STAT 234 Lecture 13
Central Limit Theorem
Section 6.1-6.2

Yibi Huang
Department of Statistics
University of Chicago

Terminology

For **i.i.d.** random variables $X_1, \ldots, X_n$ with mean μ and variance σ^2,

- *i.i.d.* = "*independent* and have an *identical distribution*"

Terminology

For **i.i.d.** random variables $X_1, \ldots, X_n$ with mean μ and variance σ^2,

- *i.i.d.* = "*independent* and have an *identical distribution*"
- the common probability distribution of individual X_i's is called the *population distribution*

Terminology

For **i.i.d.** random variables $X_1, \ldots, X_n$ with mean μ and variance σ^2,

- *i.i.d.* = "*independent* and have an *identical distribution*"
- the common probability distribution of individual X_i's is called the *population distribution*
- the collection of $\{X_1, \ldots, X_n\}$ is called a *random sample* from the population distribution

Terminology

For **i.i.d.** random variables $X_1, \ldots, X_n$ with mean μ and variance σ^2,

- *i.i.d.* = "*independent* and have an *identical distribution*"
- the common probability distribution of individual X_i's is called the *population distribution*
- the collection of $\{X_1, \ldots, X_n\}$ is called a *random sample* from the population distribution
- the mean μ of the population distribution is called the *population mean*

Terminology

For **i.i.d.** random variables $X_1, \dots, X_n$ with mean μ and variance σ^2,

- *i.i.d.* = "*independent* and have an *identical distribution*"
- the common probability distribution of individual X_i's is called the *population distribution*
- the collection of $\{X_1, \dots, X_n\}$ is called a *random sample* from the population distribution
- the mean μ of the population distribution is called the *population mean*
- the average of random sample $\{X_1, \dots, X_n\}$, $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is called the *sample mean*

Terminology

For **i.i.d.** random variables $X_1, \ldots, X_n$ with mean μ and variance σ^2,

- *i.i.d.* = "*independent* and have an *identical distribution*"
- the common probability distribution of individual X_i's is called the *population distribution*
- the collection of $\{X_1, \ldots, X_n\}$ is called a *random sample* from the population distribution
- the mean μ of the population distribution is called the *population mean*
- the average of random sample $\{X_1, \ldots, X_n\}$, $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is called the *sample mean*
- Observe that the sample mean $\overline{X}$ is also a random variable, which has a probability distribution, called the *sampling distribution of the (sample) mean*.

Weak Law of Large Number

In Lectured 11, we showed if $X_1, \ldots, X_n$ are **i.i.d.** random variables with *mean* μ and *variance* σ^2, then

$$\mathrm{E}(\overline{X}) = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

from which we can prove the *Weak Law of Large Numbers*:

$$\text{as } n \to \infty, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu.$$

Intuitively, this is clear from the mean and the variance of $\overline{X}$; the "center" of the distribution $\overline{X}$ is μ, and the "spread" around it becomes smaller and smaller as n grows.

Sampling Distribution of the (Sample) Mean

Note that the sample mean $\overline{X}$ itself is a random variable, and hence it has a probability distribution, called the *sampling distribution of the (sample) mean*.

The sampling distribution of $\overline{X}$ depends on the **population distribution**. Here are some examples.

- If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, then $\overline{X} \sim N(\mu, \sigma^2/n)$.
- If $\overline{X}$ is the average of n Bernoulli random variables $X_1, \ldots, X_n \sim$ Bernoulli(p), then $n\overline{X} \sim Bin(n, p)$, i.e.,

$$P\left(\overline{X} = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

and so on.

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be **i.i.d.** random variables with *mean* μ and *variance* σ^2. CLT asserts that, when n is large,

- the distribution of the *sample mean* $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is approximately

$$N\left(\mu_{\overline{x}} = \mu,\ \sigma^2_{\overline{x}} = \frac{\sigma^2}{n}\right).$$

- the distribution of the *total* $T = \sum_{i=1}^{n} X_i$ is approximately

$$N\left(\mu_T = n\mu,\ \sigma^2_T = n\sigma^2\right).$$

Example 1: Card Game

Recall the card game in Lecture 4, draw ONE card from a well-shuffled deck of cards and get a reward based on the card drawn as follows.

Event	reward X	$p(x)$
Heart (not ace)	\$1	$12/52$
Ace	\$5	$4/52$
King of spades	\$10	$1/52$
All else	\$0	$35/52$
Total		1

- The card drawn is placed back to the deck before he draws the card for the next game.
- Let X_i be the reward he get in the ith game, then X_i's are i.i.d. and his total reward from the 300 games is

$$X_1 + X_2 + \cdots + X_{300}$$

Example 1: Card Game

Recall the pmf for the reward X_i from one game is

x	0	1	5	10
$p_X(x)$	$35/52$	$12/52$	$4/52$	$1/52$

The expected reward from one game and the variance are

$$\mu = \mathrm{E}(X) = 0 \cdot \frac{35}{52} + 1 \cdot \frac{12}{52} + 5 \cdot \frac{4}{52} + 10 \cdot \frac{1}{52} = \frac{21}{26}$$

$$\mathrm{E}(X^2) = 0^2 \cdot \frac{35}{52} + 1^2 \cdot \frac{12}{52} + 5^2 \cdot \frac{4}{52} + 10^2 \cdot \frac{1}{52} = \frac{53}{13}$$

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2 = \frac{53}{13} - \left(\frac{21}{26}\right)^2 = \frac{2315}{26^2}$$

So if a gambler played the game 300 times, his expected value, variance of his total reward is

$$\mathrm{E}(X_1 + \cdots + X_{300}) = 300\mu = 300 \times \frac{21}{26} \approx 243.308$$

$$\mathrm{Var}(X_1 + \cdots + X_{300}) = 300\sigma^2 = 300 \times \frac{2315}{26^2}$$

$$\mathrm{SD}(X_1 + \cdots + X_{300}) = \sqrt{300 \times \frac{2315}{26^2}} = 32.052$$

The gambler is expected to get \$243.308 from the 300 games, with a standard deviation \$32.052.

