CS70 @ UC Berkeley, Spring 2026

Lecture 23

Concentration Inequalities and LLN

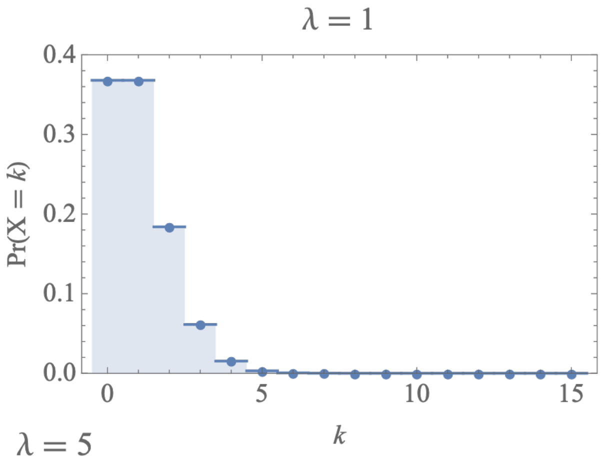
April 16, 2026

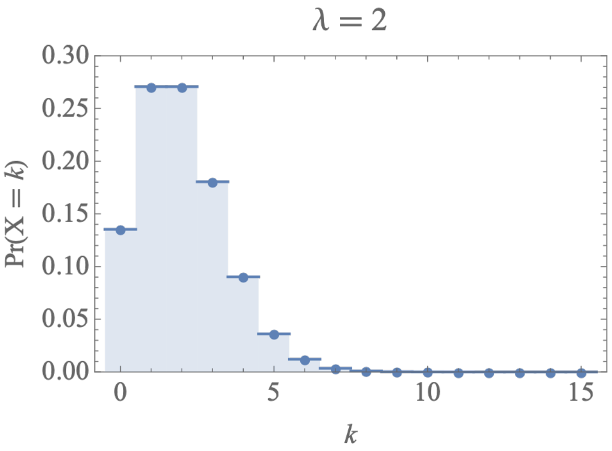
Poisson Distribution (Lecture 22)

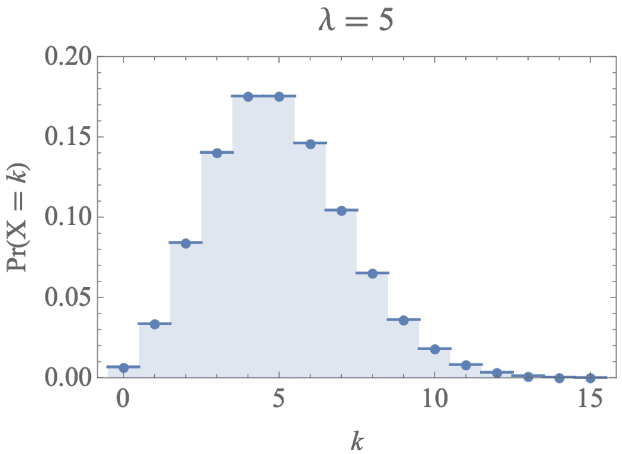
$N \sim \text{Poisson}(\lambda)$, where intensity $\lambda > 0$.

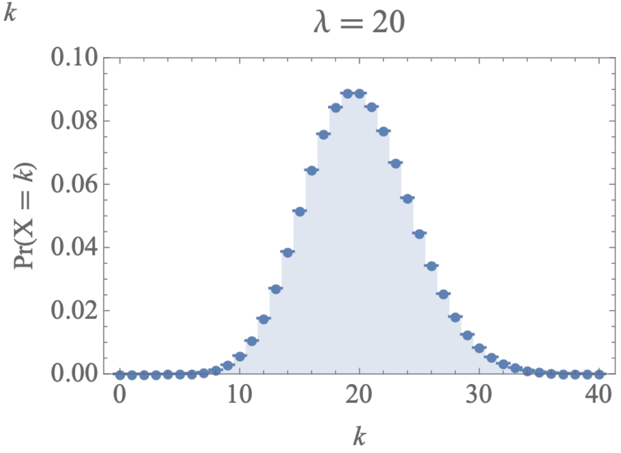
$\mathbb{P}(N = k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$, for $k \in \mathbb{N}$.

- # rain drops hitting a surface per second
- # radioactive particles emitted by radio active material during an interval of time









Looks more and more like a "bell curve" as λ gets large. We will see why this happens.

Poisson Approximation of Binomial(n, p) (Lecture 22)

This corresponds to a continuous time limit

- $(*)$ Limit as $n \to \infty$ and $p \to 0$, while $np = \lambda$ is held fixed.
- Define $Q_k = \binom{n}{k} p^k (1-p)^{n-k}$
- $Q_0 = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$ in the limit $(*)$.
- $\frac{Q_k}{Q_{k-1}} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{n-k+1}{k}\left(\frac{p}{1-p}\right) = \frac{np}{k}\left[\frac{1-\frac{k-1}{n}}{1-p}\right] \to \frac{\lambda}{k}$ in the limit $(*)$.
- $Q_k = Q_0 \frac{Q_1}{Q_0}\frac{Q_2}{Q_1}\cdots\frac{Q_k}{Q_{k-1}} \to e^{-\lambda}\frac{\lambda}{1}\frac{\lambda}{2}\cdots\frac{\lambda}{k} = e^{-\lambda}\frac{\lambda^k}{k!}$ in the limit $(*)$.

