STAT 24400 Lecture 12
Section 5.2 The Law of Large Numbers (LLN)
Section 5.3 Convergence in Distribution and the Central Limit Theorem

Yibi Huang
Department of Statistics
University of Chicago

Markov Inequality (p.121 in Textbook)

If X is a random variable that only take nonnegative values $\mathrm{P}(X \geq 0) = 1$ and for which $\mathsf{E}(X)$ exists, then

$$\mathrm{P}(X \geq t) \leq \frac{\mathsf{E}(X)}{t}.$$

Proof. We will prove this for the discrete case; the continuous case is entirely analogous.

$$\begin{aligned}
\mathsf{E}(X) = \sum_x x p(x) = \overbrace{\sum_{x<t} x p(x)}^{\geq 0 \text{ since } X \geq 0} + \sum_{x \geq t} x p(x) \\
\geq \sum_{x \geq t} x p(x) \\
\geq \sum_{x \geq t} t p(x) = t\mathrm{P}(X \geq t)
\end{aligned}$$

Chebyshev's Inequality (p.133, Textbook)

Let X be a random variable with mean μ and variance σ^2. Then, for any $t > 0$,

$$\mathrm{P}(|X - \mu| > k) \leq \frac{\sigma^2}{k^2}.$$

Proof. Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $t = k^2$) to obtain

$$\mathrm{P}(|X - \mu| \geq k) = \mathrm{P}((X - \mu)^2 \geq k^2) \leq \frac{\overbrace{\mathsf{E}[(X - \mu)^2]}^{=\mathsf{Var}(X)=\sigma^2}}{k^2} = \frac{\sigma^2}{k^2}.$$

The Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \ldots, X_n, \ldots$ be indep. r.v.s with $\mathsf{E}(X_i) = \mu$ and $\mathsf{Var}(X_i) = \sigma^2$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then, for any $\varepsilon > 0$,

$$\mathrm{P}(|\overline{X}_n - \mu| > \varepsilon) \to 0 \quad \text{as } n \to \infty.$$

Proof. We first find $\mathsf{E}(\overline{X}_n)$ and $\mathsf{Var}(\overline{X}_n)$:

$$\mathsf{E}(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^n \mathsf{E}(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \mu.$$

Since the X_i are independent,

$$\mathsf{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^n \underbrace{\mathsf{Var}(X_i)}_{=\sigma^2} = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

The desired result now follows immediately from Chebyshev's inequality, which

$$\mathrm{P}(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\mathsf{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad \text{as } n \to \infty.$$

Counterexample for WLLN — Cauchy

- The WLLN above is proved assuming the existence of $\mathrm{Var}(X_i)$. WLLN can be proved only assuming the existence of $\mathrm{E}(X_i)$.
- If $\mathrm{E}(X_i)$ doesn't exist, WLLN might not work. A counterexample the Cauchy distribution. If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. Cauchy, we can show using the Characteristic function[1] in the next page that

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \text{Cauchy}$$

which doesn't converge to a single value and this implies

$$\mathrm{P}(|\overline{X}_n| > \varepsilon) = 2 \int_{\varepsilon}^{\infty} \frac{1}{\pi(1+x^2)} \mathrm{d}x = 2 \left[\frac{\arctan(x)}{\pi} \right]_{x=\varepsilon}^{x=\infty} = 1 - \frac{2\arctan(\varepsilon)}{\pi}$$

For example, $\mathrm{P}(|\overline{X}_n| > 1) = 1 - \frac{2\arctan(1)}{\pi} = \frac{1}{2}$ for all n, which doesn't converge to 0 as $n \to \infty$.

[1] In fact, we have proved a special case in L07 that if X_1 and X_2 are indep. Cauchy, then $\frac{1}{2}(X_1 + X_2)$ also has the Cauchy distribution

Distribution of Sample Mean of Cauchy R.V.'s

Recall we mentioned at the end of L11 that the **characteristic function** of the Cauchy distribution is

$$\phi_X(t) = \mathrm{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi(1+x^2)} \mathrm{d}x = e^{-|t|}, \quad -\infty < t < \infty.$$

If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. Cauchy, then the characteristic function for $\sum_{i=1}^n X_i$ is

$$\phi_{\sum_{i=1}^n X_i}(t) = [\phi_X(t)]^n = e^{-n|t|}, \quad -\infty < t < \infty.$$

Thus, the characteristic function for $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is

$$\phi_{\overline{X}_n}(t) = \phi_{\sum_{i=1}^n X_i}(t/n) = e^{-n|t/n|} = e^{-|t|}, \quad -\infty < t < \infty,$$

which is exactly the characteristic function for Cauchy. As the characteristic function uniquely determines the distribution, we know $\overline{X}_n$ is has the Cauchy PDF below for all n:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty \le x < \infty.$$

Definition: Convergence in Distribution (p.181, Textbook)

Let $X_1, X_2, \ldots$ be a sequence of r.v.s with CDFs $F_1, F_2, \ldots$, and let X be a r.v. with CDF F. We say that *X_n converges in distribution to X* if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

at every point at which F is continuous.