Example 1: Card Game

What is the probability that the gambler can earn \$250 or more from the 300 games?

Example 1: Card Game

What is the probability that the gambler can earn \$250 or more from the 300 games?

Solution: By CLT, as $n = 300$ is large, the distribution of the total rewards $T = \sum_{i=1}^{300} X_i$ is approx. normal w/

$$\mu_T = n\mu = 300\mu = 243.308, \quad \sigma_T = \sqrt{300}\sigma = 32.052.$$

Thus

$$P(\text{total reward} > \$250) = P\left(Z > \frac{250 - 243.308}{32.052}\right)$$
$$\approx P(Z > 0.21) \approx 1 - 0.5832 \approx 0.417$$

```
1- pnorm(250, m = 243.308, s = 32.052)
[1] 0.4173
```

Example 2: Shipping Packages

Suppose a company ships packages that vary in weight:

- Packages have mean 15 lb and standard deviation 10 lb.
- Packages weights are independent from each other

Q: What is the probability that the average weight of 100 packages exceeds 17 lb?

Example 2: Shipping Packages — Solutions

Let W_i be the weight of the ith package and the total weights of 100 packages is

$$\overline{W} = \frac{1}{100}\sum_{i=1}^{100} W_i,$$

where W_i's are i.i.d. with mean $\mu_W = 15$ and SD $\sigma_W = 10$.

Example 2: Shipping Packages — Solutions

Let W_i be the weight of the ith package and the total weights of 100 packages is

$$\overline{W} = \frac{1}{100} \sum_{i=1}^{100} W_i,$$

where W_i's are i.i.d. with mean $\mu_W = 15$ and SD $\sigma_W = 10$. Then

$$\mu_{\overline{w}} = \mu_W = 15, \text{ and } \sigma_{\overline{w}} = \frac{\sigma_W}{\sqrt{100}} = \frac{10}{\sqrt{100}} = 1\,.$$

Example 2: Shipping Packages — Solutions

Let W_i be the weight of the ith package and the total weights of 100 packages is

$$\overline{W} = \frac{1}{100}\sum_{i=1}^{100} W_i,$$

where W_i's are i.i.d. with mean $\mu_W = 15$ and SD $\sigma_W = 10$. Then

$$\mu_{\overline{w}} = \mu_W = 15, \text{ and } \sigma_{\overline{w}} = \frac{\sigma_W}{\sqrt{100}} = \frac{10}{\sqrt{100}} = 1\,.$$

By CLT, $\overline{W}$ is approx. $N(\mu_{\overline{w}} = 15, \sigma^2_{\overline{w}} = 1^2)$,

Example 2: Shipping Packages — Solutions

Let W_i be the weight of the ith package and the total weights of 100 packages is

$$\overline{W} = \frac{1}{100}\sum_{i=1}^{100} W_i,$$

where W_i's are i.i.d. with mean $\mu_W = 15$ and SD $\sigma_W = 10$. Then

$$\mu_{\overline{W}} = \mu_W = 15, \text{ and } \sigma_{\overline{W}} = \frac{\sigma_W}{\sqrt{100}} = \frac{10}{\sqrt{100}} = 1\,.$$

By CLT, $\overline{W}$ is approx. $N(\mu_{\overline{W}} = 15, \sigma^2_{\overline{W}} = 1^2)$,

$$P(\overline{W} > 17) = P\left(\frac{\overline{W} - \mu_{\overline{W}}}{\sigma_{\overline{W}}} > \frac{17 - \mu_{\overline{W}}}{\sigma_{\overline{W}}}\right)$$