$X \sim \text{Binomial}(n, p)$

$\mathbb{E}[X] = np \to \lambda$ in the limit $(*)$.

$\text{Var}(X) = np(1-p) \to \lambda$ in the limit $(*)$.

Independent Poisson Random Variables

Theorem: Suppose $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are independent random variables. Then, $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Proof:

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n - k)$$

By additivity.
$(X = 0) \cap (Y = n), \dots, (X = n) \cap (Y = 0)$
partition $X + Y = n$

$$= \sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n - k)$$

By Independence

$$= \sum_{k=0}^{n} \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{e^{-\mu}\mu^{n-k}}{(n-k)!}$$

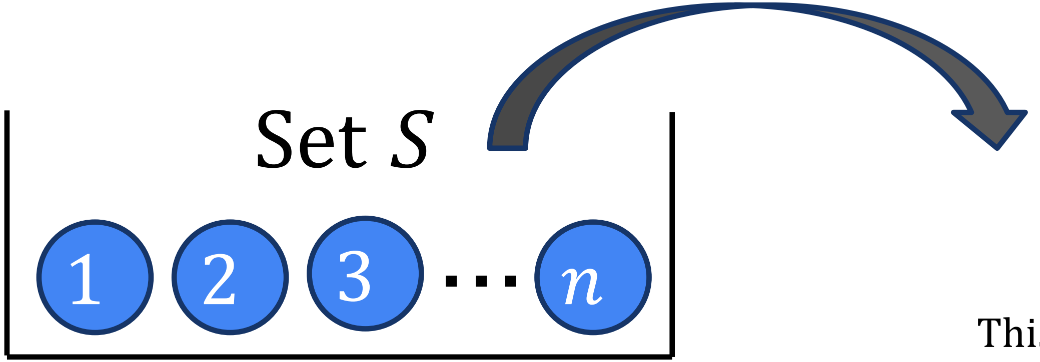
$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \lambda^k \mu^{n-k}$$

Binomial Theorem (Lecture 15)

$$= \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^n}{n!} \qquad \square$$

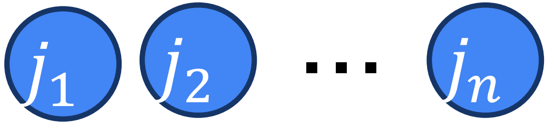
Corollary: Suppose $X_1, \dots, X_n$ are independent RVs with $X_i \sim \text{Poisson}(\lambda_i)$, for $i = 1, \dots, n$. Then, $X_1 + \dots + X_n \sim \text{Poisson}(\lambda_1 + \dots + \lambda_n)$.

Permutations and Derangements (Lecture 15)



Sample without replacement

n-element ordered object



This generates a permutation of $\{1,2,\dots n\}$.

Total number of permutations $=$ $n!$

Iteration i:	1	2	$\boxed{3}$	4	$\boxed{5}$	…	n
Sampled Label π_i:	7	1	$\boxed{3}$	8	$\boxed{5}$	…	2

Fixed points

Bijection $\pi: S \longrightarrow S$

$\pi: i \longmapsto \pi_i$

Definition: A derangement is a permutation with no fixed points.

Let D_n denote the number of derangements of $\{1,2,\dots,n\}$.

Number of Fixed Points in a Random Permutation

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \qquad \lim_{n\to\infty} \frac{D_n}{n!} = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{(-1)^k}{k!} = e^{-1}$$

- Define $F_{k,n} = \#$ permutations of $\{1,2,\ldots,n\}$ with exactly k fixed points.
- $F_{0,n} = D_n$
- $F_{k,n} = \binom{n}{k} D_{n-k}$
- $\lim_{n\to\infty} \frac{F_{k,n}}{n!} = \lim_{n\to\infty} \frac{\binom{n}{k}}{n!} (n-k)! \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}$

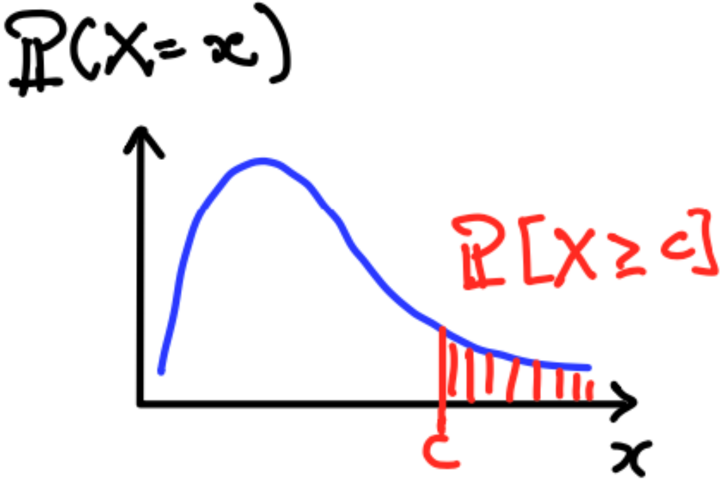
Proportion of permutations with k fixed points

$$= \lim_{n\to\infty} \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} = \frac{1}{k!} e^{-1}$$