Convergence in MGF Implies Convergence in Distribution

Suppose $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of r.v.s, each with MGF $M_{X_i}(t)$. Furthermore, suppose that

$$\lim_{i \to \infty} M_{X_i}(t) = M_X(t),$$

for all t in an open interval containing 0, and $M_X(t)$ is an MGF. Then there is a unique CDF F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{n \to \infty} F_n(x) = F(x).$$

Proof. Too advance for STAT 244.

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, each having mean μ and variance σ^2 and let $S_n = X_1 + \cdots + X_n$. The distribution of

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$\mathrm{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq a\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} \mathrm{d}x \quad \text{as } n \to \infty.$$

Proof of CLT (Using MGF)

We will first prove CLT for the case $\mu = \mathrm{E}(X_i) = 0$.

Let $M(t)$ be the common MGF of the X_i's. Since S_n is a sum of independent r.v.'s, we know the MGF of S_n is

$$M_{S_n}(t) = [M(t)]^n$$

The MGF for $Z_n = \frac{S_n}{\sqrt{n}\sigma}$ is a linear transformation of S_n, so

$$M_{Z_n}(t) = M_{S_n}\left(\frac{t}{\sqrt{n}\sigma}\right) = \left[M\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n$$

Take the Taylor series expansion of $M(s)$ about zero:

$$\begin{aligned} M(s) &= \overbrace{M(0)}^{=1} + s \overbrace{M'(0)}^{=\mathrm{E}(X)=0} + \frac{1}{2}s^2 \overbrace{M''(0)}^{=\mathrm{E}[X^2]=\mathrm{Var}(X)=\sigma^2} + \varepsilon \\ &= 1 + \frac{\sigma^2}{2}s^2 + \varepsilon \end{aligned}$$

where $\varepsilon / s^2 \to 0$ as $s \to 0$.

As $M(s) = 1 + \frac{\sigma^2}{2}s^2 + \varepsilon$, we have

$$M\left(\frac{t}{\sqrt{n}\sigma}\right) = 1 + \frac{\sigma^2}{2}\left(\frac{t}{\sqrt{n}\sigma}\right)^2 + \varepsilon_n = 1 + \frac{t^2}{2n} + \varepsilon_n$$

where $\varepsilon_n/(t^2/(n\sigma^2)) \to 0$ as $n \to \infty$.

$$M_{Z_n}(t) = \left[M\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n = \left(1 + \frac{t^2}{2n} + \varepsilon_n\right)^n \to e^{t^2/2} \quad \text{as } n \to \infty.$$

The last limit comes from the fact that

$$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a \quad \text{if} \quad \lim_{n\to\infty} a_n = a.$$

Here $e^{t^2/2}$ is the MGF of the standard normal, as was to be shown.

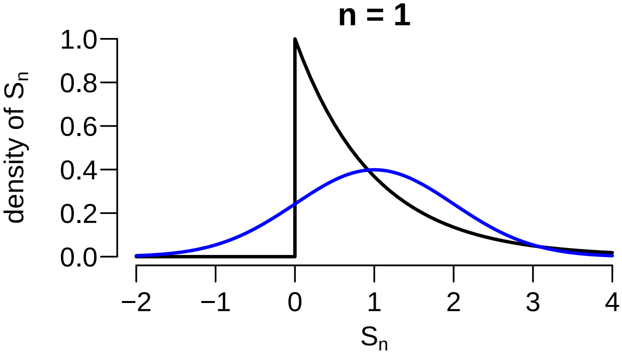
For the case $\mu = \mathrm{E}(X_i) \neq 0$, we can define $X_i' = X_i - \mu$, and let $S_n' = X_1' + \cdots + X_n'$. Then $S_n - n\mu = S_n'$ and the proof goes as the case for $\mu = 0$.

Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is $\text{Gamma}(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

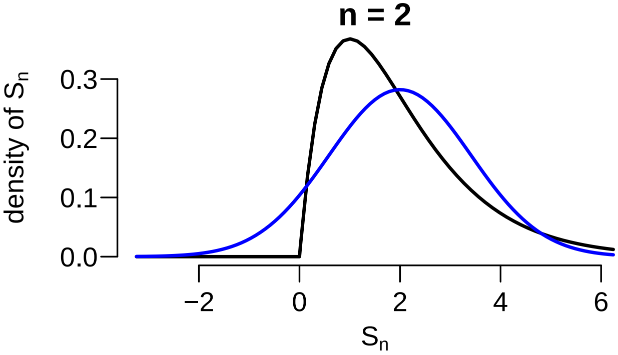


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is Gamma$(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