$$= P\left(Z > \frac{17 - 15}{1}\right) \approx 1 - \Phi(2) \approx 0.023$$

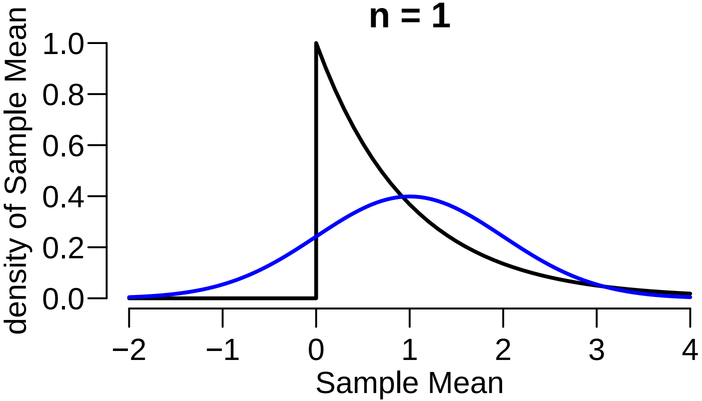
```
1- pnorm(2)
[1] 0.02275
```

If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

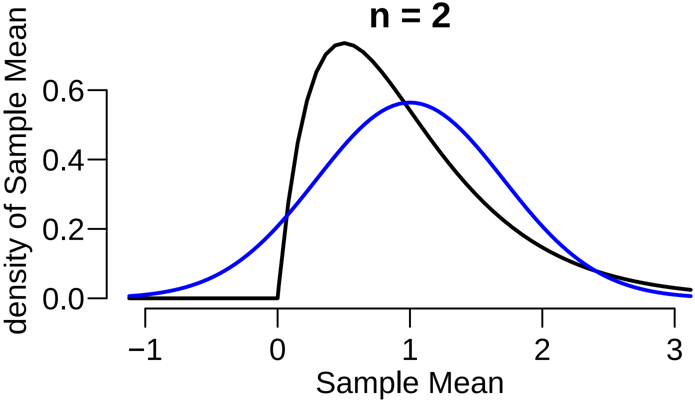


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

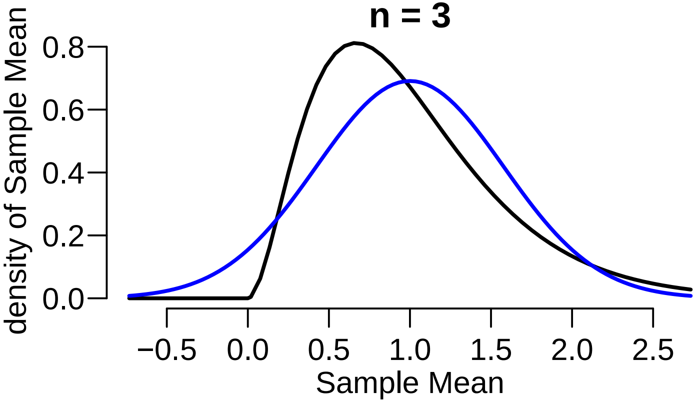


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

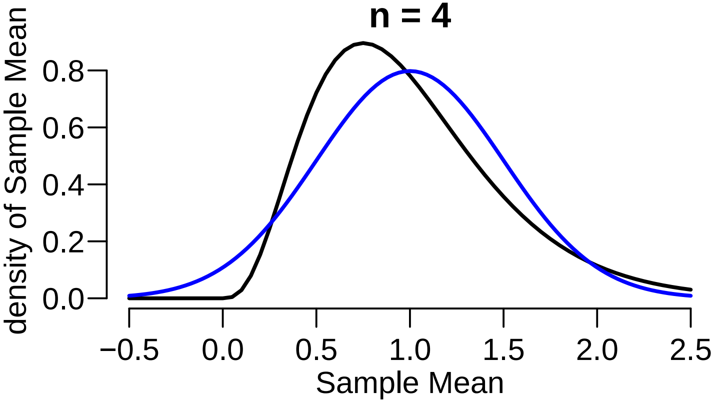


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

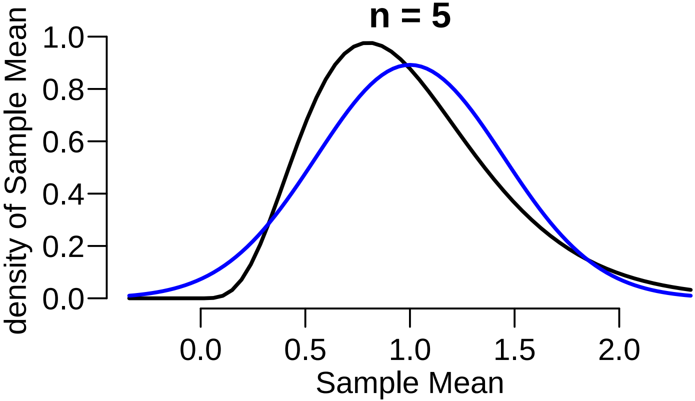


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

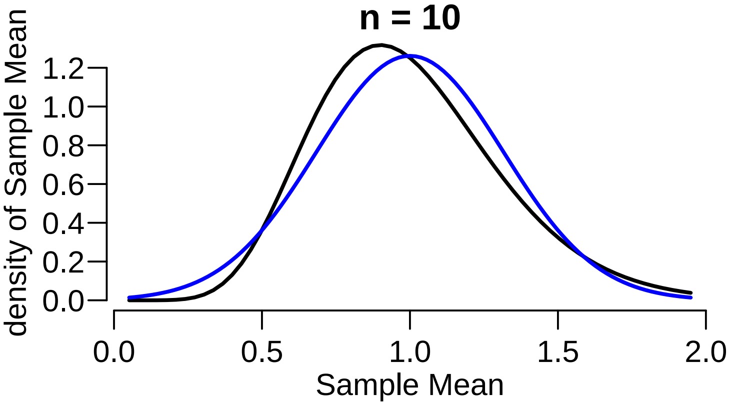