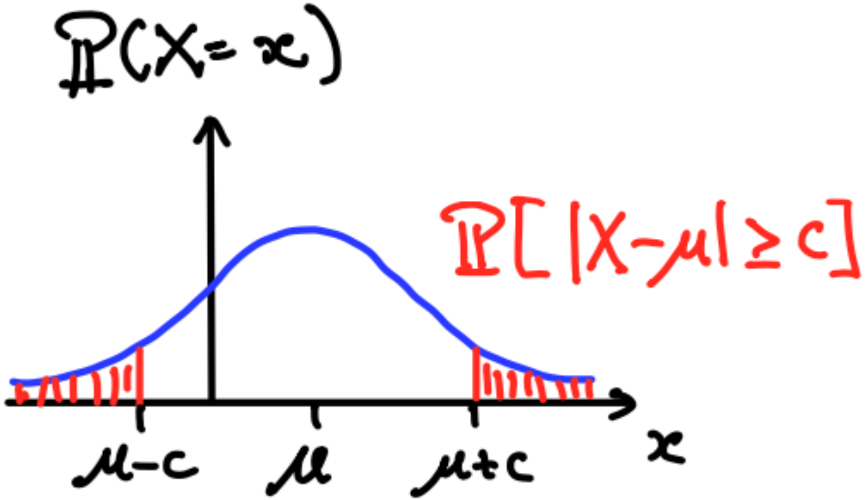
So, in the limit $n \to \infty$, #fixed points $\sim$ Poisson(1)

Concentration Inequalities

Bound tail probabilities



Quantify deviations from the mean (how tightly values cluster around it)



What do $\mathbb{E}[X]$ and $\mathrm{Var}[X]$ tell us about these tail probabilities?

Concentration Inequalities

- Provide high-probability guarantees. (e.g., $\mathbb{P}(\text{bad event}) \leq \varepsilon$)
- Machine learning & statistics
  - Generalization error bounds
- Randomized algorithms
  - Guarantee that randomized algorithms work with extremely high probability
- Signal processing and compressed sensing
  - Prove that certain random matrices satisfy the Restricted Isometry Property (RIP).
  - This property ensures that a sparse signal can be perfectly reconstructed from far fewer samples than the Nyquist-Shannon theorem would traditionally suggest.
- Non-asymptotic analysis
  - Bounds for finite sample sizes
- As the sample size increases, concentration often improves

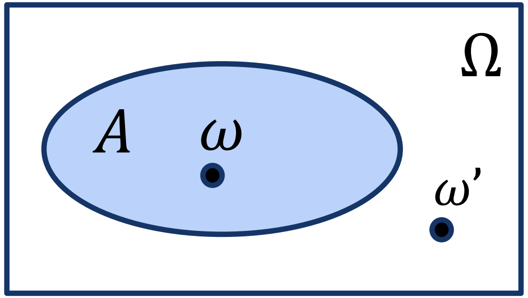
Indicator Random Variable

The **Indicator** RV for an event A is defined as

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

$$(I_A = 1) = \{\omega \in \Omega |\ I_A(\omega) = 1\} = A$$

So, $\mathbb{E}[I_A] = \mathbb{P}(I_A = 1) = \mathbb{P}(A)$.



Lemma. Let X be a non-negative random variable. Then, for all $\omega \in \Omega$ and for all constant $c > 0$,

$$X(\omega) \geq c I_{\{X \geq c\}}(\omega). \quad (*)$$

Proof: First note that the event $(X \geq c) = \{\omega \in \Omega |\ X(\omega) \geq c\}$.

- $X(\omega) < c \Rightarrow I_{\{X \geq c\}}(\omega) = 0$
  X non-negative $\Rightarrow X(\omega) \geq 0$ for all $\omega \in \Omega$ — $(*)$ holds
- $X(\omega) \geq c \Rightarrow I_{\{X \geq c\}}(\omega) = 1$ — $(*)$ holds again $\square$

Markov's Inequality

Lemma. Let X be a non-negative random variable. Then, for all $\omega \in \Omega$ and for all constant $c > 0$, $X(\omega) \geq cI_{\{X \geq c\}}(\omega)$.

Theorem (Markov' Inequality). Let X be a non-negative random variable with $\mathbb{E}[|X|] < \infty$. Then, for all constant c > 0,

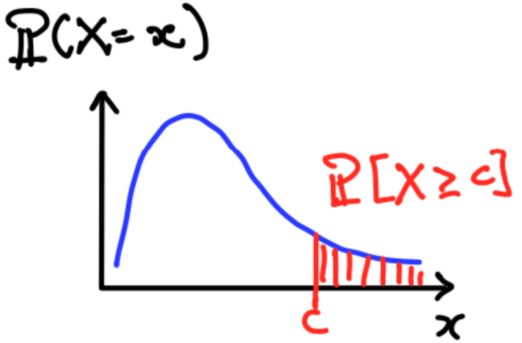
$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}.$$

Proof: Lemma $\Rightarrow \forall \omega \in \Omega$ and $c > 0$, $X(\omega) \geq cI_{\{X \geq c\}}(\omega)$.
Taking the expectation of both sides gives

$$\mathbb{E}[X] \geq \mathbb{E}\left[cI_{\{X \geq c\}}\right]$$

$$= c\mathbb{E}\left[I_{\{X \geq c\}}\right] = c\mathbb{P}(X \geq c)$$

Why doesn't the inequality direction change when we take the expectation?



