound $\Pr[10 < X < 20]$.

# Chebyshev's Theorem - Example

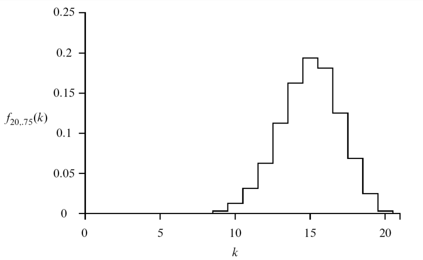
**Q**: Consider a r.v. $X \sim \text{Bin}(20, 0.75)$. Plot the $\text{PDF}_X$, compute its mean and standard deviation and bound $\Pr[10 < X < 20]$.

Range$(X) = \{0, 1, \ldots, 20\}$ and for $k \in$ Range$(X)$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



# Chebyshev's Theorem - Example

**Q**: Consider a r.v. $X \sim \text{Bin}(20, 0.75)$. Plot the $\text{PDF}_X$, compute its mean and standard deviation and bound $\Pr[10 < X < 20]$.



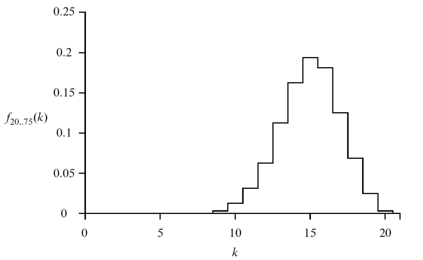

$\text{Range}(X) = \{0, 1, \ldots, 20\}$ and for $k \in \text{Range}(X)$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

$\mathbb{E}[X] = np = (20)(0.75) = 15$

$\text{Var}[X] = np(1-p) = 20(0.75)(0.25) = 3.75$ and hence $\sigma_X = \sqrt{3.75} \approx 1.94$.

# Chebyshev's Theorem - Example

**Q**: Consider a r.v. $X \sim \text{Bin}(20, 0.75)$. Plot the $\text{PDF}_X$, compute its mean and standard deviation and bound $\Pr[10 < X < 20]$.

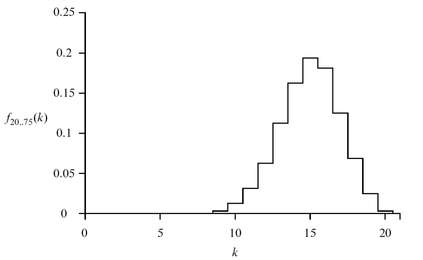



$\text{Range}(X) = \{0, 1, \ldots, 20\}$ and for $k \in \text{Range}(X)$,
$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
$\mathbb{E}[X] = np = (20)(0.75) = 15$
$\text{Var}[X] = np(1-p) = 20(0.75)(0.25) = 3.75$ and hence
$\sigma_X = \sqrt{3.75} \approx 1.94$.

$$\begin{aligned}
\Pr[10 < X < 20] &= 1 - \Pr[X \leq 10 \cup X \geq 20] \\
&= 1 - \Pr[|X - 15| \geq 5] \\
&= 1 - \Pr[|X - \mathbb{E}[X]| \geq 5] \\
&\geq 1 - \frac{\text{Var}[X]}{(5)^2} = 1 - \frac{3.75}{25} = 0.85.
\end{aligned}$$

# Chebyshev's Theorem - Example

**Q**: Consider a r.v. $X \sim \text{Bin}(20, 0.75)$. Plot the $\text{PDF}_X$, compute its mean and standard deviation and bound $\Pr[10 < X < 20]$.




$\text{Range}(X) = \{0, 1, \ldots, 20\}$ and for $k \in \text{Range}(X)$,
$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
$\mathbb{E}[X] = np = (20)(0.75) = 15$
$\text{Var}[X] = np(1-p) = 20(0.75)(0.25) = 3.75$ and hence
$\sigma_X = \sqrt{3.75} \approx 1.94$.

$$
\begin{aligned}
\Pr[10 < X < 20] &= 1 - \Pr[X \le 10 \cup X \ge 20] \\
&= 1 - \Pr[|X - 15| \ge 5] \\
&= 1 - \Pr[|X - \mathbb{E}[X]| \ge 5] \\
&\ge 1 - \frac{\text{Var}[X]}{(5)^2} = 1 - \frac{3.75}{25} = 0.85.
\end{aligned}
$$

Hence, the "probability mass" of $X$ is "concentrated" around its mean.