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

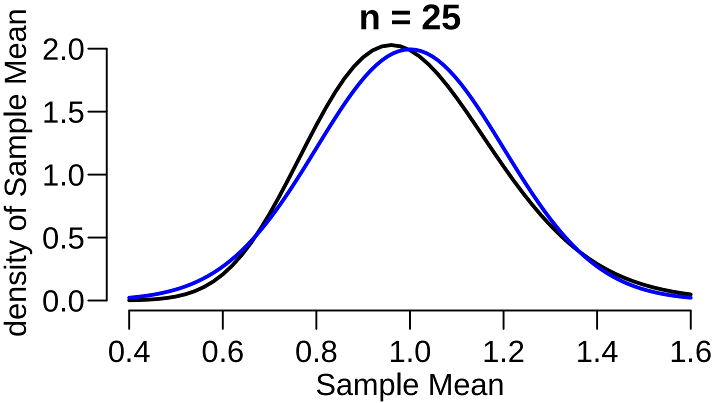


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

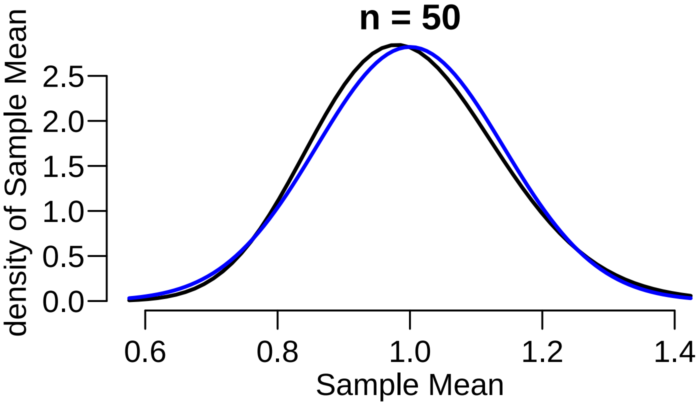


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

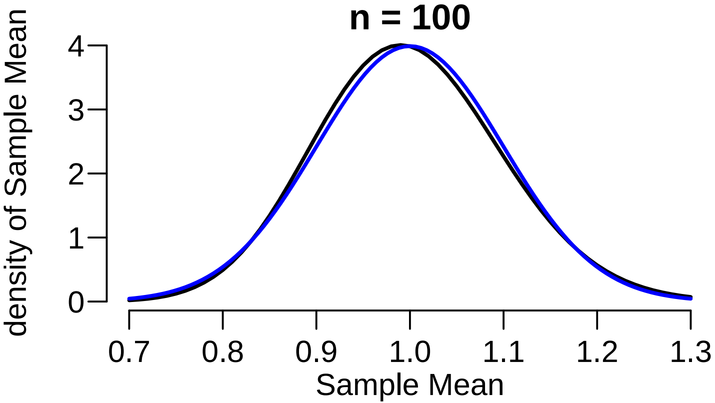


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

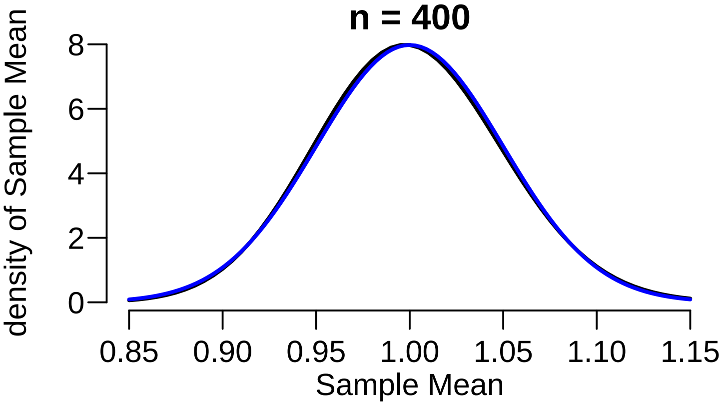


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

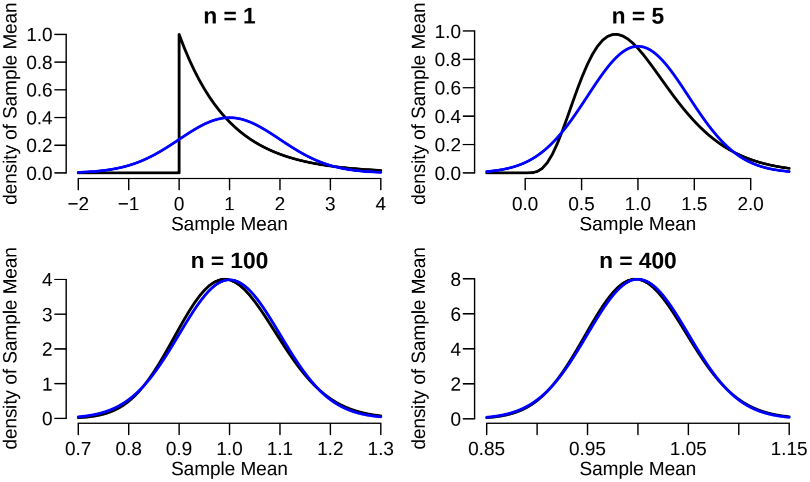


If the population distribution is exponential with density

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

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