Example

Theorem (Markov' Inequality). Let X be a non-negative random variable with $\mathbb{E}[|X|] < \infty$. Then, for all constant c > 0,

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}.$$

- Suppose $X_1, \ldots, X_n$ are i.i.d. Uniform$\{0,1,\ldots,9\}$ random variables.
- Let $S_n = X_1 + \cdots + X_n$
- Upper bound for $\mathbb{P}(S_n \geq 8n)$?
- $\mathbb{E}[S_n] = n\mathbb{E}[X_1] = n \sum_{i=0}^{9} \frac{i}{10} = n\frac{9}{2}$.
- So, Markov's Inequality implies $\mathbb{P}(S_n \geq 8n) \leq \frac{\mathbb{E}[S_n]}{8n} = \frac{9}{16}$.

Does not depend on n. We will see that this is not a good upper bound.

Generalized Markov's Inequality

Theorem (Generalized Markov' Inequality). Let X be an arbitrary random variable. Then, for all constant c > 0,

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[|X|^k]}{c^k}.$$

Proof: Apply similar arguments as in the proof of Markov's Inequality to $|X(\omega)|^k \geq c^k \, I_{\{X \geq c\}}(\omega)$.

Chebyshev's Inequality

Theorem (Chebyshev's Inequality). For all random variables X with $\mathbb{E}[X] = \mu < \infty$ and for all constants $c > 0$,

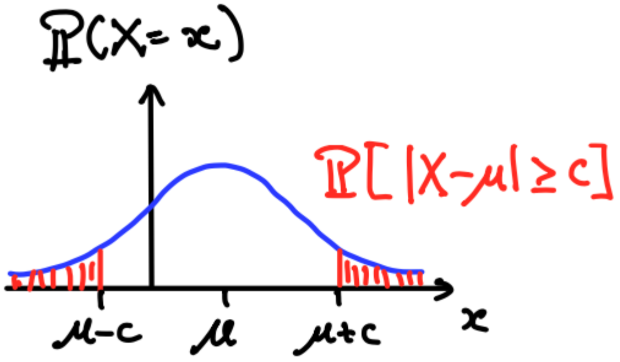
$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\mathrm{Var[X]}}{c^2}.$$

Proof: Since $|X - \mu| \geq c \Leftrightarrow |X - \mu|^2 \geq c^2$

$$\mathbb{P}[|X - \mu| \geq c] = \mathbb{P}[|X - \mu|^2 \geq c^2]$$

$$\leq \frac{\mathbb{E}[|X - \mu|^2]}{c^2} = \frac{\mathrm{Var}(X\}}{c^2} \quad \square$$

Markov's inequality



Also follows from the Generalized Markov's Inequality for $k = 2$.

Example

Theorem (Chebyshev's Inequality). For all random variables X with $\mathbb{E}[X] = \mu < \infty$ and for all constants $c > 0$,

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

- Suppose $X_1, \dots, X_n$ are i.i.d. Uniform$\{0,1,\dots,9\}$ random variables.
- Let $S_n = X_1 + \cdots + X_n$
- Upper bound for $\mathbb{P}(S_n \geq 8n)$?
- $\mathbb{E}[S_n] = n\mathbb{E}[X_1] = n \sum_{i=0}^{9} \frac{i}{10} = n\frac{9}{2}$.
- $\mathrm{Var}(S_n) = n\mathrm{Var}(X_1) = n\frac{33}{4}$.
- $\mathbb{P}(S_n \geq 8n) = \mathbb{P}\left(S_n - \frac{9n}{2} \geq \frac{7n}{2}\right) \leq \mathbb{P}\left(\left|S_n - \frac{9n}{2}\right| \geq \frac{7n}{2}\right).$

$$\leq \frac{33n/4}{(7n/2)^2} = \frac{33}{49}\frac{1}{n}$$

Vanishes as $n \to \infty$. Much better than the bound from Markov's Inequality.

Estimating the Mean of a Distribution

- Suppose $X_1, \ldots, X_n$ are i.i.d. random variables from some distribution.
- $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$ for all $i = 1, \ldots n$.
- Suppose μ is unknown, whereas σ^2 is known.
- **Goal:** Estimate μ.
- Let $S_n = X_1 + \cdots + X_n$. Then, $\mathbb{E}\left[\frac{S_n}{n}\right] = \mu$
- **Estimator** $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n}$
- **What sample size n should one use to achieve a desired error control with confidence $1 - \delta$?**
- $\mathrm{Var}(\hat{\mu}_n) = \frac{1}{n^2}(\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)) = \frac{\sigma^2}{n}$
- By Chebyshev's Inequality, $\mathbb{P}(|\hat{\mu}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\hat{\mu}_n)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n}$
- To guarantee $\mathbb{P}(|\hat{\mu}_n - \mu| \geq \varepsilon) \leq \delta$, sufficient to find n such that $\frac{\sigma^2}{\varepsilon^2 n} \leq \delta$

$$\mathbb{P}(|\hat{\mu}_n - \mu| < \varepsilon) \geq 1 - \delta$$ (this is called confidence)

$$\boxed{n \geq \frac{\sigma^2}{\varepsilon^2 \delta}}$$

Estimating the Mean of a Distribution

What if σ^2 is also unknown?