# Voter Poll

**Q**: Suppose there is an election between two candidates Donald Trump and Joe Biden, and we are hired by candidate Biden's election campaign to estimate his chances of winning the election. In particular, we want to estimate $p$, the fraction of voters favoring Biden before the election. We conduct a voter poll – selecting (typically calling) people uniformly at random (with replacement so that we can choose a person twice) and try to estimate $p$. What is the number of people we should poll to estimate $p$ reasonably accurately and with reasonably high probability?

# Voter Poll

**Q**: Suppose there is an election between two candidates Donald Trump and Joe Biden, and we are hired by candidate Biden's election campaign to estimate his chances of winning the election. In particular, we want to estimate $p$, the fraction of voters favoring Biden before the election. We conduct a voter poll – selecting (typically calling) people uniformly at random (with replacement so that we can choose a person twice) and try to estimate $p$. What is the number of people we should poll to estimate $p$ reasonably accurately and with reasonably high probability?

Define $X_i$ to be the indicator r.v. equal to 1 iff person $i$ that we called favors Biden.

**Assumption (1)**: The $X_i$ r.v's are mutually independent since the people we poll are chosen randomly and we assume that their opinions do not affect each other.

**Assumption (2)**: The people we call are identically distributed i.e. $X_i = 1$ with probability $p$.

# Voter Poll

**Q**: Suppose there is an election between two candidates Donald Trump and Joe Biden, and we are hired by candidate Biden's election campaign to estimate his chances of winning the election. In particular, we want to estimate $p$, the fraction of voters favoring Biden before the election. We conduct a voter poll – selecting (typically calling) people uniformly at random (with replacement so that we can choose a person twice) and try to estimate $p$. What is the number of people we should poll to estimate $p$ reasonably accurately and with reasonably high probability?

Define $X_i$ to be the indicator r.v. equal to 1 iff person $i$ that we called favors Biden.

**Assumption (1)**: The $X_i$ r.v's are mutually independent since the people we poll are chosen randomly and we assume that their opinions do not affect each other.

**Assumption (2)**: The people we call are identically distributed i.e. $X_i = 1$ with probability $p$.

Suppose we poll $n$ people and define $S_n := \sum_{i=1}^{n} X_i$ as the r.v. equal to the total number of people (amongst the ones we polled) that prefer Biden. $\frac{S_n}{n}$ is the *statistical estimate* of $p$.

# Voter Poll

**Q**: Suppose there is an election between two candidates Donald Trump and Joe Biden, and we are hired by candidate Biden's election campaign to estimate his chances of winning the election. In particular, we want to estimate $p$, the fraction of voters favoring Biden before the election. We conduct a voter poll – selecting (typically calling) people uniformly at random (with replacement so that we can choose a person twice) and try to estimate $p$. What is the number of people we should poll to estimate $p$ reasonably accurately and with reasonably high probability?

Define $X_i$ to be the indicator r.v. equal to 1 iff person $i$ that we called favors Biden.

**Assumption (1)**: The $X_i$ r.v's are mutually independent since the people we poll are chosen randomly and we assume that their opinions do not affect each other.

**Assumption (2)**: The people we call are identically distributed i.e. $X_i = 1$ with probability $p$.

Suppose we poll $n$ people and define $S_n := \sum_{i=1}^{n} X_i$ as the r.v. equal to the total number of people (amongst the ones we polled) that prefer Biden. $\frac{S_n}{n}$ is the *statistical estimate* of $p$.

Q: What is the distribution of $S_n$?

# Voter Poll

We want to find for what $n$ is our estimate for $p$ accurate up to an error $\epsilon > 0$ and with probability $1 - \delta$ (for $\delta \in (0, 1)$). Formally, for what $n$ is,

$$\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$$

# Voter Poll

We want to find for what $n$ is our estimate for $p$ accurate up to an error $\epsilon > 0$ and with probability $1 - \delta$ (for $\delta \in (0, 1)$). Formally, for what $n$ is,

$$\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$$

Since $S_n \sim \text{Bin}(n, p)$, $\mathbb{E}[S_n] = np$ and hence, $\mathbb{E}\left[\frac{S_n}{n}\right] = p$, meaning that our estimate is *unbiased* – in expectation, the estimate is equal to $p$. Hence, the above statement is equivalent to,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] \geq 1 - \delta$$