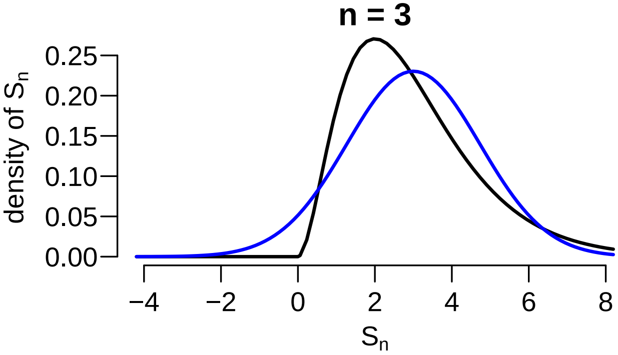


Example of CLT — Exponential

If $X_i \sim$ Exponential$(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is Gamma$(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

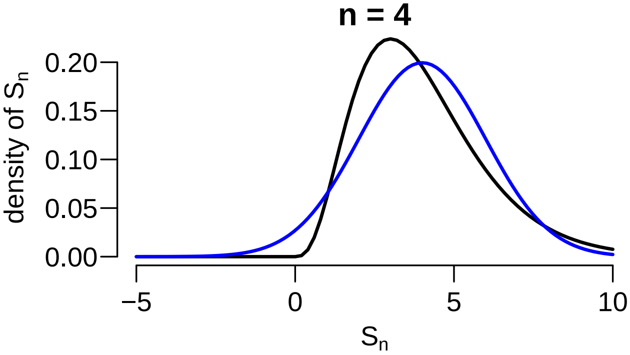


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is $\text{Gamma}(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

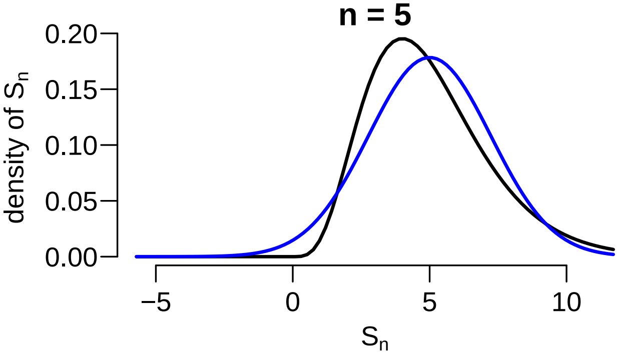


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^n X_i$ is $\text{Gamma}(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

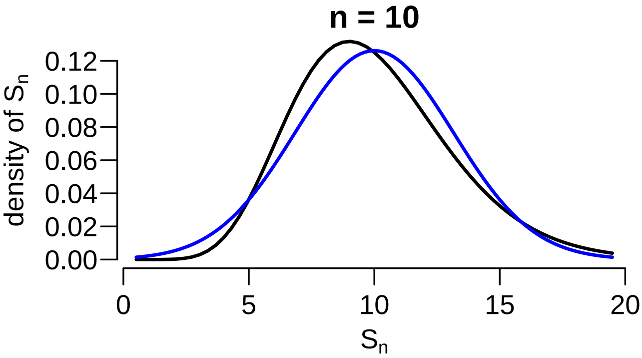


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is Gamma$(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

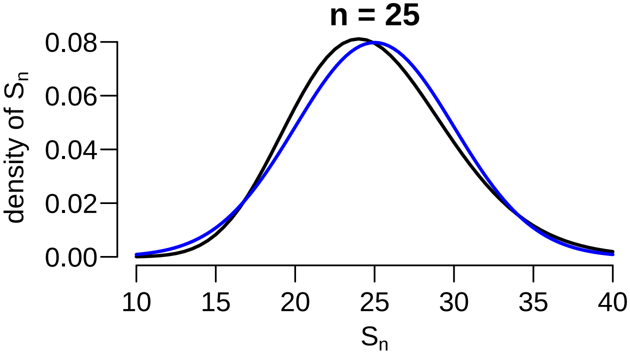


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is Gamma$(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



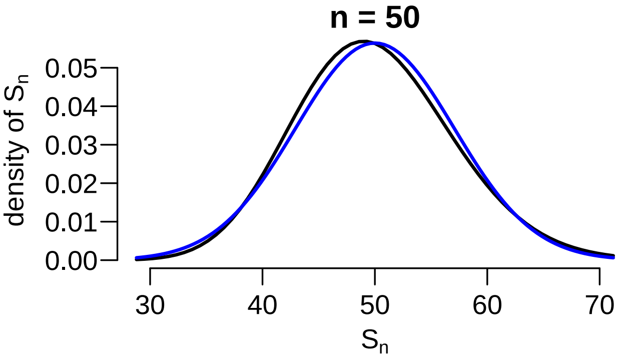
Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is $\text{Gamma}(\alpha = n, \lambda = 1)$.

Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

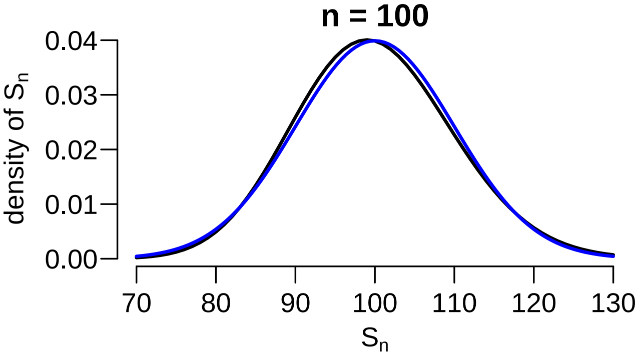


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is Gamma($\alpha = n, \lambda = 1$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

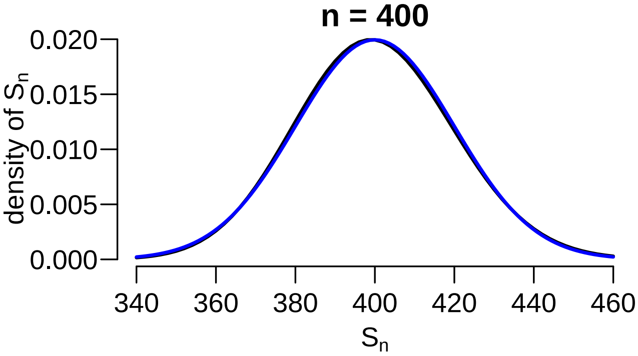


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1, \ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is $\text{Gamma}(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

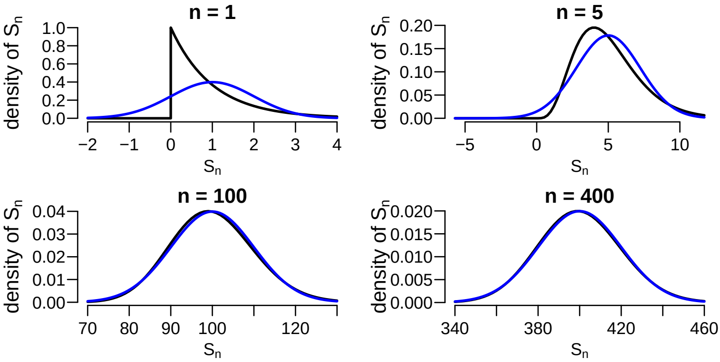


Example of CLT — Exponential

If $X_i \sim \text{Exponential}(\lambda = 1)$ with the PDF

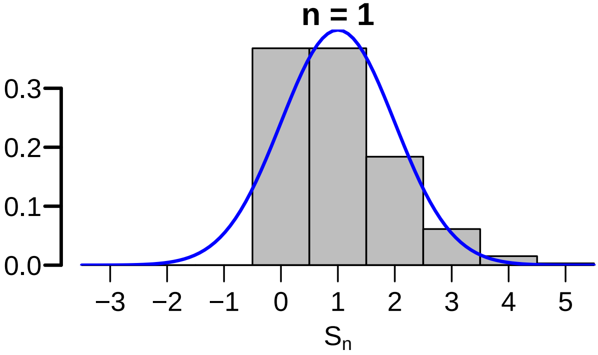
$$f(x) = e^{-x}, \quad \text{for } x > 0, \quad \mu = 1,\ \sigma^2 = 1$$

Black curve: the exact distribution of $S_n = \sum_{i=1}^{n} X_i$ is $\text{Gamma}(\alpha = n, \lambda = 1)$.
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



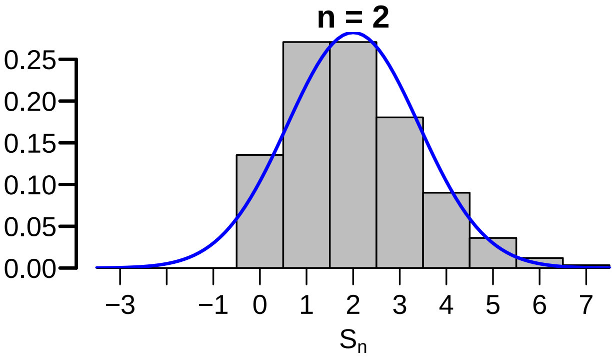
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



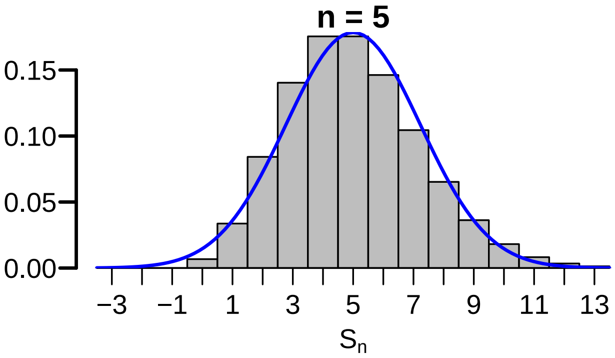
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



