Random Variables so far.

Probability Space: $\Omega$, $Pr: \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$. 
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \to [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \to R$. 
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.
Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.
  Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
$X$ and $Y$ independent $\iff$ all associated events are independent.
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \to [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \to R$.
 Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a aPr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$. 
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0,1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow \mathbb{R}$.
Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
$X$ and $Y$ independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \to [0,1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.

Random Variables: $X : \Omega \to R$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_{a} a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$. 
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.

Random Variables: $X : \Omega \rightarrow R$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.

Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$. 

Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.
  Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
$X$ and $Y$ independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.
Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$
  For independent $X$, $Y$, $Var(X + Y) = Var(X) + Var(Y)$.
  Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.
Poisson: $X \sim P(\lambda)$
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.

Random Variables: $X : \Omega \rightarrow \mathbb{R}$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.

Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$. 
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.
  Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
$X$ and $Y$ independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a aPr[X = a] = \sum_{\omega \in \Omega} X(\omega)Pr(\omega)$.

Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$
  For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.
  Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.
Binomial: $X \sim B(n, p)$
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.

Associated event: $Pr[X = a] = \sum_{\omega \in \Omega} 1_{X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_{a} a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.

Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.

Binomial: $X \sim B(n, p)$ $E(X) = np$, $Var(X) = np(1 - p)$.
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.

Random Variables: $X : \Omega \rightarrow R$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.

Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.

Binomial: $X \sim B(n, p)$ $E(X) = np$, $Var(X) = np(1 - p)$

Uniform: $X \sim U\{1, \ldots, n\}$
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \to [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.

Random Variables: $X : \Omega \to R$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_a aPr[X = a] = \sum_{\omega \in \Omega} X(\omega)Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.

Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.

Binomial: $X \sim B(n, p)$ $E(X) = np$, $Var(X) = np(1 - p)$

Uniform: $X \sim U\{1, \ldots, n\}$ $E[X] = \frac{n+1}{2}$, $Var(X) = \frac{n^2-1}{12}$.
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(\omega) = 1$.
Random Variables: $X : \Omega \rightarrow R$.
   Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
$X$ and $Y$ independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a aPr[X = a] = \sum_{\omega \in \Omega} X(\omega)Pr(\omega)$.
Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$
   For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.
   Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.
Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.
Binomial: $X \sim B(n, p)$ $E(X) = np$, $Var(X) = np(1 - p)$
Uniform: $X \sim U\{1, \ldots, n\}$ $E[X] = \frac{n+1}{2}$, $Var(X) = \frac{n^2-1}{12}$.
Geometric: $X \sim G(p)$
Random Variables so far.

Probability Space: $\Omega$, $Pr : \Omega \rightarrow [0, 1]$, $\sum_{\omega \in \Omega} Pr(w) = 1$.

Random Variables: $X : \Omega \rightarrow R$.

Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.

Expectation: $E[X] = \sum_{a} a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.


Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$

For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.

Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.

Binomial: $X \sim B(n, p)$ $E(X) = np$, $Var(X) = np(1 - p)$

Uniform: $X \sim U\{1, \ldots, n\}$ $E[X] = \frac{n+1}{2}$, $Var(X) = \frac{n^2-1}{12}$.

Geometric: $X \sim G(p)$ $E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$.
**Definition** The covariance of $X$ and $Y$ is

$$cov(X, Y) := E[(X - E[X])(Y - E[Y])].$$
Definition The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Definition The correlation of $X$, $Y$, $\text{Cor}(X, Y)$ is

$$\text{corr}(X, Y) : \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$
**Definition** The covariance of $X$ and $Y$ is

$$cov(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Definition** The correlation of $X$, $Y$, $Cor(X, Y)$ is

$$corr(X, Y) : \frac{cov(X, Y)}{\sigma(X)\sigma(Y)}.$$

Note: $|corr(X, Y)| \leq 1$. 

\[corr(X, X) = \frac{cov(X, X)}{\sigma(X)\sigma(X)} = 1\]

\[corr(X, Y) = \frac{cov(X, X)}{\sigma(X)\sigma(Y)} = \frac{\sigma(X)\sigma(Y)}{\sigma(X)\sigma(Y)} = 1\]

\[corr(X, X/2) = \frac{cov(X, X/2)}{\sigma(X)\sigma(X/2)} = \frac{\sigma(X)\sigma(X/2)}{\sigma(X)\sigma(X/2)} = 1\]

\[corr(X, 5X) = \frac{cov(X, 5X)}{\sigma(X)\sigma(5X)} = \frac{5\sigma(X)\sigma(X)}{\sigma(X)\sigma(5X)} = \frac{5}{\sqrt{2}} = 2\]

\[corr(X, X+Y) = \frac{cov(X, X+Y)}{\sigma(X)\sigma(X+Y)} = \frac{\sigma(X)\sigma(X+Y)}{\sigma(X)\sigma(X+Y)} = \frac{\sqrt{2}}{\sigma(X)} = \frac{1}{\sqrt{2}}\]
\textbf{Definition} The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

\textbf{Definition} The correlation of $X$, $Y$, $\text{Cor}(X, Y)$ is

$$\text{corr}(X, Y) : \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$ 

Note: $|\text{corr}(X, Y)| \leq 1$. 

$\text{corr}(X, X) = ?$
Definition The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Definition The correlation of $X$, $Y$, $\text{Cor}(X, Y)$ is

$$\text{corr}(X, Y) : \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$  

Note: $|\text{corr}(X, Y)| \leq 1.$  

$\text{corr}(X, X)$? 1  
$\text{corr}(X, -X)$?
**Definition** The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Definition** The correlation of $X, Y$, $\text{Cor}(X, Y)$ is

$$\text{corr}(X, Y) : = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Note: $|\text{corr}(X, Y)| \leq 1$.

$\text{corr}(X, X)$? 1
$\text{corr}(X, -X)$? -1
$\text{corr}(X, X/2)$?
**Definition** The covariance of $X$ and $Y$ is

$$cov(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Definition** The correlation of $X$, $Y$, $Cor(X, Y)$ is

$$corr(X, Y) : \frac{cov(X, Y)}{\sigma(X)\sigma(Y)}.$$ 

Note: $|corr(X, Y)| \leq 1$.

$corr(X, X)\ ? 1$

$corr(X, -X)\ ? -1$

$corr(X, X/2)\ ? 1$

$corr(X, 5X)\ ?$
Definition The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Definition The correlation of $X$, $Y$, $\text{Cor}(X, Y)$ is

$$\text{corr}(X, Y) : \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$ 

Note: $|\text{corr}(X, Y)| \leq 1$. 

$\text{corr}(X, X) ? 1$
$\text{corr}(X, -X) ? -1$
$\text{corr}(X, X/2) ? 1$
$\text{corr}(X, 5X) ? 1$
$\text{corr}(X, X + Y)$ with $\text{var}(X) = \text{Var}(Y)$, and $X$, $Y$ independent?
**Definition** The covariance of $X$ and $Y$ is

\[ \text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])]. \]

**Definition** The correlation of $X$, $Y$, $Cor(X, Y)$ is

\[ \text{corr}(X, Y) : \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}. \]

Note: $|\text{corr}(X, Y)| \leq 1$.

\[
\begin{align*}
\text{corr}(X, X) & \geq 1 \\
\text{corr}(X, -X) & \leq -1 \\
\text{corr}(X, X/2) & \geq 1 \\
\text{corr}(X, 5X) & \leq 1 \\
\text{corr}(X, X + Y) & \text{ with } \text{var}(X) = \text{Var}(Y), \text{ and } X, Y \text{ independent? } \frac{1}{\sqrt{2}}
\end{align*}
\]
**Definition** The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Definition** The correlation of $X, Y$, $\text{Cor}(X, Y)$ is

$$\text{corr}(X, Y) : = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Note: $|\text{corr}(X, Y)| \leq 1$.

- $\text{corr}(X, X) = 1$
- $\text{corr}(X, -X) = -1$
- $\text{corr}(X, X/2) = 1$
- $\text{corr}(X, 5X) = 1$
- $\text{corr}(X, X + Y)$ with $\text{var}(X) = \text{Var}(Y)$, and $X, Y$ independent? $\frac{1}{\sqrt{2}}$

$$\text{cov}(X, X + Y) = E[(X - E[X])(X - E[X] + Y - E[Y])] = \text{Var}(X) + \text{cov}(X, Y) = \text{Var}(X).$$
Definition The covariance of $X$ and $Y$ is
\[ \text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])]. \]

Definition The correlation of $X$, $Y$, $\text{Cor}(X, Y)$ is
\[ \text{corr}(X, Y) := \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}. \]

Note: $|\text{corr}(X, Y)| \leq 1$.

$\text{corr}(X, X) = 1$
$\text{corr}(X, -X) = -1$
$\text{corr}(X, X/2) = 1$
$\text{corr}(X, 5X) = 1$
$\text{corr}(X, X + Y)$ with $\text{var}(X) = \text{Var}(Y)$, and $X$, $Y$ independent? $\frac{1}{\sqrt{2}}$

\[ \text{cov}(X, X + Y) = E[(X - E[X])(X - E[X] + Y - E[Y])] = \text{Var}(X) + \text{cov}(X, Y) = \text{Var}(X). \]

\[ \text{corr}(X, X + Y) = \frac{\text{var}X}{\sigma(X)\sigma(X+Y)} = \frac{\text{var}X}{\sigma(X)\sqrt{2}\sigma(X)} = \frac{1}{\sqrt{2}}. \]
**Definition** The covariance of $X$ and $Y$ is

\[
cov(X, Y) := E[(X - E[X])(Y - E[Y])].
\]

**Definition** The correlation of $X$, $Y$, $Cor(X, Y)$ is

\[
corr(X, Y) : \frac{cov(X, Y)}{\sigma(X)\sigma(Y)}.
\]

Note: $|corr(X, Y)| \leq 1$.

\[
corr(X, X) = 1
\]
\[
corr(X, -X) = -1
\]
\[
corr(X, X/2) = 1
\]
\[
corr(X, 5X) = 1
\]
\[
corr(X, X + Y) \text{ with } var(X) = Var(Y), \text{ and } X, Y \text{ independent? } \frac{1}{\sqrt{2}}
\]

\[
cov(X, X + Y) = E[(X - E[X])(X - E[X] + Y - E[Y])] = Var(X) + cov(X, Y) = Var(X).
\]
\[
corr(X, X + Y) = \frac{varX}{\sigma(X)\sigma(X+Y)} = \frac{varX}{\sigma(X)\sqrt{2}\sigma(X)} = \frac{1}{\sqrt{2}}
\]

$r^2 = corr(X, Y)^2$ is fraction of variance of $Y$ explained by $X$. 
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.

When $\text{cov}(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller. $X$ and $Y$ are said to be negatively correlated.

When $\text{cov}(X, Y) = 0$, we say that $X$ and $Y$ are uncorrelated.

Four equally likely pairs of values

$\text{cov}(X, Y) = 1/2$  $\text{cov}(X, Y) = -1/2$  $\text{cov}(X, Y) = 0$
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$. 

Four equally likely pairs of values

$\text{cov}(X, Y) = 1/2$  $\text{cov}(X, Y) = -1/2$  $\text{cov}(X, Y) = 0$
Examples of Covariance

Note that \( E[X] = 0 \) and \( E[Y] = 0 \) in these examples. Then \( \text{cov}(X, Y) = E[XY] \).