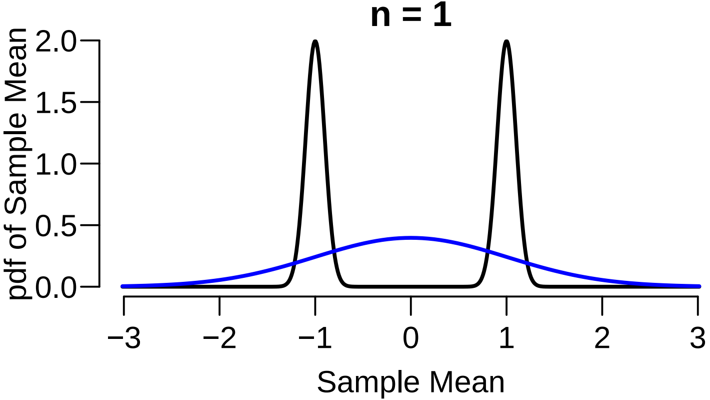


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

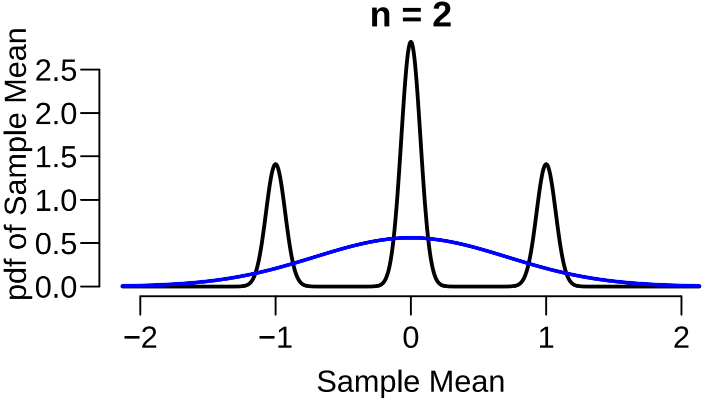


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

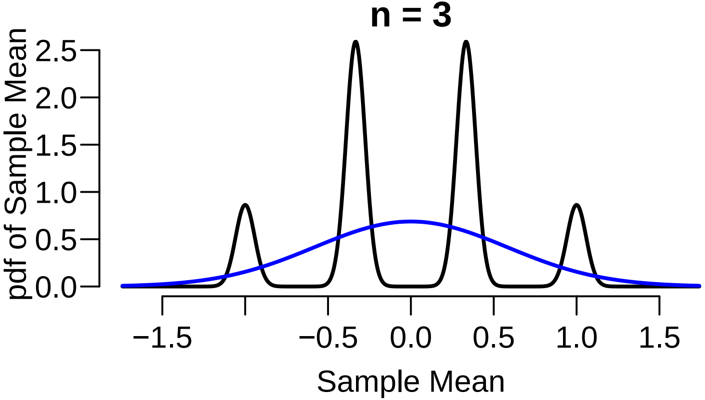


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

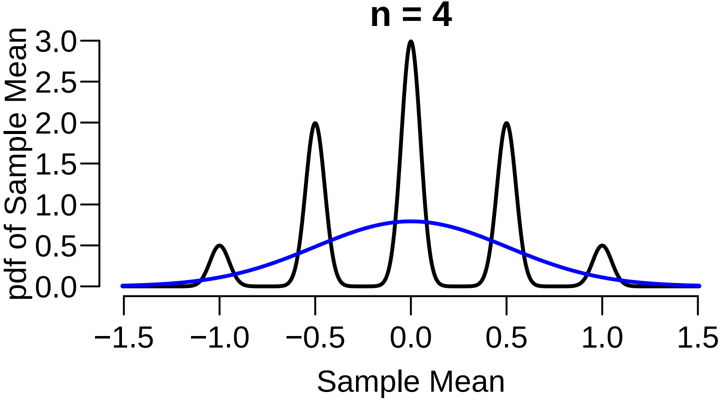


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

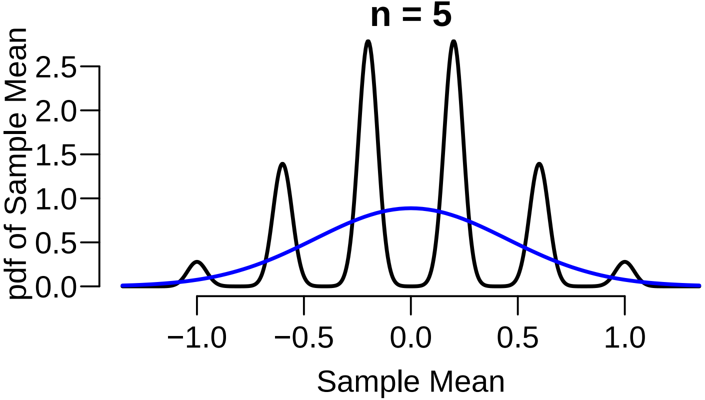


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

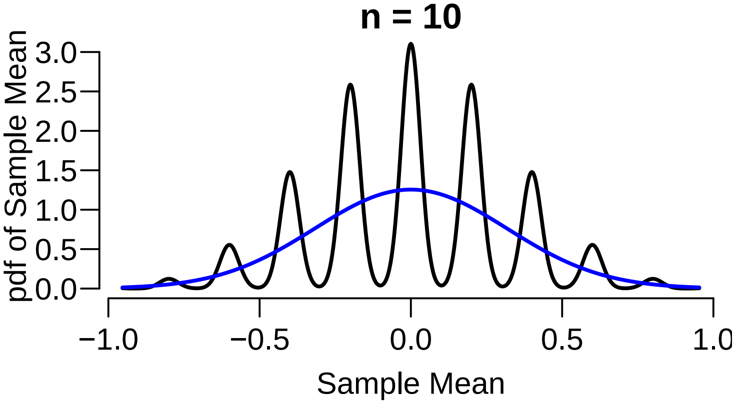


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

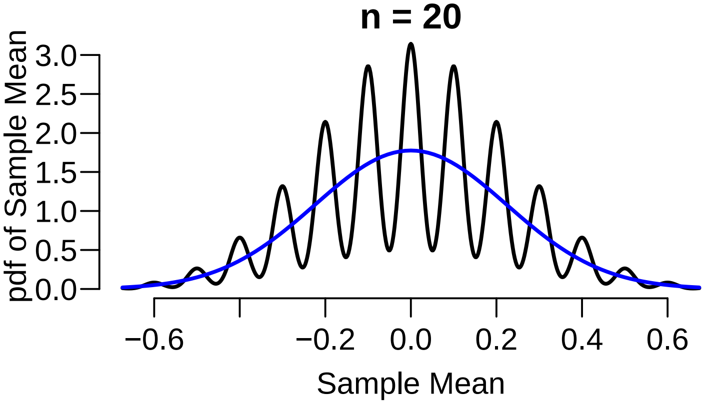