# Voter Poll

We want to find for what $n$ is our estimate for $p$ accurate up to an error $\epsilon > 0$ and with probability $1 - \delta$ (for $\delta \in (0, 1)$). Formally, for what $n$ is,

$$\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$$

Since $S_n \sim \text{Bin}(n, p)$, $\mathbb{E}[S_n] = np$ and hence, $\mathbb{E}\left[\frac{S_n}{n}\right] = p$, meaning that our estimate is *unbiased* – in expectation, the estimate is equal to $p$. Hence, the above statement is equivalent to,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] \geq 1 - \delta$$

Hence, we can use Chebyshev's Theorem for the r.v. $\frac{S_n}{n}$ with $x = \epsilon$ to bound the LHS

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] = 1 - \Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] \geq 1 - \frac{\text{Var}[S_n/n]}{\epsilon^2}.$$

# Voter Poll

We want to find for what $n$ is our estimate for $p$ accurate up to an error $\epsilon > 0$ and with probability $1 - \delta$ (for $\delta \in (0, 1)$). Formally, for what $n$ is,

$$\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$$

Since $S_n \sim \text{Bin}(n, p)$, $\mathbb{E}[S_n] = np$ and hence, $\mathbb{E}\left[\frac{S_n}{n}\right] = p$, meaning that our estimate is *unbiased* – in expectation, the estimate is equal to $p$. Hence, the above statement is equivalent to,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] \geq 1 - \delta$$

Hence, we can use Chebyshev's Theorem for the r.v. $\frac{S_n}{n}$ with $x = \epsilon$ to bound the LHS

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] = 1 - \Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] \geq 1 - \frac{\text{Var}[S_n/n]}{\epsilon^2}.$$

Hence, the problem now is to find $n$ such that,

$$1 - \frac{\text{Var}[S_n/n]}{\epsilon^2} \geq 1 - \delta \implies \frac{\text{Var}[S_n/n]}{\epsilon^2} < \delta$$

# Voter Poll

Let us calculate the $\mathrm{Var}[S_n/n]$.

$$\begin{aligned}
\mathrm{Var}[S_n/n] &= \frac{1}{n^2}\mathrm{Var}[S_n] && \text{(Using the property of variance)} \\
&= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} && \text{(Using the variance of the Binomial distribution)}
\end{aligned}$$

# Voter Poll

Let us calculate the $\mathrm{Var}[S_n/n]$.

$$\begin{aligned}\mathrm{Var}[S_n/n] &= \frac{1}{n^2}\mathrm{Var}[S_n] && \text{(Using the property of variance)}\\ &= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} && \text{(Using the variance of the Binomial distribution)}\end{aligned}$$

Hence, we want to find $n$ s.t.

$$\frac{p\,(1-p)}{n\epsilon^2} < \delta \implies n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$$

# Voter Poll

Let us calculate the $\mathrm{Var}[S_n/n]$.

$$\begin{aligned} \mathrm{Var}[S_n/n] &= \frac{1}{n^2}\mathrm{Var}[S_n] && \text{(Using the property of variance)} \\ &= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} && \text{(Using the variance of the Binomial distribution)} \end{aligned}$$

Hence, we want to find $n$ s.t.

$$\frac{p\,(1-p)}{n\epsilon^2} < \delta \implies n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$$

But we do not know $p$! If $n \geq \max_p \frac{p(1-p)}{\epsilon^2\,\delta}$, then for any $p$, $n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$. So the problem is to compute $\max_p \frac{p(1-p)}{\epsilon^2\,\delta}$. This is a concave function and is maximized at $p = 1/2$.

# Voter Poll

Let us calculate the $\mathrm{Var}[S_n/n]$.

$$\begin{aligned}
\mathrm{Var}[S_n/n] &= \frac{1}{n^2}\mathrm{Var}[S_n] && \text{(Using the property of variance)}\\
&= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} && \text{(Using the variance of the Binomial distribution)}
\end{aligned}$$

Hence, we want to find $n$ s.t.

$$\frac{p\,(1-p)}{n\epsilon^2} < \delta \implies n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$$

But we do not know $p$! If $n \geq \max_p \frac{p(1-p)}{\epsilon^2\,\delta}$, then for any $p$, $n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$. So the problem is to compute $\max_p \frac{p(1-p)}{\epsilon^2\,\delta}$. This is a concave function and is maximized at $p = 1/2$.

Hence, if $n \geq \frac{1}{4\epsilon^2\delta}$, then $\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$ meaning that we have estimated $p$ upto an error $\epsilon$ and this bound is true with high probability equal to $1 - \delta$.