When \( \text{cov}(X, Y) > 0 \), the RVs \( X \) and \( Y \) tend to be large or small together.
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.

When $\text{cov}(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller.
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.

When $\text{cov}(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller. $X$ and $Y$ are said to be negatively correlated.
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.

When $\text{cov}(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller. $X$ and $Y$ are said to be negatively correlated.

When $\text{cov}(X, Y) = 0$, we say that $X$ and $Y$ are uncorrelated.

\[\text{cov}(X, Y) = \frac{1}{2} \quad \text{cov}(X, Y) = -\frac{1}{2} \quad \text{cov}(X, Y) = 0\]
Examples of Covariance

\[
E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3
\]

\[
E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8
\]

\[
E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2
\]

\[
E[Y^2] = 1^2 \times 0.2 + 2^2 \times 0.6 + 3^2 \times 0.2 = 4.4
\]

\[
E[XY] = 1 \times 0.1 \times 0.2 + 1 \times 0.25 \times 0.25 + ... + 3 \times 3 \times 0.2 = 4.85
\]

\[
\text{cov}(X,Y) = E[XY] - E[X]E[Y] = 4.85 - 2.3 \times 2 = 0.25
\]

\[
\text{var}(X) = E[X^2] - (E[X])^2 = 5.8 - 2.3^2 = 0.51
\]

\[
\text{var}(Y) = E[Y^2] - (E[Y])^2 = 4.4 - 2^2 = 0.4
\]

\[
\text{corr}(X,Y) \approx 0.55
\]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 4.85 - 2.3 \times 2 = 0.25 \]

\[ \text{var}(X) = E[X^2] - (E[X])^2 = 5.8 - 2.3^2 = 0.51 \]

\[ \text{var}(Y) = E[Y^2] - (E[Y])^2 = 4.4 - 2^2 = 0.4 \]

\[ \text{corr}(X, Y) \approx 0.55 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]

\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]

\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]

\[ E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4 \]

\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4 \]
\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = .25 \]
Examples of Covariance

\[
E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3
\]
\[
E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8
\]
\[
E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2
\]
\[
E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4
\]
\[
E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85
\]
\[
\]
\[
var[X] = E[X^2] - E[X]^2 = .51
\]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4 \]
\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 0.25 \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = 0.51 \]
\[ \text{var}[Y] = E[Y^2] - E[Y]^2 = 0.4 \]
Examples of Covariance

\[
E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3
\]

\[
E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8
\]

\[
E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2
\]

\[
E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4
\]

\[
E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85
\]

\[
\]

\[
var[X] = E[X^2] - E[X]^2 = .51
\]

\[
var[Y] = E[Y^2] - E[Y]^2 = .4
\]

\[
corr(X, Y) \approx 0.55
\]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4 \]
\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = .25 \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = .51 \]
\[ \text{var}[Y] = E[Y^2] - E[Y]^2 = .4 \]
\[ \text{corr}(X, Y) \approx 0.55 \]
Inequalities: An Overview

- **Distribution**
  - $P_n \sim p_n$
  - $\mu$
  - $Pr[X > a]$

- **Markov**
  - $a$
  - $Pr[|X - \mu| > \epsilon]$

- **Chebyshev**
  - $\epsilon \cdot \epsilon$
  - $\mu$
  - $Pr[|X - \mu| > \epsilon]$
Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.

Pafnuty Chebyshev was one of his teachers. Markov was an atheist. In 1912 he protested Leo Tolstoy's excommunication from the Russian Orthodox Church by requesting his own excommunication. The Church complied with his request.
Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.
Andrey Markov

Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.

Pafnuty Chebyshev was one of his teachers.
Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes. Pafnuty Chebyshev was one of his teachers. Markov was an atheist. In 1912 he protested Leo Tolstoy’s excommunication from the Russian Orthodox Church by requesting his own excommunication.
Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.

Pafnuty Chebyshev was one of his teachers.

Markov was an atheist. In 1912 he protested Leo Tolstoy’s excommunication from the Russian Orthodox Church by requesting his own excommunication. The Church complied with his request.
Markov’s inequality
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev.
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

\[
\Pr[ X \geq a ] \leq \frac{\mathbb{E}[ f(X) ]}{f(a)}, \quad \text{for all } a \text{ such that } f(a) > 0.
\]
Markov’s inequality
The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality
Assume $f : \mathbb{R} \to [0, \infty)$ is nondecreasing.
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality
Assume $f : \mathbb{R} \rightarrow [0, \infty)$ is nondecreasing. Then,

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

for all $a$ such that $f(a) > 0$. 

Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

Assume $f : \mathbb{R} \rightarrow [0, \infty)$ is nondecreasing. Then,

$$
Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.
$$

**Proof:**
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

Assume $f : \mathbb{R} \to [0, \infty)$ is nondecreasing. Then,

$$
Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.
$$

**Proof:**

Observe that

$$
1\{X \geq a\} \leq \frac{f(X)}{f(a)}.
$$
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality
Assume $f : \mathbb{R} \to [0, \infty)$ is nondecreasing. Then,

$$Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)},$$

for all $a$ such that $f(a) > 0$.

**Proof:**

Observe that

$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(x)/f(a)$, which holds since $f(\cdot) \geq 0$. 
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

Assume \( f : \mathbb{R} \to [0, \infty) \) is nondecreasing. Then,

\[
Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.
\]

**Proof:**

Observe that

\[
1\{X \geq a\} \leq \frac{f(X)}{f(a)}.
\]

Indeed, if \( X < a \), the inequality reads \( 0 \leq f(x)/f(a) \), which holds since \( f(\cdot) \geq 0 \). Also, if \( X \geq a \), it reads \( 1 \leq f(x)/f(a) \), which holds since \( f(\cdot) \) is nondecreasing.
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

Assume $f : \mathbb{R} \to [0, \infty)$ is nondecreasing. Then,

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

for all $a$ such that $f(a) > 0$.

**Proof:**

Observe that

$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(x)/f(a)$, which holds since $f(\cdot) \geq 0$. Also, if $X \geq a$, it reads $1 \leq f(x)/f(a)$, which holds since $f(\cdot)$ is nondecreasing.

Taking the expectation yields the inequality,
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality
Assume \( f : \mathbb{R} \rightarrow [0, \infty) \) is nondecreasing. Then,

\[
Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.
\]

**Proof:**

Observe that

\[
1 \{X \geq a\} \leq \frac{f(X)}{f(a)}.
\]

Indeed, if \( X < a \), the inequality reads \( 0 \leq f(x)/f(a) \), which holds since \( f(\cdot) \geq 0 \). Also, if \( X \geq a \), it reads \( 1 \leq f(x)/f(a) \), which holds since \( f(\cdot) \) is nondecreasing.

Taking the expectation yields the inequality,

expectation of an indicator is the probability.
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

Assume $f : \mathbb{R} \to [0, \infty)$ is nondecreasing. Then,

$$\Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)},$$

for all $a$ such that $f(a) > 0$.

**Proof:**

Observe that

$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(x)/f(a)$, which holds since $f(\cdot) \geq 0$. Also, if $X \geq a$, it reads $1 \leq f(x)/f(a)$, which holds since $f(\cdot)$ is nondecreasing.

Taking the expectation yields the inequality,

- expectation of an indicator is the probability.
- and expectation is monotone, e.g., weighted sum of points.
Markov’s inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev’s first inequality.

**Theorem** Markov’s Inequality

Assume $f : \mathbb{R} \to [0, \infty)$ is nondecreasing. Then,

$$Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)},$$

for all $a$ such that $f(a) > 0$.

**Proof:**

Observe that

$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(x)/f(a)$, which holds since $f(\cdot) \geq 0$. Also, if $X \geq a$, it reads $1 \leq f(x)/f(a)$, which holds since $f(\cdot)$ is nondecreasing.

Taking the expectation yields the inequality, expectation of an indicator is the probability. and expectation is monotone, e.g., weighted sum of points.
Markov's inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev's first inequality.

**Theorem** Markov's Inequality
Assume \( f : \mathbb{R} \to [0, \infty) \) is nondecreasing. Then,

\[
Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.
\]

**Proof:**

Observe that

\[
1 \{X \geq a\} \leq \frac{f(X)}{f(a)}.
\]

Indeed, if \( X < a \), the inequality reads \( 0 \leq f(x)/f(a) \), which holds since \( f(\cdot) \geq 0 \). Also, if \( X \geq a \), it reads \( 1 \leq f(x)/f(a) \), which holds since \( f(\cdot) \) is nondecreasing.

Taking the expectation yields the inequality,

- expectation of an indicator is the probability.
- and expectation is monotone, e.g., weighted sum of points.

That is,

\[
\sum_v Pr[X = v] 1\{v \geq a\} \leq \sum_v Pr[X = v] \frac{f(v)}{f(a)}.
\]
\[ f(a)1\{X \geq a\} \leq f(x) \Rightarrow 1\{X \geq a\} \leq \frac{f(X)}{f(a)} \]

\[ \Rightarrow Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)} \]
Markov Inequality Example: \( G(p) \)

Let \( X = G(p) \).
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] =$
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] =$
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$ 

Choosing $f(x) = x^2$, we get
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$  

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{2-p}{p^2 a^2}.$$
Markov Inequality Example: $G(p)$

Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$  

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{2-p}{p^2 a^2}.$$
Markov Inequality Example: $P(\lambda)$

Let $X = P(\lambda)$. 
Markov Inequality Example: $P(\lambda)$

Let $X = P(\lambda)$. Recall that $E[X] =$
Markov Inequality Example: $P(\lambda)$

Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] =$
Markov Inequality Example: \( P(\lambda) \)

Let \( X = P(\lambda) \). Recall that \( E[X] = \lambda \) and \( E[X^2] = \lambda + \lambda^2 \).
Markov Inequality Example: \( P(\lambda) \)

Let \( X = P(\lambda) \). Recall that \( E[X] = \lambda \) and \( E[X^2] = \lambda + \lambda^2 \).

Choosing \( f(x) = x \), we get
Markov Inequality Example: $P(\lambda)$

Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.$$
Markov Inequality Example: $P(\lambda)$

Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.$$  

Choosing $f(x) = x^2$, we get
Markov Inequality Example: $P(\lambda)$

Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.$$  

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$
Markov Inequality Example: \( P(\lambda) \)

Let \( X = P(\lambda) \). Recall that \( E[X] = \lambda \) and \( E[X^2] = \lambda + \lambda^2 \).