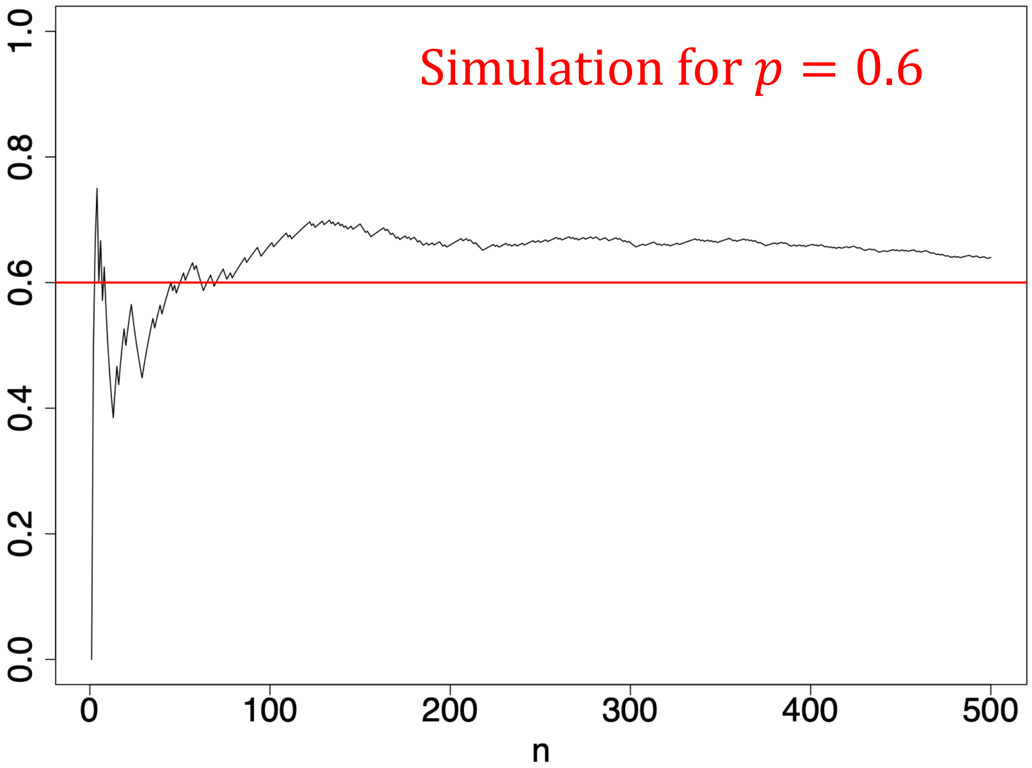
- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p) random variables
- $\mathbb{E}[X_i] = p$ and $\mathrm{Var}[X_i] = p(1-p)$ for all $i = 1, \ldots n$.
- Let $S_n = X_1 + \cdots + X_n$.
- $\mathbb{E}[S_n] = np$, so $\mathbb{E}\left[\frac{S_n}{n}\right] = p$.
- **Goal:** Estimate p.
- **Estimator** $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n}$
- By Chebyshev's Inequality, $\mathbb{P}(|\hat{p}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n}$.
- To guarantee $\mathbb{P}(|\hat{\mu}_n - \mu| \geq \varepsilon) \leq \delta$, choose

$$\boxed{n \geq \frac{1}{4\varepsilon^2 \delta}}$$

Sample Estimate of p

- Let $X_1, \dots, X_n$ be i.i.d. Bernoulli(p) random variables
- $\mathbb{E}[X_i] = p$ for all $i = 1, \dots n$
- Let $S_n = X_1 + \cdots + X_n$.
- $\mathbb{E}[S_n] = np$, so $\mathbb{E}\left[\frac{S_n}{n}\right] = p$.