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

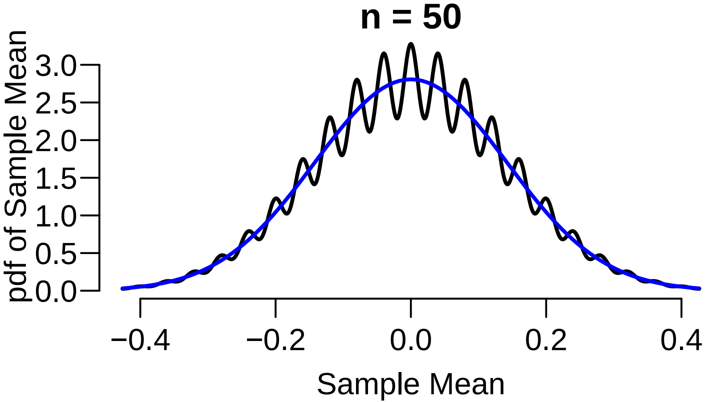


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(- \frac{(x-1)^2}{2(0.1)^2} \Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(- \frac{(x+1)^2}{2(0.1)^2} \Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

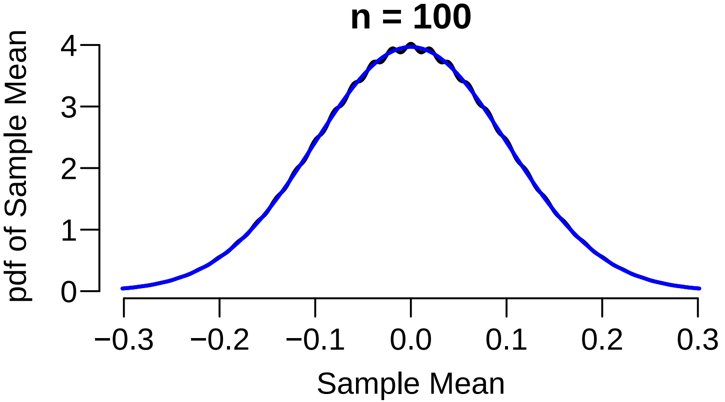


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

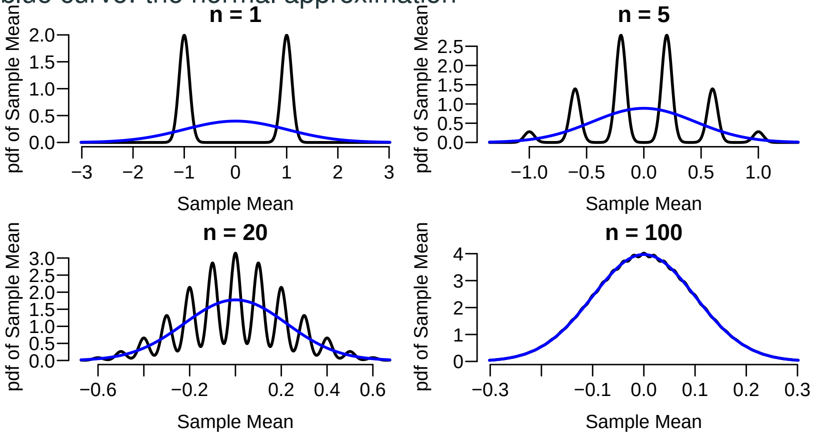


If the population distribution is Bimodal with density

$$f(x) = \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.5}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