Choosing \( f(x) = x \), we get

\[
Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.
\]

Choosing \( f(x) = x^2 \), we get

\[
Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.
\]
Chebyshev’s Inequality

This is Pafnuty’s inequality:

\[
\Pr\left[|X - \mathbb{E}[X]| > a\right] \leq \frac{\text{var}[X]}{a^2},
\]

for all \(a > 0\). This result confirms that the variance measures the “deviations from the mean.”
Chebyshev’s Inequality

This is Pafnuty’s inequality:

**Theorem:**

\[
Pr[|X - E[X]| > a] \leq \frac{\text{var}[X]}{a^2}, \text{ for all } a > 0.
\]
Chebyshev’s Inequality

This is Pafnuty’s inequality:

**Theorem:**

\[
Pr[|X - E[X]| > a] \leq \frac{\text{var}[X]}{a^2}, \text{ for all } a > 0.
\]

**Proof:** Let \( Y = |X - E[X]| \) and \( f(y) = y^2 \).
This is Pafnuty’s inequality:

**Theorem:**

\[ \Pr[|X - E[X]| > a] \leq \frac{\text{var}[X]}{a^2}, \text{ for all } a > 0. \]

**Proof:** Let \( Y = |X - E[X]| \) and \( f(y) = y^2 \). Then,

\[ \Pr[Y \geq a] \leq \frac{E[f(Y)]}{f(a)} \]
This is Pafnuty’s inequality:

**Theorem:**

\[ Pr[|X - E[X]| > a] \leq \frac{\text{var}[X]}{a^2}, \text{ for all } a > 0. \]

**Proof:** Let \( Y = |X - E[X]| \) and \( f(y) = y^2 \). Then,

\[ Pr[Y \geq a] \leq \frac{E[f(Y)]}{f(a)} = \frac{\text{var}[X]}{a^2}. \]
Chebyshev’s Inequality

This is Pafnuty’s inequality:

**Theorem:**

\[
Pr[|X - E[X]| > a] \leq \frac{\text{var}[X]}{a^2}, \text{ for all } a > 0.
\]

**Proof:** Let \(Y = |X - E[X]|\) and \(f(y) = y^2\). Then,

\[
Pr[Y \geq a] \leq \frac{E[f(Y)]}{f(a)} = \frac{\text{var}[X]}{a^2}.
\]

This result confirms that the variance measures the “deviations from the mean.”
Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $\text{var}[X] = \lambda$.

Thus, $\Pr[|X - \lambda| \geq n] \leq \frac{\text{var}[X]}{n^2} = \frac{\lambda}{n^2}$. 
Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$.
Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $\text{var}[X] = \lambda$. 
Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. Thus,

$$Pr[|X - \lambda| \geq n] \leq \frac{var[X]}{n^2} = \frac{\lambda}{n^2}.$$
Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. Thus,

$$Pr[|X - \lambda| \geq n] \leq \frac{var[X]}{n^2} = \frac{\lambda}{n^2}.$$
Chebyshev and Poisson (continued)

Let \( X = P(\lambda) \). Then, \( E[X] = \lambda \) and \( var[X] = \lambda \).
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. By Markov’s inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $\text{var}[X] = \lambda$. By Markov's inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$ 

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0$
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $\text{var}[X] = \lambda$. By Markov's inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$ 

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$. 
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. By Markov’s inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$ 

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$.

Hence, for $a > \lambda$, 

Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $\text{var}[X] = \lambda$. By Markov’s inequality,

$$
\Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.
$$

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$.

Hence, for $a > \lambda$, $\Pr[X \geq a] \leq \Pr[|X - \lambda| \geq a - \lambda] \leq \frac{\lambda}{(a - \lambda)^2}$.
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. By Markov’s inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$  

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$.

Hence, for $a > \lambda$, $Pr[X \geq a] \leq Pr[|X - \lambda| \geq a - \lambda] \leq \frac{\lambda}{(a - \lambda)^2}$. 

![Graph showing Markov and Chebyshev inequalities for $X = P(\lambda)$, $\lambda = 10$. The graph illustrates the probability $Pr[X \geq a]$ as a function of $a$, comparing Markov's inequality with Chebyshev's inequality.](image-url)
Here is a classical application of Chebyshev’s inequality.
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality.
How likely is it that the fraction of $H$’s differs from 50%?
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from 50%? Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise.
Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise. Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$
Fraction of \( H \)'s

Here is a classical application of Chebyshev’s inequality.
How likely is it that the fraction of \( H \)'s differs from 50%?
Let \( X_m = 1 \) if the \( m \)-th flip of a fair coin is \( H \) and \( X_m = 0 \) otherwise.
Define
\[
Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.
\]
We want to estimate
\[
Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].
\]
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise. Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$ 

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise. Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$ 

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{\text{var}[Y_n]}{(0.1)^2} = 100\text{var}[Y_n].$$
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise.
Define 

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$ 

We want to estimate 

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev, 

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{\text{var}[Y_n]}{(0.1)^2} = 100\text{var}[Y_n].$$

Now,
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise.

Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$  

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{\text{var}[Y_n]}{(0.1)^2} = 100\text{var}[Y_n].$$

Now,

$$\text{var}[Y_n] = \frac{1}{n^2} (\text{var}[X_1] + \cdots + \text{var}[X_n])$$
Fraction of $H$'s

Here is a classical application of Chebyshev's inequality. How likely is it that the fraction of $H$'s differs from 50%? Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise. Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$ 

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{\text{var}[Y_n]}{(0.1)^2} = 100\text{var}[Y_n].$$

Now,

$$\text{var}[Y_n] = \frac{1}{n^2}(\text{var}[X_1] + \cdots + \text{var}[X_n]) = \frac{1}{n}\text{var}[X_1].$$
Fraction of $H$’s

Here is a classical application of Chebyshev’s inequality. How likely is it that the fraction of $H$’s differs from $50\%$?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise.

Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$  

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{\text{var}[Y_n]}{(0.1)^2} = 100 \text{var}[Y_n].$$

Now,

$$\text{var}[Y_n] = \frac{1}{n^2}(\text{var}[X_1] + \cdots + \text{var}[X_n]) = \frac{1}{n} \text{var}[X_1] \leq \frac{1}{4n}.$$
Fraction of $H$'s

Here is a classical application of Chebyshev’s inequality.

How likely is it that the fraction of $H$'s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise.

Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$ 

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{\text{var}[Y_n]}{(0.1)^2} = 100\text{var}[Y_n].$$

Now,

$$\text{var}[Y_n] = \frac{1}{n^2} (\text{var}[X_1] + \cdots + \text{var}[X_n]) = \frac{1}{n} \text{var}[X_1] \leq \frac{1}{4n}.$$ 

$$\text{Var}(X_i) = p(1 - lp) \leq (.5)(.5) = \frac{1}{4}$$
Fraction of $H$'s

\[ Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1. \]

\[ Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}. \]
Fraction of $H$’s

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$  

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$  

For $n = 1,000$, we find that this probability is less than 2.5%. 

As $n \to \infty$, this probability goes to zero. In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of $H$’s is within $\varepsilon > 0$ of $50$% approaches $1$:  

$$Pr[|Y_n - 0.5| \leq \varepsilon] \to 1.$$ 

This is an example of the Law of Large Numbers. We look at a general case next.
Fraction of $H$'s

\[ Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1. \]

\[ Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}. \]

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.
Fraction of $H$'s

\[ Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1. \]

\[ Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}. \]

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$,
Fraction of $H$’s

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$ 

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$ 

For $n = 1,000$, we find that this probability is less than 2.5%. 

As $n \to \infty$, this probability goes to zero. 

In fact, for any $\varepsilon > 0$, as $n \to \infty$, 


Fraction of $H$’s

\[ Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1. \]

\[ Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}. \]

For \( n = 1,000 \), we find that this probability is less than 2.5%.

As \( n \to \infty \), this probability goes to zero.

In fact, for any \( \varepsilon > 0 \), as \( n \to \infty \), the probability that the fraction of \( Hs \) is within \( \varepsilon > 0 \) of 50% approaches 1:
Fraction of $H$’s

\[ Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1. \]

\[ Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}. \]

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of $H$s is within $\varepsilon > 0$ of 50% approaches 1:

\[ Pr[|Y_n - 0.5| \leq \varepsilon] \to 1. \]
Fraction of $H$’s

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$ 

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \rightarrow \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \rightarrow \infty$, the probability that the fraction of $H$s is within $\varepsilon > 0$ of 50% approaches 1:

$$Pr[|Y_n - 0.5| \leq \varepsilon] \rightarrow 1.$$ 

This is an example of the Law of Large Numbers.
Fraction of $H$'s

\[ Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1. \]

\[ Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}. \]

For $n = 1,000$, we find that this probability is less than $2.5\%$.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of $H$s is within $\varepsilon > 0$ of $50\%$ approaches $1$:

\[ Pr[|Y_n - 0.5| \leq \varepsilon] \to 1. \]

This is an example of the **Law of Large Numbers**.

We look at a general case next.
Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$\Pr\left[ |X_1 + \cdots + X_n - \mu| \geq \varepsilon \right] \to 0,$$

as $n \to \infty$. 

Proof: Let $Y_n = X_1 + \cdots + X_n$. Then

$$\Pr\left[ |Y_n - \mu| \geq \varepsilon \right] \leq \text{var} \left[ Y_n \right] \varepsilon^2 \leq \text{var} \left[ X_1 + \cdots + X_n \right] n \varepsilon^2 \leq n \varepsilon^2 \to 0,$$

as $n \to \infty$. 

Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. 
Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$Pr\left[ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right] \to 0, \text{ as } n \to \infty.$$
Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$Pr\left[\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right] \to 0, \text{ as } n \to \infty.$$ 

**Proof:**
Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then
Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$\Pr\left[ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right] \to 0, \text{ as } n \to \infty.$$  

**Proof:**
Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$\Pr[|Y_n - \mu| \geq \varepsilon] \leq \frac{\text{var}[Y_n]}{\varepsilon^2}$$
Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$Pr[|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon] \to 0, \text{ as } n \to \infty.$$ 

**Proof:**

Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$Pr[|Y_n - \mu| \geq \varepsilon] \leq \frac{\text{var}[Y_n]}{\varepsilon^2} = \frac{\text{var}[X_1 + \cdots + X_n]}{n^2 \varepsilon^2}.$$
**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$Pr\left[\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right] \to 0, \text{ as } n \to \infty.$$  

**Proof:**

Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$Pr[|Y_n - \mu| \geq \varepsilon] \leq \frac{\text{var}[Y_n]}{\varepsilon^2} = \frac{\text{var}[X_1 + \cdots + X_n]}{n^2 \varepsilon^2} = \frac{n \text{var}[X_1]}{n^2 \varepsilon^2}.$$
Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$\Pr[|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon] \to 0,$$

as $n \to \infty$.

**Proof:**

Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$\Pr[|Y_n - \mu| \geq \varepsilon] \leq \frac{\text{var}[Y_n]}{\varepsilon^2} = \frac{\text{var}[X_1 + \cdots + X_n]}{n^2 \varepsilon^2}$$

$$= \frac{n\text{var}[X_1]}{n^2 \varepsilon^2} = \frac{\text{var}[X_1]}{n \varepsilon^2}.$$
Theorem Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$Pr[|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon] \to 0, \text{ as } n \to \infty.$$  

Proof:

Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$Pr[|Y_n - \mu| \geq \varepsilon] \leq \frac{\text{var}[Y_n]}{\varepsilon^2} = \frac{\text{var}[X_1 + \cdots + X_n]}{n^2 \varepsilon^2} = \frac{n \text{var}[X_1]}{n^2 \varepsilon^2} = \frac{\text{var}[X_1]}{n \varepsilon^2} \to 0, \text{ as } n \to \infty.$$
Summary

Variance; Inequalities; WLLN

- **Variance:** \( \text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2 \)
Summary

- **Variance:** \( \text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2 \)
- **Fact:** \( \text{var}[aX + b] = a^2 \text{var}[X] \)
Variance; Inequalities; WLLN

- **Variance:** \( \text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2 \)
- **Fact:** \( \text{var}[aX + b] = a^2 \text{var}[X] \)
- **Sum:** \( X, Y, Z \) pairwise ind. \( \Rightarrow \text{var}[X + Y + Z] = \cdots \)
Summary

Variance; Inequalities; WLLN

- **Variance:** \( \text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2 \)
- **Fact:** \( \text{var}[aX + b] = a^2 \text{var}[X] \)
- **Sum:** \( X, Y, Z \) pairwise ind. \( \Rightarrow \text{var}[X + Y + Z] = \cdots \)
- **Markov:** \( \Pr[X \geq a] \leq E[f(X)]/f(a) \) where ...
Summary

Variance; Inequalities; WLLN

- **Variance**: \( \text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2 \)
- **Fact**: \( \text{var}[aX + b] = a^2 \text{var}[X] \)
- **Sum**: \( X, Y, Z \) pairwise ind. \( \Rightarrow \text{var}[X + Y + Z] = \cdots \)
- **Markov**: \( \Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)} \) where ...
- **Chebyshev**: \( \Pr[|X - E[X]| \geq a] \leq \frac{\text{var}[X]}{a^2} \)
Summary

Variance; Inequalities; WLLN

- **Variance**: \( \text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2 \)
- **Fact**: \( \text{var}[aX + b] = a^2 \text{var}[X] \)
- **Sum**: \( X, Y, Z \) pairwise ind. \( \Rightarrow \text{var}[X + Y + Z] = \cdots \)
- **Markov**: \( Pr[X \geq a] \leq \frac{E[f(X)}/f(a)] \text{ where } \cdots \)
- **Chebyshev**: \( Pr[|X - E[X]| \geq a] \leq \frac{\text{var}[X]}{a^2} \)
- **WLLN**: \( X_m \) i.i.d. \( \Rightarrow \frac{X_1 + \cdots + X_n}{n} \approx E[X] \)
Confidence?