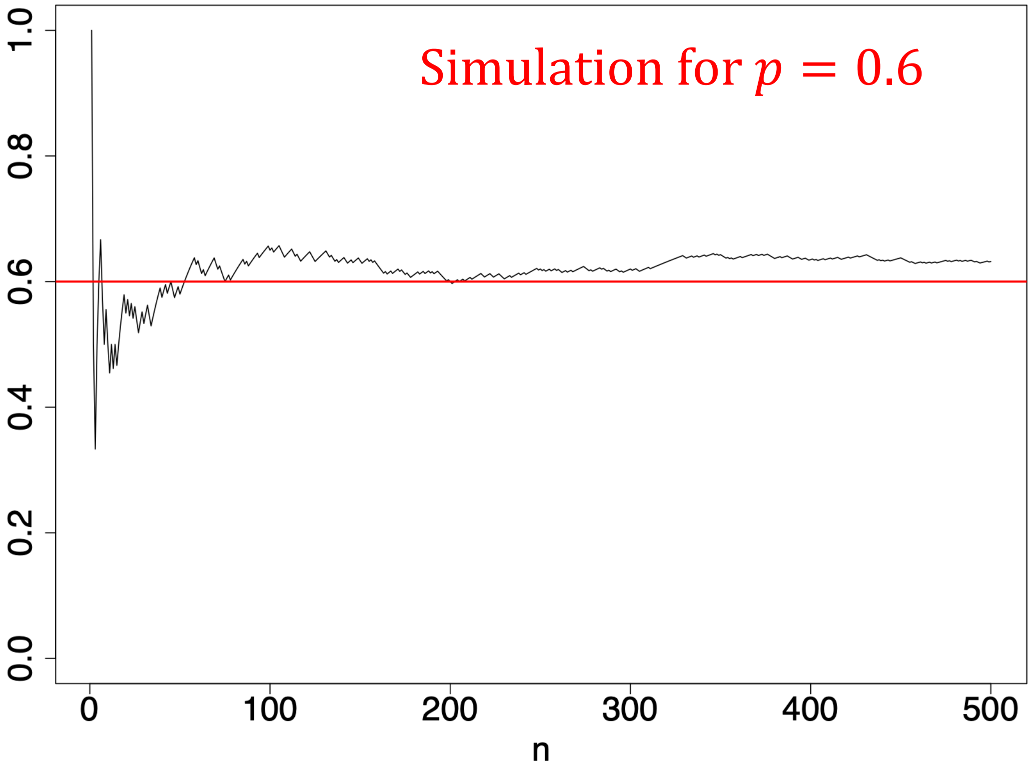
Estimator $\hat{p}_n = \frac{S_n}{n}$



Sample Estimate of p

- Let $X_1, \dots, X_n$ be i.i.d. Bernoulli(p) random variables
- $\mathbb{E}[X_i] = p$ for all $i = 1, \dots n$
- Let $S_n = X_1 + \cdots + X_n$.
- $\mathbb{E}[S_n] = np$, so $\mathbb{E}\left[\frac{S_n}{n}\right] = p$.

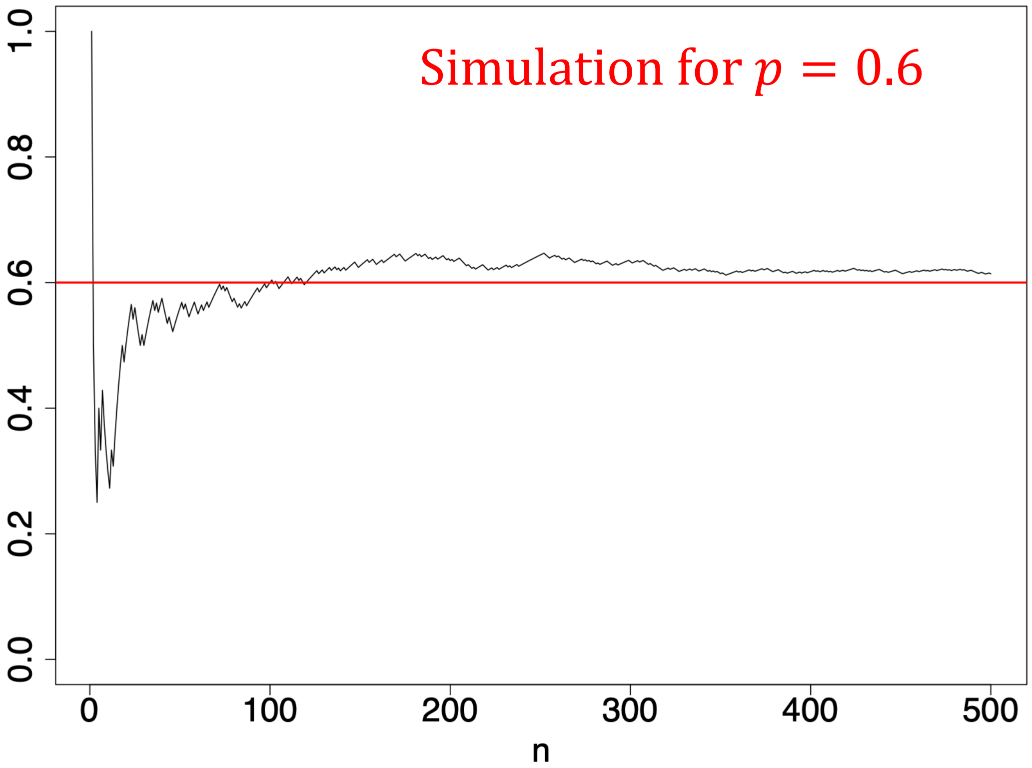
Estimator $\hat{p}_n = \frac{S_n}{n}$



Sample Estimate of p

- Let $X_1, \dots, X_n$ be i.i.d. Bernoulli(p) random variables
- $\mathbb{E}[X_i] = p$ for all $i = 1, \dots n$
- Let $S_n = X_1 + \cdots + X_n$.
- $\mathbb{E}[S_n] = np$, so $\mathbb{E}\left[\frac{S_n}{n}\right] = p$.