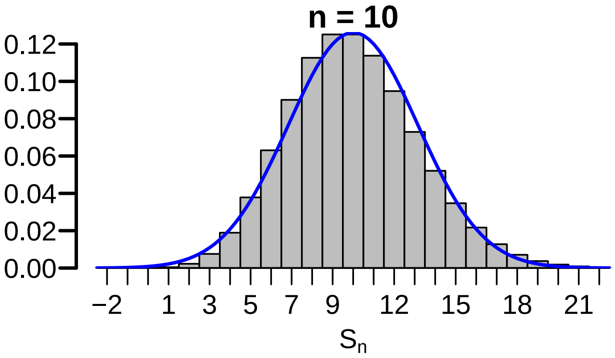
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson$(\lambda = 1)$, $\mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson$(\lambda = n)$
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



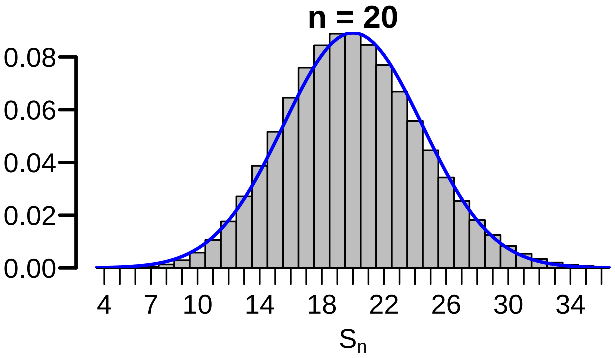
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



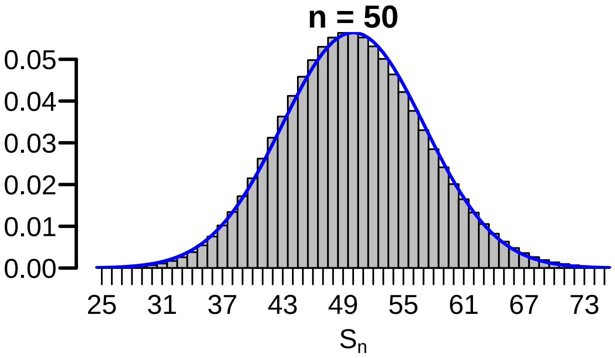
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^n X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



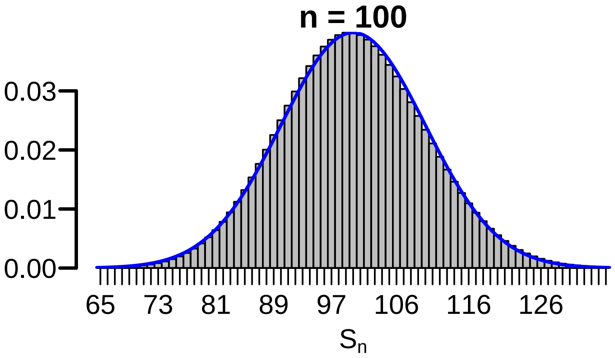
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\quad \mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



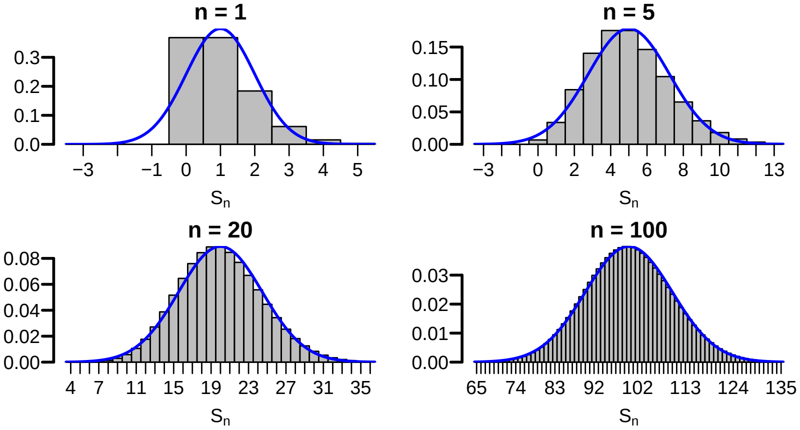
Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.



Example of CLT — Poisson

If X_i's are i.i.d. $\sim$ Poisson($\lambda = 1$), $\quad \mu = 1$, $\sigma^2 = 1$
Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is Poisson($\lambda = n$)
Blue curve: By CLT, S_n is approx. $\sim N(\mu = n, \sigma^2 = n)$.