You flip a coin once and get $H$. Do you think that $\Pr[H] = 1$?

You flip a coin 10 times and get 5 $H$s. Are you sure that $\Pr[H] = 0.5$?

You flip a coin 106 times and get 35% of $H$s. How much are you willing to bet that $\Pr[H]$ is exactly 0.35?

How much are you willing to bet that $\Pr[H] \in [0.3, 0.4]$?

Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$. Your estimate is $\hat{\theta}$. How much confidence do you have in your estimate?
Confidence?

- You flip a coin once and get $H$. 

- You flip a coin 10 times and get 5 $H$s. Are you sure that $\Pr[H] = 0.5$?

- You flip a coin 106 times and get 35% of $H$s. How much are you willing to bet that $\Pr[H]$ is exactly 0.35? How much are you willing to bet that $\Pr[H] \in [0.3, 0.4]$?

- Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$. Your estimate is $\hat{\theta}$. How much confidence do you have in your estimate?
Confidence?

- You flip a coin once and get $H$.
  Do think that $Pr[H] = 1$?
Confidence?

- You flip a coin once and get $H$. Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.

How much are you willing to bet that $Pr[H]$ is exactly 0.35?

How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$?

Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$. Your estimate is $\hat{\theta}$. How much confidence do you have in your estimate?
Confidence?

- You flip a coin once and get $H$. Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s. Are you sure that $Pr[H] = 0.5$?
Confidence?

- You flip a coin once and get $H$.
  Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s.

Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$.
Your estimate is $\hat{\theta}$.
How much confidence do you have in your estimate?
Confidence?

- You flip a coin once and get $H$.
  Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s.
  How much are you willing to bet that $Pr[H]$ is exactly 0.35?
Confidence?

- You flip a coin once and get $H$.
  Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s.
  How much are you willing to bet that $Pr[H]$ is exactly 0.35?
  How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$?
Confidence?

- You flip a coin once and get $H$.
  Do you think that $Pr[H] = 1$?

- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $Pr[H] = 0.5$?

- You flip a coin $10^6$ times and get 35% of $H$s.
  How much are you willing to bet that $Pr[H]$ is exactly 0.35?
  How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$?

Did different exam rooms perform differently?
Confidence?

- You flip a coin once and get $H$.
  Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s.
  How much are you willing to bet that $Pr[H]$ is exactly 0.35?
  How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$?
  Did different exam rooms perform differently? (6 afraid of 7?)
Confidence?

- You flip a coin once and get $H$.
  Do you think that $\Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $\Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s.
  How much are you willing to bet that $\Pr[H]$ is exactly 0.35?
  How much are you willing to bet that $\Pr[H] \in [0.3, 0.4]$?
  Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$. 
Confidence?

- You flip a coin once and get $H$. Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s. Are you sure that $Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s. How much are you willing to bet that $Pr[H]$ is exactly 0.35? How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$? Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$. Your estimate is $\hat{\theta}$. 
Confidence?

- You flip a coin once and get $H$.
  Do you think that $Pr[H] = 1$?
- You flip a coin 10 times and get 5 $H$s.
  Are you sure that $Pr[H] = 0.5$?
- You flip a coin $10^6$ times and get 35% of $H$s.
  How much are you willing to bet that $Pr[H]$ is exactly 0.35?
  How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$?
  Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$.
Your estimate is $\hat{\theta}$.
How much confidence do you have in your estimate?
Confidence?

Confidence is essential in many applications:

▶ How effective is a medication?
▶ Are we sure of the mileage of a car?
▶ Can we guarantee the lifespan of a device?
▶ We simulated a system. Do we trust the simulation results?
▶ Is an algorithm guaranteed to be fast?
▶ Do we know that a program has no bug?

As scientists and engineers, be convinced of this fact:
An estimate without confidence level is useless!
Confidence?

Confidence is essential in many applications:
Confidence?

Confidence is essential in many applications:

➤ How effective is a medication?
Confidence?

Confidence is essential in many applications:

- How effective is a medication?
- Are we sure of the mileage of a car?
Confidence?

Confidence is essential in many applications:

- How effective is a medication?
- Are we sure of the mileage of a car?
- Can we guarantee the lifespan of a device?

As scientists and engineers, be convinced of this fact: An estimate without confidence level is useless!
Confidence?

Confidence is essential in many applications:

▶ How effective is a medication?
▶ Are we sure of the mileage of a car?
▶ Can we guarantee the lifespan of a device?
▶ We simulated a system. Do we trust the simulation results?

As scientists and engineers, be convinced of this fact: An estimate without confidence level is useless!
Confidence?

Confidence is essential in many applications:

- How effective is a medication?
- Are we sure of the mileage of a car?
- Can we guarantee the lifespan of a device?
- We simulated a system. Do we trust the simulation results?
- Is an algorithm guaranteed to be fast?
Confidence?

Confidence is essential in many applications:

- How effective is a medication?
- Are we sure of the mileage of a car?
- Can we guarantee the lifespan of a device?
- We simulated a system. Do we trust the simulation results?
- Is an algorithm guaranteed to be fast?
- Do we know that a program has no bug?
Confidence?

Confidence is essential in many applications:

- How effective is a medication?
- Are we sure of the mileage of a car?
- Can we guarantee the lifespan of a device?
- We simulated a system. Do we trust the simulation results?
- Is an algorithm guaranteed to be fast?
- Do we know that a program has no bug?

As scientists and engineers, be convinced of this fact:
Confidence?

Confidence is essential in many applications:
- How effective is a medication?
- Are we sure of the mileage of a car?
- Can we guarantee the lifespan of a device?
- We simulated a system. Do we trust the simulation results?
- Is an algorithm guaranteed to be fast?
- Do we know that a program has no bug?

As scientists and engineers, be convinced of this fact:

An estimate without confidence level is useless!
Confidence Interval

Definition: Confidence Interval

An interval \([a, b]\) is a 95\% confidence interval for an unknown quantity \(\theta\) if

\[
\Pr[\theta \in [a, b]] \geq 95\%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework.

Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).

For instance, \(X_n = B(\theta)\).

Thus, more precisely, given \(\theta\), the random variables \(X_n\) are i.i.d. with a known distribution (that depends on \(\theta\)).

\[\triangleright\]

We observe \(X_1, \ldots, X_n\)

\[\triangleright\]

We calculate \(a = a(X_1, \ldots, X_n)\) and \(b = b(X_1, \ldots, X_n)\)

\[\triangleright\]

If we can guarantee that \(\Pr[\theta \in [a, b]] \geq 95\%\), then \([a, b]\) is a 95\% CI for \(\theta\).
Confidence Interval

The following definition captures precisely the notion of confidence.
**Confidence Interval**

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\% confidence interval for an unknown quantity \(\theta\) if

\[\Pr[\theta \in [a, b]] \geq 95\%\]

The interval \([a, b]\) is calculated on the basis of observations. Here is a typical framework.

Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).

For instance, \(X_n = B(\theta)\).

Thus, more precisely, given \(\theta\), the random variables \(X_n\) are i.i.d. with a known distribution (that depends on \(\theta\)).

▶ We observe \(X_1, \ldots, X_n\)
▶ We calculate \(a = a(X_1, \ldots, X_n)\) and \(b = b(X_1, \ldots, X_n)\)
▶ If we can guarantee that \(\Pr[\theta \in [a, b]] \geq 95\%\), then \([a, b]\) is a 95\% CI for \(\theta\).
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval $[a, b]$ is a 95%-confidence interval for an unknown quantity $\theta$ if

$$\Pr[\theta \in [a, b]] \geq 95\%.$$
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]

The interval \([a, b]\) is calculated on the basis of observations.
The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework.
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95\%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework. Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).
Confidence Interval

The following definition captures precisely the notion of confidence.

Definition: Confidence Interval

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework. Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).

For instance, \(X_n = B(\theta)\).
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval $[a, b]$ is a 95%-confidence interval for an unknown quantity $\theta$ if

$$Pr[\theta \in [a, b]] \geq 95\%.$$ 

The interval $[a, b]$ is calculated on the basis of observations. Here is a typical framework. Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. and have a distribution that depends on some parameter $\theta$.

For instance, $X_n = B(\theta)$.

Thus, more precisely,
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval $[a, b]$ is a 95%-confidence interval for an unknown quantity $\theta$ if

$$Pr[\theta \in [a, b]] \geq 95\%.$$ 

The interval $[a, b]$ is calculated on the basis of observations.

Here is a typical framework. Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. and have a distribution that depends on some parameter $\theta$.

For instance, $X_n = B(\theta)$.

Thus, more precisely, given $\theta$, the random variables $X_n$ are i.i.d. with a known distribution (that depends on $\theta$).
**Confidence Interval**

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval $[a, b]$ is a 95%-confidence interval for an unknown quantity $\theta$ if

$$Pr[\theta \in [a, b]] \geq 95\%.$$ 

The interval $[a, b]$ is calculated on the basis of observations.

Here is a typical framework. Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. and have a distribution that depends on some parameter $\theta$.

For instance, $X_n = B(\theta)$.

Thus, more precisely, given $\theta$, the random variables $X_n$ are i.i.d. with a known distribution (that depends on $\theta$).

- We observe $X_1, \ldots, X_n$
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework. Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).

For instance, \(X_n = B(\theta)\).

Thus, more precisely, given \(\theta\), the random variables \(X_n\) are i.i.d. with a known distribution (that depends on \(\theta\)).

- We observe \(X_1, \ldots, X_n\)
- We calculate \(a = a(X_1, \ldots, X_n)\) and \(b = b(X_1, \ldots, X_n)\)
Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework. Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).

For instance, \(X_n = B(\theta)\).

Thus, more precisely, given \(\theta\), the random variables \(X_n\) are i.i.d. with a known distribution (that depends on \(\theta\)).

- We observe \(X_1, \ldots, X_n\)
- We calculate \(a = a(X_1, \ldots, X_n)\) and \(b = b(X_1, \ldots, X_n)\)
- If we can guarantee that \(Pr[\theta \in [a, b]] \geq 95\%\),
**Confidence Interval**

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval \([a, b]\) is a 95\%-confidence interval for an unknown quantity \(\theta\) if

\[
Pr[\theta \in [a, b]] \geq 95%.
\]

The interval \([a, b]\) is calculated on the basis of observations.