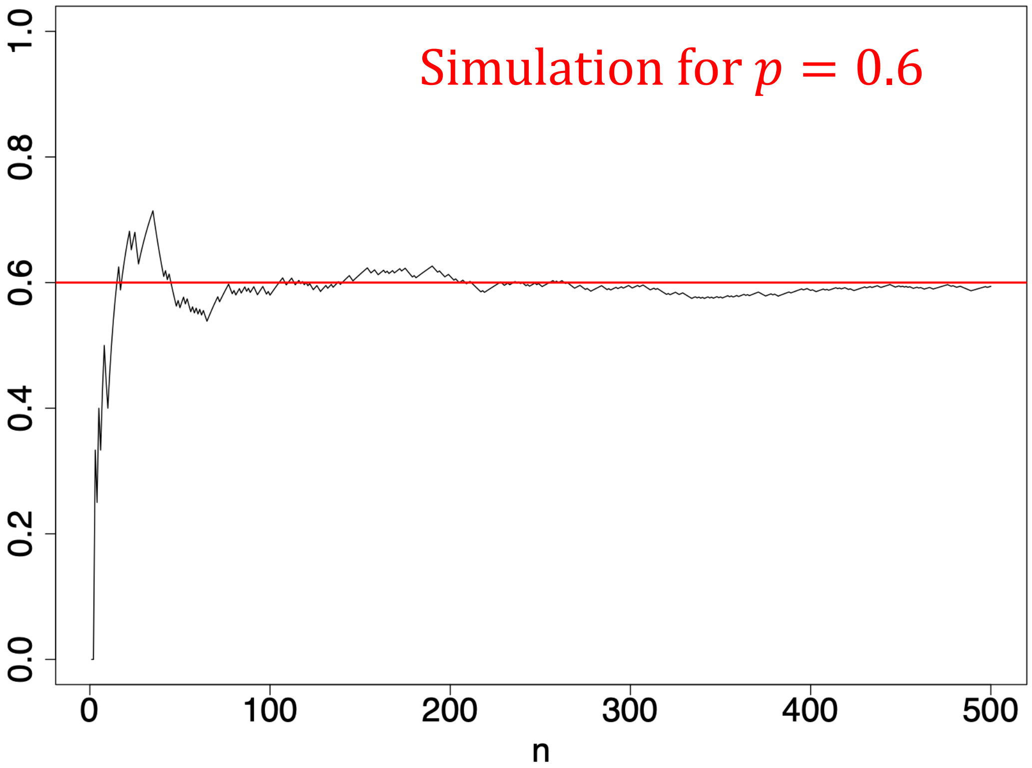
Estimator $\hat{p}_n = \frac{S_n}{n}$



Sample Estimate of p

- Let $X_1, \dots, X_n$ be i.i.d. Bernoulli(p) random variables
- $\mathbb{E}[X_i] = p$ for all $i = 1, \dots n$
- Let $S_n = X_1 + \cdots + X_n$.
- $\mathbb{E}[S_n] = np$, so $\mathbb{E}\left[\frac{S_n}{n}\right] = p$.