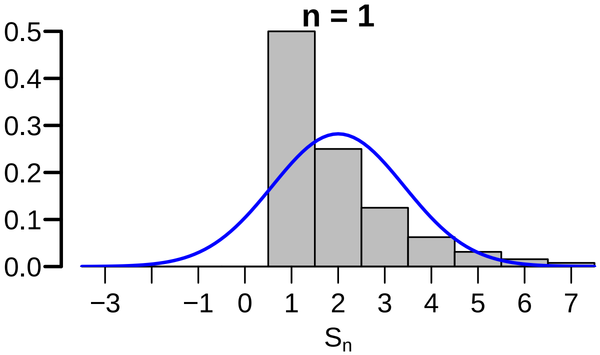


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x,\ x = 1, 2, 3, \ldots \Rightarrow \mu = \frac{1}{p} = 2,\ \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

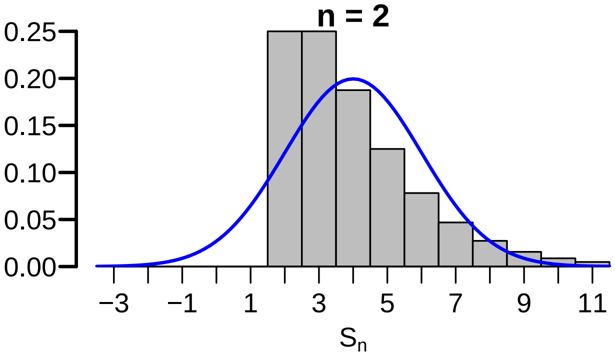


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x, \; x = 1, 2, 3, \ldots \; \Rightarrow \mu = \frac{1}{p} = 2, \; \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

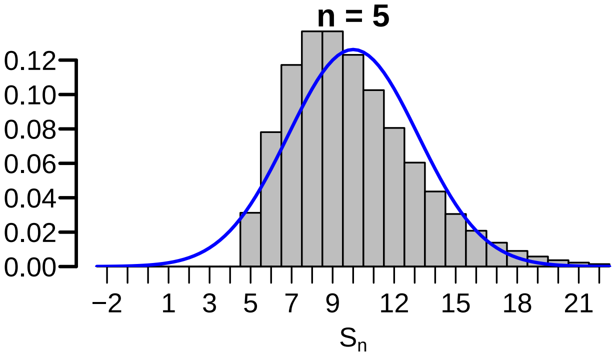


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x, \; x = 1, 2, 3, \ldots \;\Rightarrow \mu = \frac{1}{p} = 2, \; \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

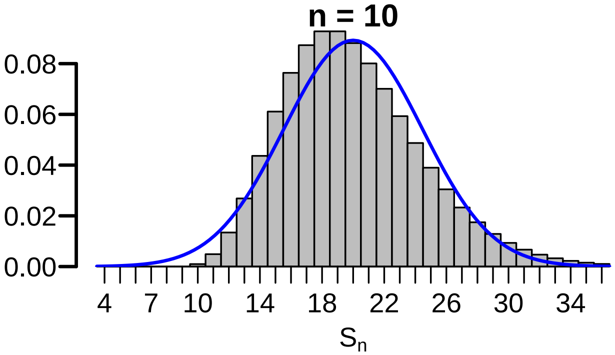


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x,\ x = 1, 2, 3, \ldots \Rightarrow \mu = \frac{1}{p} = 2,\ \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

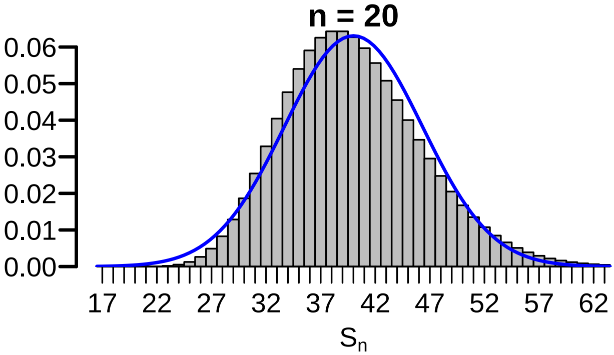


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x,\ x = 1, 2, 3, \ldots \Rightarrow \mu = \frac{1}{p} = 2,\ \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^n X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

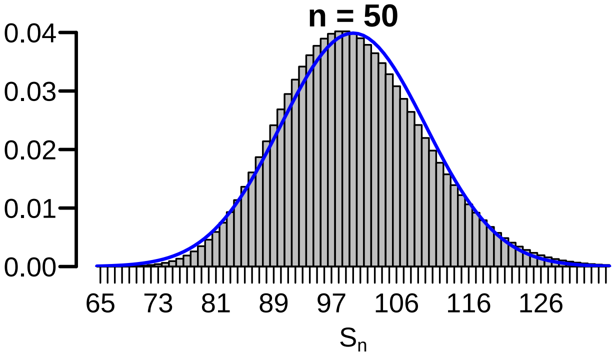


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x,\ x = 1, 2, 3, \ldots \Rightarrow \mu = \frac{1}{p} = 2,\ \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^{n} X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

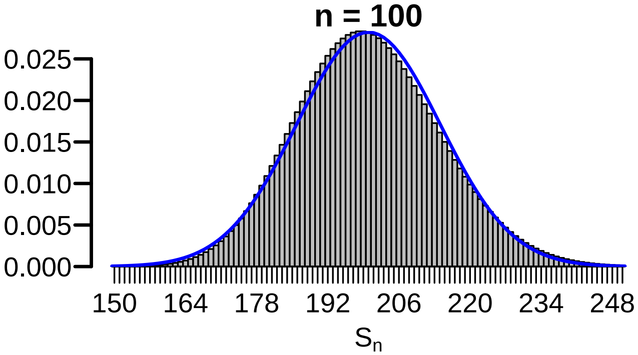


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x,\ x = 1, 2, 3, \dots \ \Rightarrow \mu = \frac{1}{p} = 2,\ \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^n X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.