Here is a typical framework. Assume that \(X_1, X_2, \ldots, X_n\) are i.i.d. and have a distribution that depends on some parameter \(\theta\).

For instance, \(X_n = B(\theta)\).

Thus, more precisely, given \(\theta\), the random variables \(X_n\) are i.i.d. with a known distribution (that depends on \(\theta\)).

- We observe \(X_1, \ldots, X_n\)
- We calculate \(a = a(X_1, \ldots, X_n)\) and \(b = b(X_1, \ldots, X_n)\)
- If we can guarantee that \(Pr[\theta \in [a, b]] \geq 95%\), then \([a, b]\) is a 95\%-CI for \(\theta\).
Confidence Interval: Applications

We poll 1000 people. Among those, 48\% declare they will vote for Trump. We do some calculations...

We conclude that $[0.43, 0.53]$ is a 95\% CI for the fraction of all the voters who will vote for Trump.

We observe 1,000 heart valve replacements that were performed by Dr. Bill. Among those, 35 patients died during surgery. (Sad example!)

We do some calculations...

We conclude that $[1\%, 5\%]$ is a 95\% CI for the probability of dying during that surgery by Dr. Bill.

We do a similar calculation for Dr. Fred. We find that $[8\%, 12\%]$ is a 95\% CI for Dr. Fred's surgery.

What surgeon do you choose?
Confidence Interval: Applications

- We poll 1000 people.

Among those, 48% declare they will vote for Trump.

We do some calculations ...

We conclude that $[0.43, 0.53]$ is a 95% CI for the fraction of all the voters who will vote for Trump.

We observe 1,000 heart valve replacements that were performed by Dr. Bill.

Among those, 35 patients died during surgery. (Sad example!)

We do some calculations ...

We conclude that $[1\%, 5\%]$ is a 95% CI for the probability of dying during that surgery by Dr. Bill.

We do a similar calculation for Dr. Fred.

We find that $[8\%, 12\%]$ is a 95% CI for Dr. Fred's surgery.

What surgeon do you choose?
We poll 1000 people.

Among those, 48% declare they will vote for Trump.
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery.
  - (Sad example!)
  - We do some calculations ...
  - We conclude that $[0.43, 0.53]$ is a 95% CI for the probability of dying during that surgery by Dr. Bill.

- We do a similar calculation for Dr. Fred.
  - We find that $[8\%, 12\%]$ is a 95% CI for Dr. Fred's surgery.

- What surgeon do you choose?
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
  - We conclude that $[1\%, 5\%]$ is a 95%-CI for the probability of dying during that surgery by Dr. Bill.

- We do a similar calculation for Dr. Fred.
  - We find that $[8\%, 12\%]$ is a 95%-CI for Dr. Fred's surgery.

- What surgeon do you choose?
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that \([0.43, 0.53]\) is a 95\%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
  - We conclude that \([1\%, 5\%]\) is a 95\%-CI for the probability of dying during that surgery by Dr. Bill.

- We do a similar calculation for Dr. Fred.
  - We find that \([8\%, 12\%]\) is a 95\%-CI for Dr. Fred’s surgery.

- What surgeon do you choose?
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that [0.43, 0.53] is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
Confidence Interval: Applications

▶ We poll 1000 people.
  ▶ Among those, 48% declare they will vote for Trump.
  ▶ We do some calculations ....
  ▶ We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.

▶ We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  ▶ Among those, 35 patients died during surgery.
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that \([0.43, 0.53]\) is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
  - We conclude that \([1\%, 5\%]\) is a 95%-CI for the probability of dying during that surgery by Dr. Bill.
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
  - We conclude that $[1\%, 5\%]$ is a 95%-CI for the probability of dying during that surgery by Dr. Bill.
  - We do a similar calculation for Dr. Fred.
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
  - We conclude that $[1\%, 5\%]$ is a 95%-CI for the probability of dying during that surgery by Dr. Bill.
  - We do a similar calculation for Dr. Fred.
  - We find that $[8\%, 12\%]$ is a 95%-CI for Dr. Fred’s surgery.
Confidence Interval: Applications

- We poll 1000 people.
  - Among those, 48% declare they will vote for Trump.
  - We do some calculations ....
  - We conclude that \([0.43, 0.53]\) is a 95%-CI for the fraction of all the voters who will vote for Trump.

- We observe 1,000 heart valve replacements that were performed by Dr. Bill.
  - Among those, 35 patients died during surgery. (Sad example!)
  - We do some calculations ...
  - We conclude that \([1\%, 5\%]\) is a 95%-CI for the probability of dying during that surgery by Dr. Bill.
  - We do a similar calculation for Dr. Fred.
  - We find that \([8\%, 12\%]\) is a 95%-CI for Dr. Fred’s surgery.
  - What surgeon do you choose?
Coin Flips: Intuition

Say that you flip a coin \( n = 100 \) times and observe 20 Hs. If \( p := \Pr[H] = 0.5 \), this event is very unlikely. Intuitively, it is unlikely that the fraction of Hs, say \( A_n \), differs a lot from \( p := \Pr[H] \). Thus, it is unlikely that \( p \) differs a lot from \( A_n \). Hence, one should be able to build a confidence interval \([A_n - \varepsilon, A_n + \varepsilon]\) for \( p \).

The key idea is that \(|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon]\). Thus, \( \Pr[|A_n - p| > \varepsilon] \leq 5\% \iff \Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\% \).

It remains to find \( \varepsilon \) such that \( \Pr[|A_n - p| > \varepsilon] \leq 5\% \).

One approach: Chebyshev.
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs. If $p := \Pr[H] = 0.5$, this event is very unlikely. Intuitively, it is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := \Pr[H]$. Thus, it is unlikely that $p$ differs a lot from $A_n$. Hence, one should be able to build a confidence interval $[A_n - \varepsilon, A_n + \varepsilon]$ for $p$. The key idea is that $|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon]$. Thus, $\Pr[|A_n - p| > \varepsilon] \leq 5\% \iff \Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\%$. It remains to find $\varepsilon$ such that $\Pr[|A_n - p| > \varepsilon] \leq 5\%$. One approach: Chebyshev.
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs.
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

The key idea is that $|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon]$.

Thus, $Pr[|A_n - p| > \varepsilon] \leq 5\% \iff Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\%$.

It remains to find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$.

One approach: Chebyshev.
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$,
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := Pr[H]$. 

Say that you flip a coin \( n = 100 \) times and observe 20 Hs.

If \( p := Pr[H] = 0.5 \), this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say \( A_n \), differs a lot from \( p := Pr[H] \).

Thus, it is unlikely that \( p \) differs a lot from \( A_n \).
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := Pr[H]$.

Thus, it is unlikely that $p$ differs a lot from $A_n$. Hence, one should be able to build a confidence interval $[A_n - \varepsilon, A_n + \varepsilon]$ for $p$. 
Say that you flip a coin \( n = 100 \) times and observe 20 Hs.

If \( p := Pr[H] = 0.5 \), this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say \( A_n \), differs a lot from \( p := Pr[H] \).

Thus, it is unlikely that \( p \) differs a lot from \( A_n \). Hence, one should be able to build a confidence interval \( [A_n - \varepsilon, A_n + \varepsilon] \) for \( p \).

The key idea is that \( |A_n - p| \leq \varepsilon \Leftrightarrow p \in [A_n - \varepsilon, A_n + \varepsilon] \).
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := Pr[H]$.

Thus, it is unlikely that $p$ differs a lot from $A_n$. Hence, one should be able to build a confidence interval $[A_n - \varepsilon, A_n + \varepsilon]$ for $p$.

The key idea is that $|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon]$.

Thus, $Pr[|A_n - p| > \varepsilon] \leq 5\% \iff Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\%$. 
Coin Flips: Intuition

Say that you flip a coin $n = 100$ times and observe $20$ Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := Pr[H]$.

Thus, it is unlikely that $p$ differs a lot from $A_n$. Hence, one should be able to build a confidence interval $[A_n - \varepsilon, A_n + \varepsilon]$ for $p$.

The key idea is that $|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon]$.

Thus, $Pr[|A_n - p| > \varepsilon] \leq 5\% \iff Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\%$.

It remains to find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$. 
Coin Flips: Intuition

Say that you flip a coin \( n = 100 \) times and observe 20 Hs.

If \( p := Pr[H] = 0.5 \), this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say \( A_n \), differs a lot from \( p := Pr[H] \).

Thus, it is unlikely that \( p \) differs a lot from \( A_n \). Hence, one should be able to build a confidence interval \([A_n - \varepsilon, A_n + \varepsilon]\) for \( p \).

The key idea is that \(|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon] \).

Thus, \( Pr[|A_n - p| > \varepsilon] \leq 5\% \iff Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\% \).

It remains to find \( \varepsilon \) such that \( Pr[|A_n - p| > \varepsilon] \leq 5\% \).

One approach:
Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := Pr[H]$.

Thus, it is unlikely that $p$ differs a lot from $A_n$. Hence, one should be able to build a confidence interval $[A_n - \varepsilon, A_n + \varepsilon]$ for $p$.

The key idea is that $|A_n - p| \leq \varepsilon \iff p \in [A_n - \varepsilon, A_n + \varepsilon]$.

Thus, $Pr[|A_n - p| > \varepsilon] \leq 5\% \iff Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\%$.

It remains to find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$.

One approach: Chebyshev.
Confidence Interval with Chebyshev

Flip a coin \( n \) times. Let \( A_n \) be the fraction of heads. Can we find \( \varepsilon \) such that \( \Pr[|A_n - p| > \varepsilon] \leq 5\% \)? Using Chebyshev, we will see that \( \varepsilon = \frac{2.25}{\sqrt{n}} \) works. Thus \([A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]\) is a 95%-CI for \( p \).

Example: If \( n = 1500 \), then \( \Pr[p \in [A_n - 0.05, A_n + 0.05]] \geq 95\% \).

In fact, \( a = \frac{1}{\sqrt{n}} \) works, so that with \( n = 1500 \), one has \( \Pr[p \in [A_n - 0.02, A_n + 0.02]] \geq 95\% \).
Flip a coin \( n \) times.
Confidence Interval with Chebyshev

- Flip a coin $n$ times. Let $A_n$ be the fraction of $H$s.
Confidence Interval with Chebyshev

- Flip a coin $n$ times. Let $A_n$ be the fraction of $H$s.
- Can we find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$?
Flip a coin \( n \) times. Let \( A_n \) be the fraction of \( Hs \).

Can we find \( \varepsilon \) such that \( Pr[|A_n - p| > \varepsilon] \leq 5\% \)?

Using Chebyshev, we will see that \( \varepsilon = 2.25 \frac{1}{\sqrt{n}} \) works.
Confidence Interval with Chebyshev

- Flip a coin $n$ times. Let $A_n$ be the fraction of $H$s.
- Can we find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$?

Using Chebyshev, we will see that $\varepsilon = 2.25 \frac{1}{\sqrt{n}}$ works. Thus

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$. 
Flip a coin \( n \) times. Let \( A_n \) be the fraction of \( H \)s.

Can we find \( \varepsilon \) such that \( Pr[|A_n - p| > \varepsilon] \leq 5\% \)?