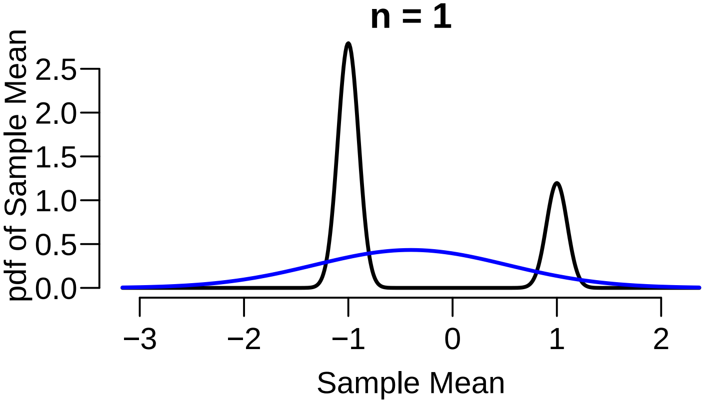


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

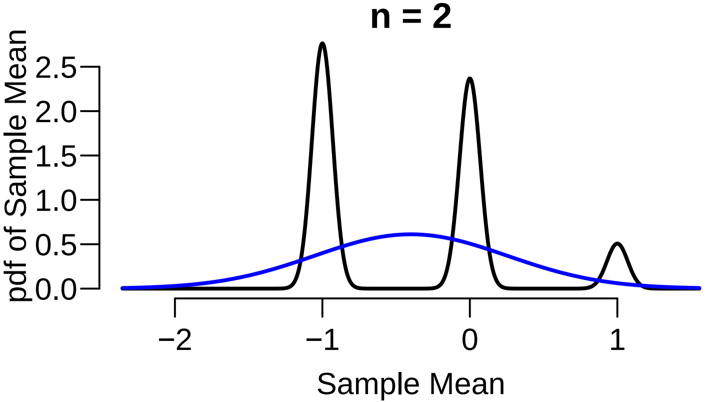


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

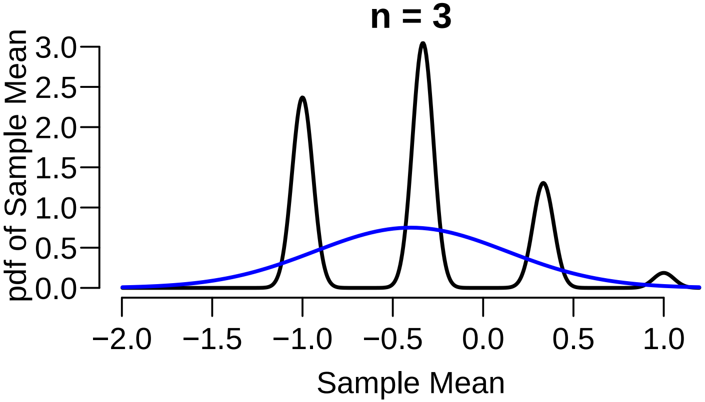


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

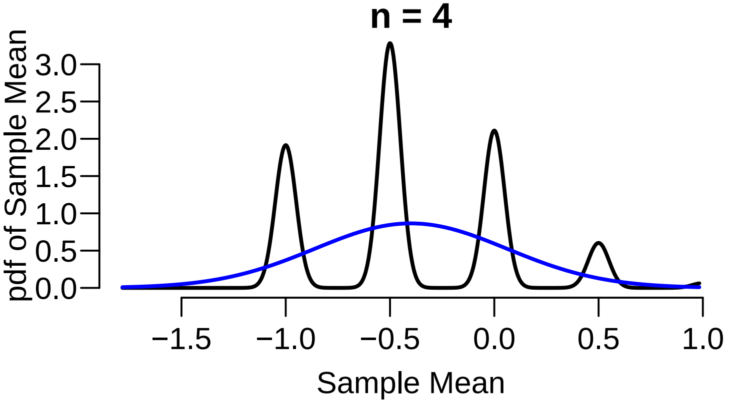


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

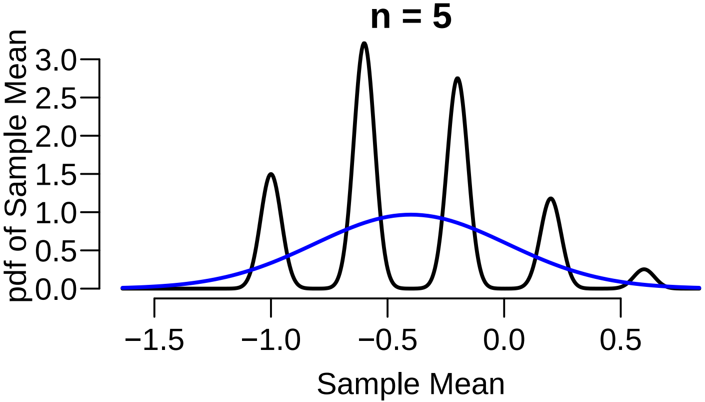


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

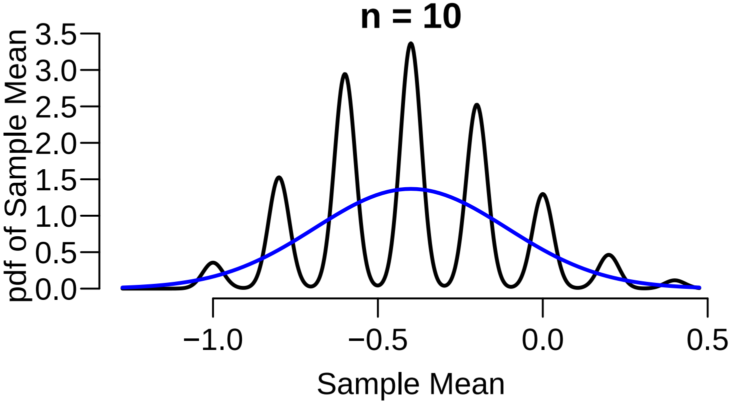


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

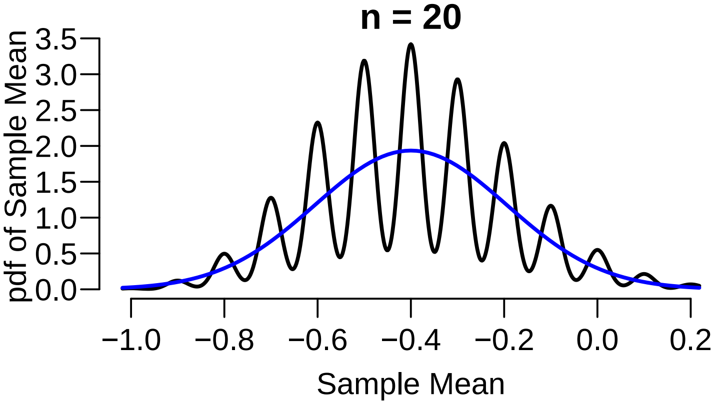


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x-1)^2}{2(0.1)^2}\right) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\left(-\frac{(x+1)^2}{2(0.1)^2}\right)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