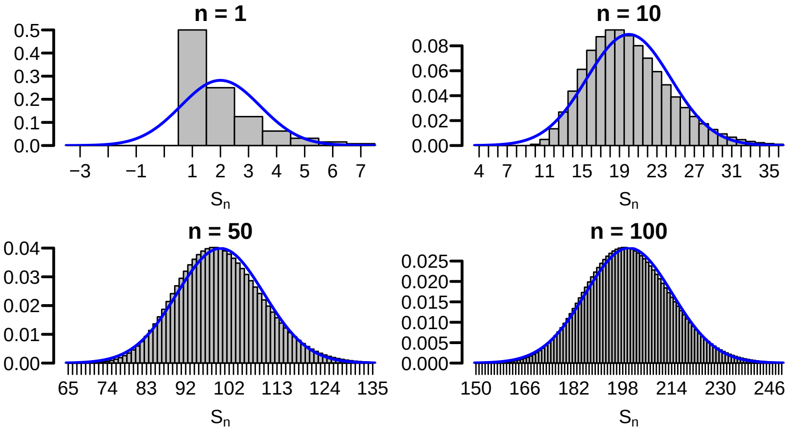


Example of CLT — Geometric

If X_i's are i.i.d. $\sim$ Geometric($p = 0.5$), with

$$\mathrm{P}(X_i = x) = (0.5)^x, \quad x = 1, 2, 3, \ldots \quad \Rightarrow \mu = \frac{1}{p} = 2, \ \sigma^2 = \frac{1-p}{p^2} = 2.$$

Histogram: exact distn. of $S_n = \sum_{i=1}^n X_i$ is NegBin($n, p = 0.5$).
Blue curve: By CLT, S_n is approx. $\sim N(\mu = 2n, \sigma^2 = 2n)$.



Normal Approximation to Binomial Distribution

Normal approximation to the Binomial distributions is a special case of CLT:

$$X = \sum_{i=1}^{n} X_i \sim Bin(n, p),$$

where $X_1, X_2, ..., X_n$ are n independent Bernoulli random variables with success probability p.

Normal Approximation to Binomial Distribution

Normal approximation to the Binomial distributions is a special case of CLT:

$$X = \sum_{i=1}^{n} X_i \sim Bin(n, p),$$

where $X_1, X_2, ..., X_n$ are n independent Bernoulli random variables with success probability p.

Therefore,

$$\mathrm{E}(X_i) = p, \quad \mathrm{Var}(X_i) = p(1-p).$$

Normal Approximation to Binomial Distribution

Normal approximation to the Binomial distributions is a special case of CLT:

$$X = \sum_{i=1}^{n} X_i \sim Bin(n, p),$$

where $X_1, X_2, ..., X_n$ are n independent Bernoulli random variables with success probability p.

Therefore,

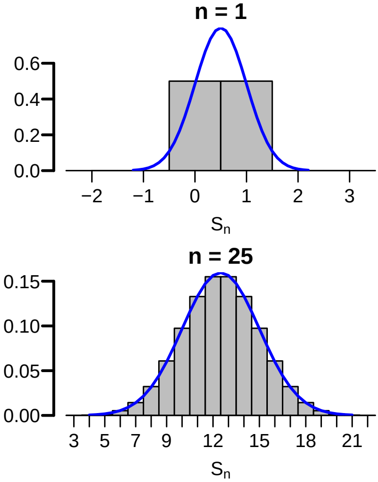
$$\mathrm{E}(X_i) = p, \quad \mathrm{Var}(X_i) = p(1-p).$$