Using Chebyshev, we will see that \( \varepsilon = 2.25 \frac{1}{\sqrt{n}} \) works. Thus

\[
[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]
\]

is a 95\%-CI for \( p \).

Example: If \( n = 1500 \), then \( Pr[p \in [A_n - 0.05, A_n + 0.05]] \geq 95\% \).
Confidence Interval with Chebyshev

- Flip a coin $n$ times. Let $A_n$ be the fraction of $H$s.
- Can we find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$?

Using Chebyshev, we will see that $\varepsilon = 2.25 \frac{1}{\sqrt{n}}$ works. Thus

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$.

Example: If $n = 1500$, then $Pr[p \in [A_n - 0.05, A_n + 0.05]] \geq 95\%$.

In fact, $a = \frac{1}{\sqrt{n}}$ works,
Confidence Interval with Chebyshev

Flip a coin $n$ times. Let $A_n$ be the fraction of Hs.

Can we find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$?

Using Chebyshev, we will see that $\varepsilon = 2.25 \frac{1}{\sqrt{n}}$ works. Thus

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95\%-CI for $p$.

Example: If $n = 1500$, then $Pr[p \in [A_n - 0.05, A_n + 0.05]] \geq 95\%$.

In fact, $a = \frac{1}{\sqrt{n}}$ works, so that with $n = 1,500$ one has

$Pr[p \in [A_n - 0.02, A_n + 0.02]] \geq 95\%$. 
Theorem: Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = X_1 + \cdots + X_n$. Then, 

$$\Pr\left[ \mu \in \left[ A_n - 4.5\sigma\sqrt{n}, A_n + 4.5\sigma\sqrt{n} \right] \right] \geq 95\%.$$ 

Thus, $[A_n - 4.5\sigma\sqrt{n}, A_n + 4.5\sigma\sqrt{n}]$ is a 95\% CI for $\mu$.

Example: Let $X_n = 1\{\text{coin } n \text{ yields H}\}$. Then $\mu = E[X_n] = p := \Pr[H]$. Also, $\sigma^2 = \text{var}(X_n) = p(1-p) \leq 1/4$. Hence, $[A_n - 4.51/2\sqrt{n}, A_n + 4.51/2\sqrt{n}]$ is a 95\% CI for $p$. 

Confidence Intervals: Result
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$.

\[ \Pr\left[ \mu \in \left[ A_n - 4.5 \sigma \sqrt{\frac{1}{n}}, A_n + 4.5 \sigma \sqrt{\frac{1}{n}} \right] \right] \geq 95\% \]

Thus, $\left[ A_n - 4.5 \sigma \sqrt{\frac{1}{n}}, A_n + 4.5 \sigma \sqrt{\frac{1}{n}} \right]$ is a 95\% CI for $\mu$. 

Example:
Let $X_n = 1\{\text{coin } n \text{ yields } H\}$.
Then $\mu = \mathbb{E}[X_n] = p = \Pr[H]$.
Also, $\sigma^2 = \text{var}(X_n) = p(1-p) \leq \frac{1}{4}$. 

Hence, $\left[ A_n - 4.5 \cdot \frac{1}{2} \sqrt{\frac{1}{n}}, A_n + 4.5 \cdot \frac{1}{2} \sqrt{\frac{1}{n}} \right]$ is a 95\% CI for $p$. 

Confidence Intervals: Result

Theorem:
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$.
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$
**Theorem:**

Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$\Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$ 

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$. 

---

**Example:**

Let $X_n = 1\{\text{coin } n \text{ yields } H\}$. Then $\mu = \mathbb{E}[X_n] = p = \Pr[H]$. Also, $\sigma^2 = \text{var}(X_n) = p(1-p) \leq 1/4$. Hence,

$$[A_n - 4.5 \frac{1/2}{\sqrt{n}}, A_n + 4.5 \frac{1/2}{\sqrt{n}}]$$

is a 95%-CI for $p$. 

---

**Confidence Intervals: Result**
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$ 

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$.

**Example:**
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$  

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$.

**Example:** Let $X_n = 1\{\text{coin } n \text{ yields } H\}$. 
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$.
Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$ 

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$.

Example: Let $X_n = 1\{\text{coin } n \text{ yields } H\}$. Then

$$\mu = E[X_n] = p := Pr[H].$$
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \ldots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$ 

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$.

**Example:** Let $X_n = 1\{\text{coin } n \text{ yields } H\}$. Then

$$\mu = E[X_n] = p := Pr[H].$$ 

Also, $\sigma^2 = \text{var}(X_n) = $
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$.

Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$.

Example: Let $X_n = 1\{\text{coin n yields } H\}$. Then

$$\mu = E[X_n] = p := Pr[H].$$
Also, $\sigma^2 = var(X_n) = p(1 - p)$
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$ 

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95\%-CI for $\mu$.

**Example:** Let $X_n = 1\{\text{coin } n \text{ yields } H\}$. Then

$$\mu = E[X_n] = p : = Pr[H].$$ 
Also, $\sigma^2 = \text{var}(X_n) = p(1 - p) \leq \frac{1}{4}$. 
Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$ 

Thus, $[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}}]$ is a 95%-CI for $\mu$.

**Example:** Let $X_n = 1\{\text{coin } n \text{ yields } H\}$. Then

$$\mu = E[X_n] = p := Pr[H]. \text{ Also, } \sigma^2 = \text{var}(X_n) = p(1-p) \leq \frac{1}{4}.$$ 

Hence, $[A_n - 4.5 \frac{1/2}{\sqrt{n}}, A_n + 4.5 \frac{1/2}{\sqrt{n}}]$ is a 95%-CI for $p$. 
We prove the theorem, i.e., that \( A_n \pm 4.5 \sigma / \sqrt{n} \) is a 95\% CI for \( \mu \).

From Chebyshev:

\[
\Pr \left[ |A_n - \mu| \geq 4.5 \sigma / \sqrt{n} \right] \leq \frac{\text{var}(A_n)}{4.5^2 \sigma^2 / n} = \frac{n}{20}.
\]

Thus,

\[
\Pr \left[ |A_n - \mu| \leq 4.5 \sigma / \sqrt{n} \right] \geq 95\%.
\]

Hence,

\[
\Pr \left[ \mu \in \left[ A_n - 4.5 \sigma / \sqrt{n}, A_n + 4.5 \sigma / \sqrt{n} \right] \right] \geq 95\%.
\]
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$. 
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma / \sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma / \sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma / \sqrt{n}]^2}$$
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma/\sqrt{n}]^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$
Confidence Interval: Analysis

We prove the theorem, i.e., that \( A_n \pm 4.5\sigma / \sqrt{n} \) is a 95\%-CI for \( \mu \).

From Chebyshev:

\[
Pr[|A_n - \mu| \geq 4.5\sigma / \sqrt{n}] \leq \frac{\text{var}(A_n)}{(4.5\sigma / \sqrt{n})^2} = \frac{n}{20\sigma^2} \text{var}(A_n).
\]

Now,

\[
\text{var}(A_n) = \text{var}(\frac{X_1 + \cdots + X_n}{n}) =
\]
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma/\sqrt{n}]^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$

Now,

$$\text{var}(A_n) = \text{var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n)$$
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95\%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma/\sqrt{n}]^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$

Now,

$$\text{var}(A_n) = \text{var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2} \times n \cdot \text{var}(X_1) = \cdots$$
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma/\sqrt{n}]^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$

Now,

$$\text{var}(A_n) = \text{var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2} \times n \cdot \text{var}(X_1) = \frac{1}{n} \sigma^2.$$
Confidence Interval: Analysis

We prove the theorem, i.e., that \( A_n \pm 4.5\sigma/\sqrt{n} \) is a 95\%-CI for \( \mu \).

From Chebyshev:

\[
Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{(4.5\sigma/\sqrt{n})^2} = \frac{n}{20\sigma^2} \text{var}(A_n).
\]

Now,

\[
\text{var}(A_n) = \text{var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n)
\]

\[
= \frac{1}{n^2} \times n \cdot \text{var}(X_1) = \frac{1}{n} \sigma^2.
\]

Hence,

\[
Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{n}{20\sigma^2} \times \frac{1}{n} \sigma^2
\]
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma/\sqrt{n}]^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$

Now,

$$\text{var}(A_n) = \text{var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2} \times n \text{var}(X_1) = \frac{1}{n} \sigma^2.$$

Hence,

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{n}{20\sigma^2} \times \frac{1}{n} \sigma^2 = 5\%.$$
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma / \sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma / \sqrt{n}] \leq \frac{\text{var}(A_n)}{[4.5\sigma / \sqrt{n}]^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$

Now,

$$\text{var}(A_n) = \text{var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2} \times n \cdot \text{var}(X_1) = \frac{1}{n} \sigma^2.$$

Hence,

$$Pr[|A_n - \mu| \geq 4.5\sigma / \sqrt{n}] \leq \frac{n}{20\sigma^2} \times \frac{1}{n} \sigma^2 = 5\%.$$  

Thus,

$$Pr[|A_n - \mu| \leq 4.5\sigma / \sqrt{n}] \geq 95\%.$$
Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{\text{var}(A_n)}{(4.5\sigma/\sqrt{n})^2} = \frac{n}{20\sigma^2} \text{var}(A_n).$$

Now,

$$\text{var}(A_n) = \text{var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n) = \frac{1}{n^2} \times n \cdot \text{var}(X_1) = \frac{1}{n} \sigma^2.$$

Hence,

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{n}{20\sigma^2} \times \frac{1}{n} \sigma^2 = 5\%.$$

Thus,

$$Pr[|A_n - \mu| \leq 4.5\sigma/\sqrt{n}] \geq 95\%.$$ 

Hence,

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$
Confidence interval for $p$ in $B(p)$
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. 
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$. 

Theorem: 

$[A_n - 1.96\sqrt{n}, A_n + 1.96\sqrt{n}]$ is a 95%-CI for $p$.

Proof: 

We have just seen that $Pr[\mu \in [A_n - 1.96\sigma/\sqrt{n}, A_n + 1.96\sigma/\sqrt{n}]] \geq 95\%$.

Here, $\mu = p$ and $\sigma^2 = p(1-p)$.

Thus, $\sigma^2 \leq 1/4$ and $\sigma \leq 1/2$.

Thus, $Pr[\mu \in [A_n - 1.960.5/\sqrt{n}, A_n + 1.960.5/\sqrt{n}]] \geq 95\%$. 

Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$ is a 95%-CI for $p$. 

Proof: We have just seen that 

$$\Pr \left[ \mu \in \left[ A_n - \frac{4}{0.5} \frac{\sqrt{n}}{\sqrt{n}}, A_n + \frac{4}{0.5} \frac{\sqrt{n}}{\sqrt{n}} \right] \right] \geq 95\%.$$ 

Here, $\mu = p$ and $\sigma^2 = p(1-p)$. Thus, $\sigma \leq 0.5$ and $\sigma \leq 1.25$. Thus, 

$$\Pr \left[ \mu \in \left[ A_n - \frac{4 \times 0.5}{\sqrt{n}}, A_n + \frac{4 \times 0.5}{\sqrt{n}} \right] \right] \geq 95\%.$$
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$.

**Proof:**
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$ is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$

Here, $\mu = p$
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$ 

Here, $\mu = p$ and $\sigma^2 = p(1 - p)$. 
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$ is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$ 

Here, $\mu = p$ and $\sigma^2 = p(1 - p)$. Thus, $\sigma^2 \leq \frac{1}{4}$
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$ 

Here, $\mu = p$ and $\sigma^2 = p(1 - p)$. Thus, $\sigma^2 \leq \frac{1}{4}$ and $\sigma \leq \frac{1}{2}$. 

Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$ [A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}] $$

is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$ Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%. $$

Here, $\mu = p$ and $\sigma^2 = p(1 - p)$. Thus, $\sigma^2 \leq \frac{1}{4}$ and $\sigma \leq \frac{1}{2}$.

Thus,

$$ Pr[\mu \in [A_n - 4.5 \times 0.5/\sqrt{n}, A_n + 4.5 \times 0.5/\sqrt{n}]] \geq 95\%. $$
Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}]$$

is a 95%-CI for $p$.

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$ 

Here, $\mu = p$ and $\sigma^2 = p(1-p)$. Thus, $\sigma^2 \leq \frac{1}{4}$ and $\sigma \leq \frac{1}{2}$.

Thus,

$$Pr[\mu \in [A_n - 4.5 \times 0.5/\sqrt{n}, A_n + 4.5 \times 0.5/\sqrt{n}]] \geq 95\%.$$
Confidence interval for $p$ in $B(p)$
Confidence interval for $p$ in $B(p)$

An illustration:
Confidence interval for $p$ in $B(p)$

An illustration:

$$95\% \text{ CI for } p = \left[ A_n - 2.25 \frac{1}{\sqrt{n}}, A_n + 2.25 \frac{1}{\sqrt{n}} \right]$$
Confidence interval for $p$ in $B(p)$

An illustration:

Good practice: You run your simulation, or experiment.
Confidence interval for $p$ in $B(p)$

An illustration:

$$95\% \text{ - CI for } p = \left[ A_n - 2.25 \frac{1}{\sqrt{n}}, A_n + 2.25 \frac{1}{\sqrt{n}} \right]$$

Good practice: You run your simulation, or experiment. You get an estimate.
Confidence interval for $p$ in $B(p)$

An illustration:

$95\% - \text{CI for } p = [A_n - 2.25 \frac{1}{\sqrt{n}}, A_n + 2.25 \frac{1}{\sqrt{n}}]$

Good practice: You run your simulation, or experiment. You get an estimate. You indicate your confidence interval.
Confidence interval for $p$ in $B(p)$
Confidence interval for $p$ in $B(p)$

Improved CI:
Improved CI: In fact, one can replace 2.25 by 1.
Confidence interval for $p$ in $B(p)$

Improved CI: In fact, one can replace $2.25$ by $1$.

Quite a bit of work to get there:
Confidence interval for $p$ in $B(p)$

Improved CI: In fact, one can replace $2.25$ by $1$.

Quite a bit of work to get there: continuous random variables;
Confidence interval for $p$ in $B(p)$

Improved CI: In fact, one can replace 2.25 by 1.

Quite a bit of work to get there: continuous random variables; Gaussian;
Confidence interval for $p$ in $B(p)$

Improved CI: In fact, one can replace 2.25 by 1.

Quite a bit of work to get there: continuous random variables; Gaussian; Central Limit Theorem.
Confidence Interval for $1/p$ in $G(p)$
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. 

Theorem: 

$$\left[ A_n - 4.5/\sqrt{n}, A_n + 4.5/\sqrt{n} \right]$$

is a 95%-CI for $1/p$.

Proof:

We know that 

$$\Pr\left[ \mu \in \left[ A_n - 4.5/\sqrt{n}, A_n + 4.5/\sqrt{n} \right] \right] \geq 95\%.$$ 

Here, $\mu = 1/p$ and $\sigma = \sqrt{1/p - p} \leq 1/p$.

Hence, 

$$\Pr\left[ 1/p \in \left[ A_n - 4.5p/\sqrt{n}, A_n + 4.5p/\sqrt{n} \right] \right] \geq 95\%.$$ 

Now, 

$$A_n - 4.5/\sqrt{n} \leq 1/p \leq A_n + 4.5/\sqrt{n}$$

is equivalent to 

$$A_n + 4.5/\sqrt{n} \leq 1/p \leq A_n - 4.5/\sqrt{n}.$$ 

Examples:

$$\left[ 0.7, 1.8 \right]_{100}, \left[ 0.96, 1.05 \right]_{10000}.$$
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$
\left[ \frac{A_n}{1+4.5/\sqrt{n}}, \frac{A_n}{1-4.5/\sqrt{n}} \right]
$$

is a 95%-CI for $\frac{1}{p}$.
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]$$

is a 95%-CI for $1/p$.

**Proof:**
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[\frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}}]$$ is a 95%-CI for $\frac{1}{p}$.

**Proof:** We know that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}] \geq 95\%.$$. 
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]$$

is a 95%-CI for $\frac{1}{p}$.

**Proof:** We know that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$ 

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. 

Examples:

$[0.7 A_{100}, 1.8 A_{100}]$ and $[0.96 A_{10000}, 1.05 A_{10000}]$. 

Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]$$

is a 95%-CI for $\frac{1}{p}$.

**Proof:** We know that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$  

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. Hence,

$$Pr\left[ \frac{1}{p} \in [A_n - 4.5 \frac{1}{p\sqrt{n}}, A_n + 4.5 \frac{1}{p\sqrt{n}}] \right] \geq 95\%.$$
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

\[
\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]
\]
is a 95%-CI for $\frac{1}{p}$.

**Proof:** We know that

\[
Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95%.
\]

Here, $\mu = \frac{1}{p}$ and $\sigma = \sqrt{\frac{1-p}{p}} \leq \frac{1}{p}$. Hence,

\[
Pr[\frac{1}{p} \in [A_n - 4.5 \frac{1}{p\sqrt{n}}, A_n + 4.5 \frac{1}{p\sqrt{n}}]] \geq 95%.
\]

Now, $A_n - 4.5 \frac{1}{p\sqrt{n}} \leq \frac{1}{p} \leq \frac{1}{p} \leq A_n + 4.5 \frac{1}{p\sqrt{n}}$ is equivalent to
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

\[
\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]
\]

is a 95%-CI for $\frac{1}{p}$.

**Proof:** We know that

\[
Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95%.
\]

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. Hence,

\[
Pr[\frac{1}{p} \in [A_n - 4.5 \frac{1}{p\sqrt{n}}, A_n + 4.5 \frac{1}{p\sqrt{n}}]] \geq 95%.
\]

Now, $A_n - 4.5 \frac{1}{p\sqrt{n}} \leq \frac{1}{p} \leq A_n + 4.5 \frac{1}{p\sqrt{n}}$ is equivalent to

\[
\frac{A_n}{1 + 4.5/\sqrt{n}} \leq \frac{1}{p} \leq \frac{A_n}{1 - 4.5/\sqrt{n}}.
\]
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]$$

is a 95%-CI for $1/p$.

**Proof:** We know that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95%.$$ 

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. Hence,

$$Pr\left[ \frac{1}{p} \in [A_n - 4.5 \frac{1}{p\sqrt{n}}, A_n + 4.5 \frac{1}{p\sqrt{n}}] \right] \geq 95%.$$ 

Now, $A_n - 4.5 \frac{1}{p\sqrt{n}} \leq \frac{1}{p} \leq A_n + 4.5 \frac{1}{p\sqrt{n}}$ is equivalent to

$$\frac{A_n}{1 + 4.5/\sqrt{n}} \leq \frac{1}{p} \leq \frac{A_n}{1 - 4.5/\sqrt{n}}.$$ 

**Examples:**

- $[0.7 A_{100}, 1.8 A_{100}]$
- $[0.96 A_{10000}, 1.05 A_{10000}]$
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]$$

is a 95%-CI for $1/p$.

**Proof:** We know that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$ 

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. Hence,

$$Pr[\frac{1}{p} \in [A_n - 4.5 \frac{1}{p\sqrt{n}}, A_n + 4.5 \frac{1}{p\sqrt{n}}]] \geq 95\%.$$ 

Now, $A_n - 4.5 \frac{1}{p\sqrt{n}} \leq \frac{1}{p} \leq A_n + 4.5 \frac{1}{p\sqrt{n}}$ is equivalent to

$$\frac{A_n}{1 + 4.5/\sqrt{n}} \leq \frac{1}{p} \leq \frac{A_n}{1 - 4.5/\sqrt{n}}.$$ 

**Examples:** $[0.7A_{100}, 1.8A_{100}]$
Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$
\left[ \frac{A_n}{1 + 4.5/\sqrt{n}}, \frac{A_n}{1 - 4.5/\sqrt{n}} \right]
$$

is a 95%-CI for $1/p$.

**Proof:** We know that

$$
Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95%.
$$

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. Hence,

$$
Pr\left[ \frac{1}{p} \in [A_n - 4.5 \frac{1}{p\sqrt{n}}, A_n + 4.5 \frac{1}{p\sqrt{n}}] \right] \geq 95%.
$$

Now, $A_n - 4.5 \frac{1}{p\sqrt{n}} \leq \frac{1}{p} \leq A_n + 4.5 \frac{1}{p\sqrt{n}}$ is equivalent to

$$
\frac{A_n}{1 + 4.5/\sqrt{n}} \leq \frac{1}{p} \leq \frac{A_n}{1 - 4.5/\sqrt{n}}.
$$

**Examples:** $[0.7A_{100}, 1.8A_{100}]$ and $[0.96A_{10000}, 1.05A_{10000}]$. 
You are given coin A and coin B. You want to find out which one has a larger \( P[H] \).

Let \( p_A \) and \( p_B \) be the values of \( P[H] \) for the two coins.

**Approach:**

▶ Flip each coin \( n \) times.
▶ Let \( A_n \) be the fraction of Hs for coin A and \( B_n \) for coin B.
▶ Assume \( A_n > B_n \). It is tempting to think that \( p_A > p_B \).

**Confidence?**

**Analysis:**

Note that 

\[
E[A_n - B_n] = p_A - p_B \quad \text{and} \quad \text{var}(A_n - B_n) = \frac{1}{n}(p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}.
\]

Thus, 

\[
\Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n}\varepsilon^2,
\]

so 

\[
\Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n}(A_n - B_n)^2.
\]

**Example:**

With \( n = 100 \) and \( A_n - B_n = 0.2 \), 

\[
\Pr[p_A > p_B] \geq 1 - \frac{1}{16} = 0.875.
\]
Which Coin is Better?

You are given coin $A$ and coin $B$. 

Approach:

▶ Flip each coin $n$ times.

▶ Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.

▶ Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

Confidence?

Analysis:

Note that $E[A_n - B_n] = p_A - p_B$ and $\text{var}(A_n - B_n) \leq 1/2n$. 

Thus, $\Pr[|A_n - B_n - (p_A - p_B)| > \epsilon] \leq 1/2n\epsilon^2$, so $\Pr[p_A - p_B \in [A_n - B_n - \epsilon, A_n - B_n + \epsilon]] \geq 1 - 1/2n\epsilon^2$, and $\Pr[p_A - p_B \geq 0] \geq 1 - 1/2n(A_n - B_n)^2$.

Example: With $n = 100$ and $A_n - B_n = 0.2$, $\Pr[p_A > p_B] \geq 1 - 1/8 = 0.875$. 

Which Coin is Better?
You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. 

Approach:
▶ Flip each coin $n$ times.
▶ Let $A_n$ be the fraction of $H$s for coin $A$ and $B_n$ for coin $B$.
▶ Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

Confidence?
Analysis:
Note that $E[A_n - B_n] = p_A - p_B$ and $\text{var}(A_n - B_n) = \frac{1}{n}(p_A(1 - p_A) + p_B(1 - p_B)) \leq \frac{1}{2n}$.
Thus, $Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n\varepsilon^2}$, so $Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n\varepsilon^2}$.