# Voter Poll

Let us calculate the $\text{Var}[S_n/n]$.

$$\begin{aligned}\text{Var}[S_n/n] &= \frac{1}{n^2}\text{Var}[S_n] && \text{(Using the property of variance)}\\ &= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} && \text{(Using the variance of the Binomial distribution)}\end{aligned}$$

Hence, we want to find $n$ s.t.

$$\frac{p\,(1-p)}{n\epsilon^2} < \delta \implies n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$$

But we do not know $p$! If $n \geq \max_p \frac{p(1-p)}{\epsilon^2\,\delta}$, then for any $p$, $n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$. So the problem is to compute $\max_p \frac{p(1-p)}{\epsilon^2\,\delta}$. This is a concave function and is maximized at $p = 1/2$.

Hence, if $n \geq \frac{1}{4\epsilon^2\delta}$, then $\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1-\delta$ meaning that we have estimated $p$ upto an error $\epsilon$ and this bound is true with high probability equal to $1-\delta$.

For example, if $\epsilon = 0.01$ and $\delta = 0.01$ meaning that we want the bound to hold 99% of the time, then, we require $n \geq 250000$.

# Pairwise Independent Sampling

**Claim**: Let $G_1, G_2, \ldots, G_n$ be pairwise independent random variables with the same mean $\mu$ and standard deviation $\sigma$. Define $S_n := \sum_{i=1}^{n} G_i$, then,

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1}{n}\left(\frac{\sigma}{\epsilon}\right)^2.$$

# Pairwise Independent Sampling

**Claim**: Let $G_1, G_2, \ldots, G_n$ be pairwise independent random variables with the same mean $\mu$ and standard deviation $\sigma$. Define $S_n := \sum_{i=1}^{n} G_i$, then,

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1}{n}\left(\frac{\sigma}{\epsilon}\right)^2.$$

*Proof*: Let us compute $\mathbb{E}[S_n/n]$ and $\mathrm{Var}[S_n/n]$.

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} G_i\right] = \sum_{i=1}^{n} \mathbb{E}[G_i] = n\mu \implies \mathbb{E}[S_n/n] = \frac{1}{n}\mathbb{E}[S_n] = \mu$$

(Using linearity of expectation)

$$\mathrm{Var}[S_n] = \mathrm{Var}\left[\sum_{i=1}^{n} G_i\right] = \sum_{i=1}^{n} \mathrm{Var}[G_i] = n\sigma^2$$

(Using linearity of variance for pairwise independent r.v's)

$$\implies \mathrm{Var}[S_n/n] = \frac{1}{n^2}\mathrm{Var}[S_n] = \frac{\sigma^2}{n}$$

# Pairwise Independent Sampling

Using Chebyshev's Theorem,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] = \Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

# Pairwise Independent Sampling

Using Chebyshev's Theorem,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] = \Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, for arbitrary pairwise independent r.v's, if $n$ increases, the probability of deviation from the mean $\mu$ decreases.

# Pairwise Independent Sampling

Using Chebyshev's Theorem,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] = \Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, for arbitrary pairwise independent r.v's, if $n$ increases, the probability of deviation from the mean $\mu$ decreases.

**Weak Law of Large Numbers**: Let $G_1, G_2, \ldots, G_n$ be pairwise independent variables with the same mean $\mu$ and (finite) standard deviation $\sigma$. Define $X_n := \frac{\sum_{i=1}^{n} G_i}{n}$, then for every $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr[|X_n - \mu| \leq \epsilon] = 1.$$

# Pairwise Independent Sampling

Using Chebyshev's Theorem,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] = \Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, for arbitrary pairwise independent r.v's, if $n$ increases, the probability of deviation from the mean $\mu$ decreases.

**Weak Law of Large Numbers**: Let $G_1, G_2, \ldots, G_n$ be pairwise independent variables with the same mean $\mu$ and (finite) standard deviation $\sigma$. Define $X_n := \frac{\sum_{i=1}^{n} G_i}{n}$, then for every $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr[|X_n - \mu| \leq \epsilon] = 1.$$

*Proof*: Follows from the theorem on pairwise independent sampling since $\lim_{n\to\infty} \Pr[|X_n - \mu| \leq \epsilon] = \lim_{n\to\infty} \left[1 - \frac{\sigma^2}{n\epsilon^2}\right] = 1.$

Questions?