By CLT, for large n, $Y \sim \mathrm{Bin}(n, p)$ is approximately distributed as

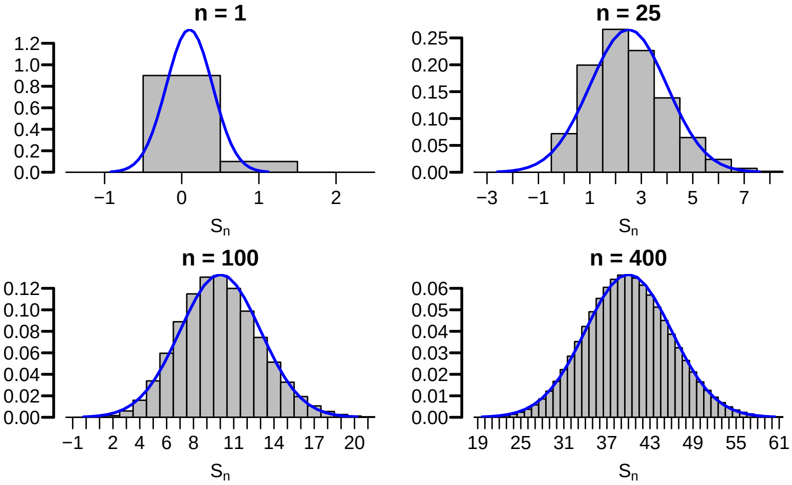
$$N(\mu_Y = np,\ \sigma_Y^2 = np(1-p)).$$

Normal Approximation to Bin($n, p = 0.5$)

When $X_1, \ldots, X_n \sim$ Bernoulli($p = 0.5$), the exact distribution of S_n is $Bin(n, p = 0.5)$



For $X_1, \ldots, X_n \sim$ Bernoulli($p = 0.1$), the exact distribution of S_n is $Bin(n, p = 0.1)$



Example 3: Roulette Calibration

With a perfectly balanced roulette wheel, red numbers should turn up 18 in 38 of the time. To test its wheel, one casino records the results of 3800 plays. Let X be the number of reds the casino got.

Q1: If the roulette wheel is perfectly balanced, what is the chance that $X \geq 1890$?

Q2 If the casino gets 1890 reds, do you think the roulette wheel should be calibrated?

Example 3: Roulette Calibration

Q1: If the roulette wheel is perfectly balanced, what is the chance that $X \geq 1890$?

Example 3: Roulette Calibration

Q1: If the roulette wheel is perfectly balanced, what is the chance that $X \geq 1890$?

Sol.: We know $X \sim \text{Bin}(n = 3800, p = \frac{18}{38})$.

Example 3: Roulette Calibration

Q1: If the roulette wheel is perfectly balanced, what is the chance that $X \geq 1890$?

Sol.: We know $X \sim \text{Bin}(n = 3800, p = \frac{18}{38})$.

Thus

$$\begin{aligned} \text{E}(X) &= np = 3800(18/38) = 1800 \\ \text{Var}(X) &= np(1-p) = 3800(18/38)(20/38) \approx 947.37 \end{aligned}$$

By CLT, X is approx. $\sim N(\mu = 1800, \sigma^2 = 947.37)$, or $Z = \frac{X-1800}{\sqrt{947.37}} \sim N(0,1)$ Thus,

$$P(X \geq 1890) \approx P\left(Z \geq \frac{1890 - 1800}{\sqrt{947.37}} \approx 2.92\right) \approx 1 - \Phi(2.92) \approx 0.00173.$$

```
1-pnorm(1890, m = 1800, s = sqrt(3800*(18/38)*(20/38)))
[1] 0.001728
```

Example 3: Roulette Calibration

As $X \sim \text{Bin}(n = 3800,\ p = \frac{18}{38})$, the exact probability of $X \geq 1890$ is

$$P(X \geq 1890) = \sum\nolimits_{k=1890}^{3800} \binom{3800}{k} \left(\frac{18}{38}\right)^{k} \left(\frac{20}{38}\right)^{3800-k} \approx 0.00183$$

We can see normal approx. to Binomial gives fairly good approx to the exact Binomial probability.

Example 3: Roulette Calibration

As $X \sim \text{Bin}(n = 3800, p = \frac{18}{38})$, the exact probability of $X \geq 1890$ is

$$P(X \geq 1890) = \sum_{k=1890}^{3800} \binom{3800}{k} \left(\frac{18}{38}\right)^k \left(\frac{20}{38}\right)^{3800-k} \approx 0.00183$$

We can see normal approx. to Binomial gives fairly good approx to the exact Binomial probability.

Q2 If the casino gets 1890 reds, do you think the roulette wheel should be calibrated?

Example 3: Roulette Calibration

As $X \sim \text{Bin}(n = 3800, p = \frac{18}{38})$, the exact probability of $X \geq 1890$ is

$$P(X \geq 1890) = \sum_{k=1890}^{3800} \binom{3800}{k} \left(\frac{18}{38}\right)^{k} \left(\frac{20}{38}\right)^{3800-k} \approx 0.00183$$

We can see normal approx. to Binomial gives fairly good approx to the exact Binomial probability.

Q2 If the casino gets 1890 reds, do you think the roulette wheel should be calibrated? Yes. $X \geq 1890$ is very unlikely to happen.

How Large n Has to Be to Use CLT?

- If the population is normal, then any n will do.
- If the population distribution is symmetric, then n should be at least 30 or so.
- The more skew or irregular the population, the larger n has to be
- For the Binomial distribution, a rule of thumb is that n should be such that

$$np \geq 10 \quad \text{and} \quad n(1-p) \geq 10.$$