Example:
With $n = 100$ and $A_n - B_n = 0.2$, $Pr[p_A > p_B] \geq 1 - \frac{1}{8} = 0.875$. 

Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

Approach:
▶ Flip each coin $n$ times.
▶ Let $A_n$ be the fraction of $H$s for coin $A$ and $B_n$ for coin $B$.
▶ Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

Confidence?

Analysis:
Note that $E[A_n - B_n] = p_A - p_B$ and $\text{var}(A_n - B_n) = \frac{1}{n}(p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}$.
Thus, $Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n \varepsilon^2}$, so $Pr[p_A - p_B \in (A_n - B_n - \varepsilon, A_n - B_n + \varepsilon)] \geq 1 - \frac{1}{2n \varepsilon^2}$.

Example:
With $n = 100$ and $A_n - B_n = 0.2$, $Pr[p_A > p_B] \geq 1 - \frac{1}{8} = 0.875$. 
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

▶ Flip each coin $n$ times.
▶ Let $A_n$ be the fraction of $H$s for coin $A$ and $B_n$ for coin $B$.
▶ Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:**

Note that $E[A_n - B_n] = p_A - p_B$ and $\text{var}(A_n - B_n) = \frac{1}{n}(p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}$.

Thus, $\Pr[|A_n - B_n - (p_A - p_B)| > \epsilon] \leq \frac{1}{2n \epsilon^2}$, so $\Pr[|p_A - p_B| \in [A_n - B_n - \epsilon, A_n - B_n + \epsilon]] \geq 1 - \frac{1}{2n \epsilon^2}$, and $\Pr[p_A - p_B \geq 0] \geq 1 - \frac{1}{2n (A_n - B_n)^2}$.

**Example:**

With $n = 100$ and $A_n - B_n = 0.2$, $\Pr[p_A > p_B] \geq 1 - \frac{1}{8} = 0.875$. 
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

Approach:

- Flip each coin $n$ times.
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$. 
Which Coin is Better?
You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

Approach:

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$. 

Confidence?

Analysis:

Note that $E[A_n - B_n] = p_A - p_B$ and $\text{var}(A_n - B_n) = \frac{1}{n}(p_A(1-p_A)+p_B(1-p_B)) \leq \frac{1}{2n}$.

Thus, $\Pr(|A_n - B_n - (p_A - p_B)| > \epsilon) \leq \frac{1}{2}\epsilon^2$,

so $\Pr[p_A - p_B \in [A_n - B_n - \epsilon, A_n - B_n + \epsilon]] \geq 1 - \frac{1}{2n}\epsilon^2$.

Example:

With $n = 100$ and $A_n - B_n = 0.2$, $\Pr[p_A > p_B] \geq 1 - \frac{1}{8} = 0.875$. 
Which Coin is Better?
You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**
- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:**
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:** Note that

$$E[A_n - B_n] =$$
Which Coin is Better?
You are given coin \(A\) and coin \(B\). You want to find out which one has a larger \(Pr[H]\). Let \(p_A\) and \(p_B\) be the values of \(Pr[H]\) for the two coins.

**Approach:**

- Flip each coin \(n\) times.
- Let \(A_n\) be the fraction of Hs for coin \(A\) and \(B_n\) for coin \(B\).
- Assume \(A_n > B_n\). It is tempting to think that \(p_A > p_B\).

**Confidence?**

**Analysis:** Note that

\[
E[A_n - B_n] = p_A - p_B \quad \text{and} \quad \text{var}(A_n - B_n) =
\]
Which Coin is Better?

You are given coin A and coin B. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin A and $B_n$ for coin B.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:** Note that

$$E[A_n - B_n] = p_A - p_B \text{ and } var(A_n - B_n) = \frac{1}{n}(p_A(1 - p_A) + p_B(1 - p_B)) \leq \frac{1}{2n}.$$
Which Coin is Better?

You are given coin \( A \) and coin \( B \). You want to find out which one has a larger \( Pr[H] \). Let \( p_A \) and \( p_B \) be the values of \( Pr[H] \) for the two coins.

**Approach:**

- Flip each coin \( n \) times.
- Let \( A_n \) be the fraction of Hs for coin \( A \) and \( B_n \) for coin \( B \).
- Assume \( A_n > B_n \). It is tempting to think that \( p_A > p_B \).

**Confidence?**

**Analysis:** Note that

\[
E[A_n - B_n] = p_A - p_B \quad \text{and} \quad \text{var}(A_n - B_n) = \frac{1}{n}(p_A(1-p_A)+p_B(1-p_B)) \leq \frac{1}{2n}.
\]

Thus, \( Pr[|A_n - B_n - (p_A - p_B)| > \epsilon] \leq \frac{1}{2n\epsilon^2}, \)
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:** Note that

$$E[A_n - B_n] = p_A - p_B$$

and

$$\text{var}(A_n - B_n) = \frac{1}{n}(p_A(1-p_A)+p_B(1-p_B)) \leq \frac{1}{2n}.$$ 

Thus,

$$Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n\varepsilon^2},$$

so

$$Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n\varepsilon^2},$$
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:** Note that

$$E[A_n - B_n] = p_A - p_B \text{ and } var(A_n - B_n) = \frac{1}{n} (p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}.$$  

Thus, $Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n\varepsilon^2}$, so

$$Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n\varepsilon^2}, \text{ and}$$

$$Pr[p_A - p_B \geq 0] \geq 1 - \frac{1}{2n(A_n - B_n)^2}.$$  

**Example:**
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:** Note that

$$E[A_n - B_n] = p_A - p_B$$

and

$$\text{var}(A_n - B_n) = \frac{1}{n} (p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}.$$

Thus,

$$Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n\varepsilon^2},$$

so

$$Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n\varepsilon^2},$$

and

$$Pr[p_A - p_B \geq 0] \geq 1 - \frac{1}{2n(A_n - B_n)^2}.$$

**Example:** With $n = 100$ and $A_n - B_n = 0.2$,
Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$.

**Confidence?**

**Analysis:** Note that

$$E[A_n - B_n] = p_A - p_B \text{ and } var(A_n - B_n) = \frac{1}{n}(p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}.$$  

Thus, $Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n\varepsilon^2}$, so

$$Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n\varepsilon^2}, \text{ and}$$

$$Pr[p_A - p_B \geq 0] \geq 1 - \frac{1}{2n(A_n - B_n)^2}.$$  

**Example:** With $n = 100$ and $A_n - B_n = 0.2$, $Pr[p_A > p_B] \geq 1 - \frac{1}{8} = 0.875$. 
For $B(p)$, we wanted to estimate $p$. The CI requires

$$\sigma = \sqrt{p(1-p)}.$$  
We replaced $\sigma$ by an upper bound: $1/2$. In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - \bar{X}_n)^2.$$  
However, in some cases, this is dangerous! The theory says it is OK if the distribution of $X_n$ is nice (Gaussian). This is used regularly in practice. However, be aware of the risk.
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. 

The CI requires $\sigma = \sqrt{p(1-p)}$.

We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance: $s^2 = \frac{1}{n} \sum_{m=1}^{n} (X_m - \bar{X}_n)^2$.

However, in some cases, this is dangerous! The theory says it is OK if the distribution of $X_n$ is nice (Gaussian). This is used regularly in practice. However, be aware of the risk.
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. 
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$. 
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - A_n)^2.$$
For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - A_n)^2.$$

However, in some cases, this is dangerous!
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - A_n)^2.$$ 

However, in some cases, this is dangerous! The theory says it is OK if the distribution of $X_n$ is nice (Gaussian).
Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - A_n)^2.$$ 

However, in some cases, this is dangerous! The theory says it is OK if the distribution of $X_n$ is nice (Gaussian). This is used regularly in practice.
**Unknown $\sigma$**

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - A_n)^2.$$ 

However, in some cases, this is dangerous! The theory says it is OK if the distribution of $X_n$ is nice (Gaussian). This is used regularly in practice. However, be aware of the risk.
Summary

Confidence Intervals

1. Estimates without confidence level are useless!
2. $[a, b]$ is a 95\% CI for $\theta$ if $\Pr[\theta \in [a, b]] \geq 95\%$.
3. Using Chebyshev: $[A_n - 4 .5 \sigma / \sqrt{n}, A_n + 4 .5 \sigma / \sqrt{n}]$ is a 95\% CI for $\mu$.
4. Using CLT, we will replace 4 .5 by 2.
5. When $\sigma$ is not known, one can replace it by an upper bound.
6. Examples: $B(p)$, $G(p)$, which coin is better?
7. In some cases, one can replace $\sigma$ by the empirical standard deviation.
Summary

Confidence Intervals

1. Estimates without confidence level are useless!
1. Estimates without confidence level are useless!
2. \([a, b]\) is a 95\%-CI for \(\theta\) if
Confidence Intervals

1. Estimates without confidence level are useless!
2. \([a, b]\) is a 95\%-CI for \(\theta\) if \(Pr[\theta \in [a, b]] \geq 95\%\).

Using Chebyshev:

\([A_n - 4 \frac{\sigma}{\sqrt{n}}, A_n + 4 \frac{\sigma}{\sqrt{n}}]\) is a 95\%-CI for \(\mu\).

Using CLT, we will replace 4\(\frac{\sigma}{\sqrt{n}}\) by 2.

5. When \(\sigma\) is not known, one can replace it by an upper bound.

6. Examples: \(B(p)\), \(G(p)\), which coin is better?

7. In some cases, one can replace \(\sigma\) by the empirical standard deviation.
Summary

1. Estimates without confidence level are useless!
2. \([a, b]\) is a 95\%-CI for \(\theta\) if \(Pr[\theta \in [a, b]] \geq 95\%\).
3. Using Chebyshev: 
   \([A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]\) is a 95\%-CI for \(\mu\).
1. Estimates without confidence level are useless!

2. \([a, b]\) is a 95%-CI for \(\theta\) if \(Pr[\theta \in [a, b]] \geq 95\%\).

3. Using Chebyshev: \([A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]\) is a 95%-CI for \(\mu\).

4. Using CLT, we will replace 4.5 by 2.
Summary

Confidence Intervals

1. Estimates without confidence level are useless!

2. \([a, b]\) is a 95%-CI for \(\theta\) if \(Pr[\theta \in [a, b]] \geq 95\%\).

3. Using Chebyshev: \([A_n - 4.5\sigma / \sqrt{n}, A_n + 4.5\sigma / \sqrt{n}]\) is a 95%-CI for \(\mu\).

4. Using CLT, we will replace 4.5 by 2.

5. When \(\sigma\) is not known, one can replace it by an upper bound.
Summary

Confidence Intervals

1. Estimates without confidence level are useless!

2. \([a, b]\) is a 95\%-CI for \(\theta\) if \(Pr[\theta \in [a, b]] \geq 95\%\).

3. Using Chebyshev: \([A_n - 4.5\sigma / \sqrt{n}, A_n + 4.5\sigma / \sqrt{n}]\) is a 95\%-CI for \(\mu\).

4. Using CLT, we will replace 4.5 by 2.

5. When \(\sigma\) is not known, one can replace it by an upper bound.

6. Examples: \(B(p), G(p)\), which coin is better?

7. In some cases, one can replace \(\sigma\) by the empirical standard deviation.