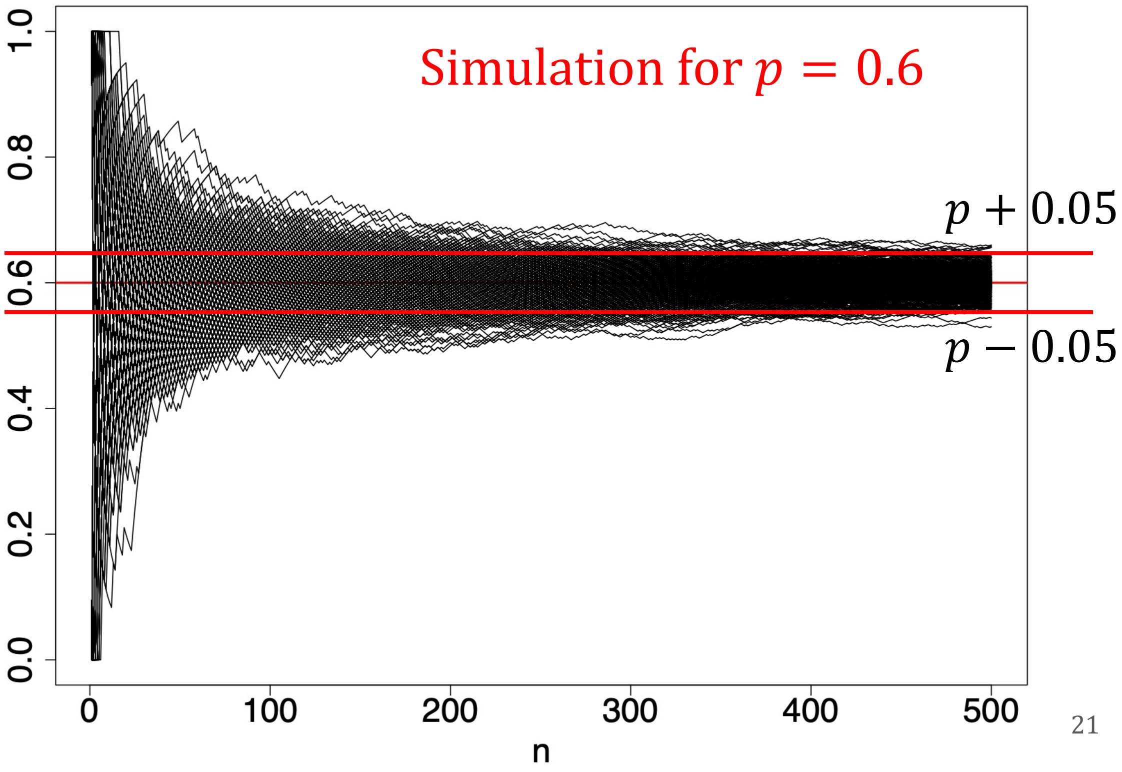
Estimator $\hat{p}_n = \frac{S_n}{n}$

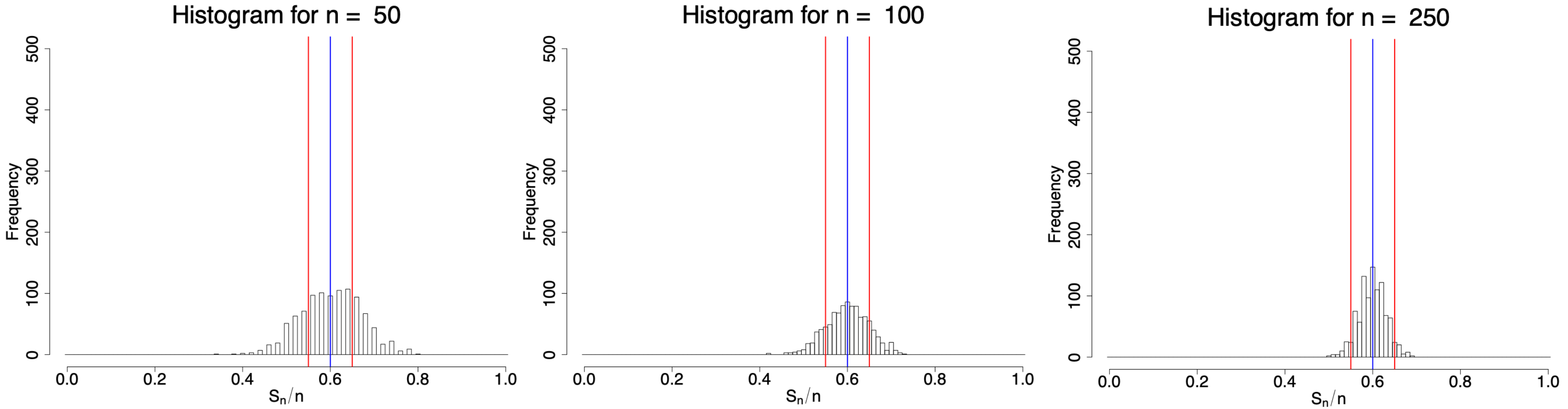


Sample Estimate of p

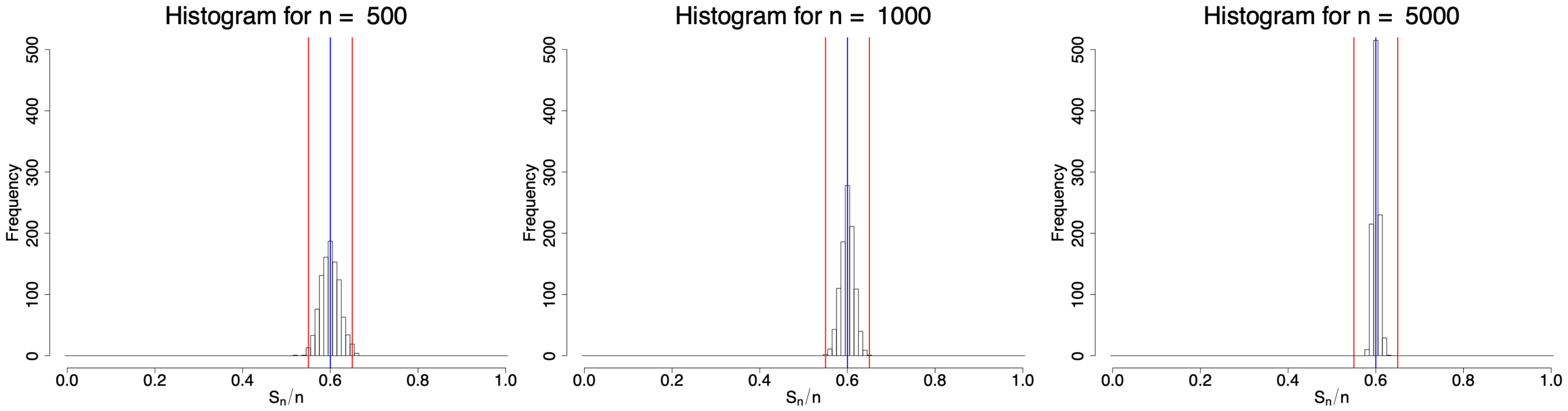
- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p) random variables
- $\mathbb{E}[X_i] = p$ for all $i = 1, \ldots n$
- Let $S_n = X_1 + \cdots + X_n$.
- $\mathbb{E}[S_n] = np$, so $\mathbb{E}\left[\frac{S_n}{n}\right] = p$.

Estimator $\hat{p}_n = \frac{S_n}{n}$





Red lines correspond to $y = 0.6 \pm 0.05$



Weak Law of Large Numbers (WLLN)

Theorem (Weak LLN). Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with a finite mean μ and finite variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

Proof: $\mathbb{E}\left[\frac{S_n}{n}\right] = \mu$ and $\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{\operatorname{Var}(S_n)}{n^2} = \frac{\operatorname{Var}(X_1)+\cdots+\operatorname{Var}(X_n)}{n^2} = \frac{\sigma^2}{n}$

By Independence

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \le \frac{\operatorname{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} \le \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ as } n \to \infty.$$

By Chebyshev's Inequality

$\square$

Remark: In fact, finite variance is NOT required for WLLN, although this proof needs it.