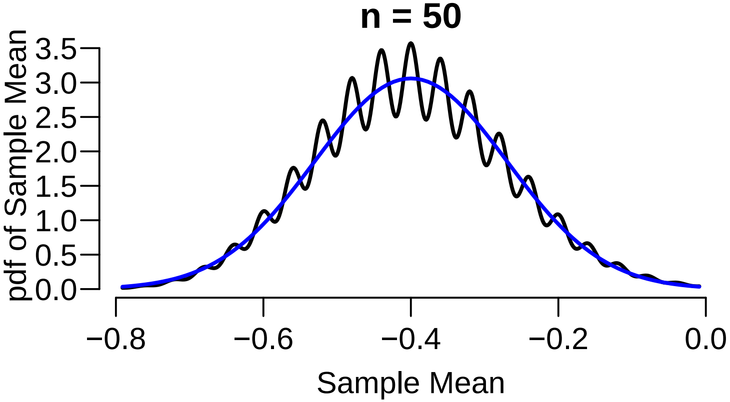


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

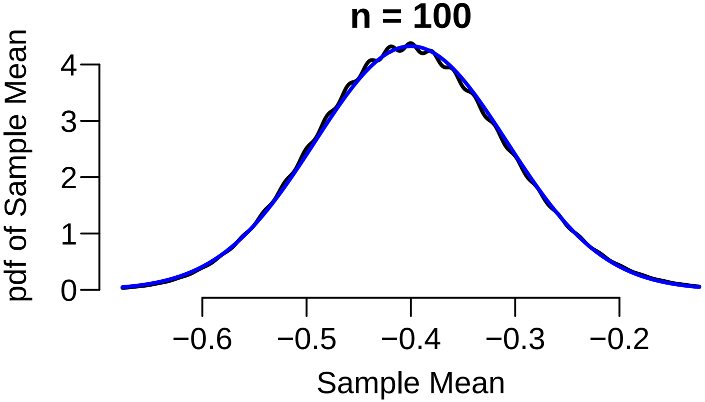


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation

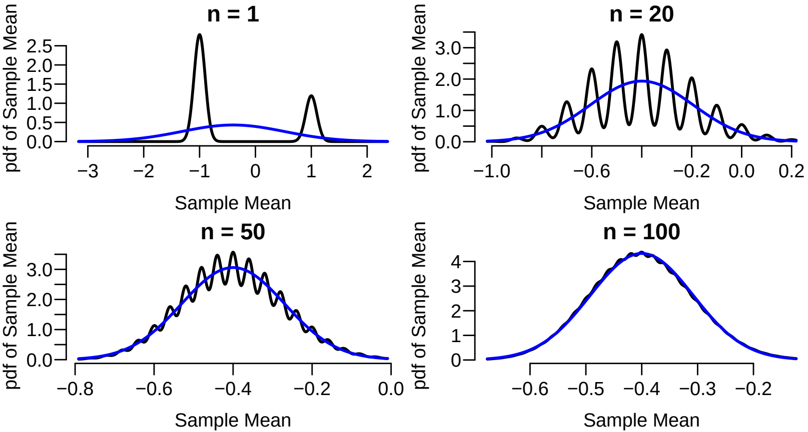


If the population distribution is Bimodal with density

$$f(x) = \frac{0.3}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x-1)^2}{2(0.1)^2}\Big) + \frac{0.7}{\sqrt{2\pi}(0.1)} \exp\Big(-\frac{(x+1)^2}{2(0.1)^2}\Big)$$

black curve: the exact sampling distribution of $\overline{X}$,

blue curve: the normal approximation



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,

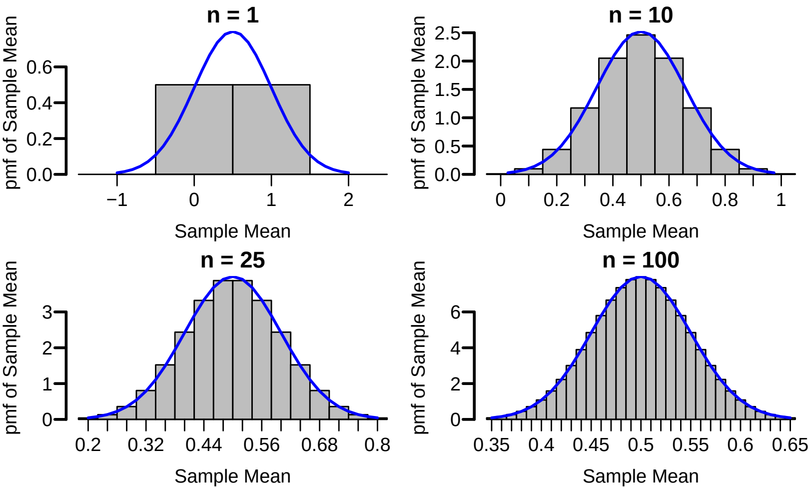
$$\mathrm{E}(X_i) = p, \quad \mathrm{Var}(X_i) = p(1-p).$$

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

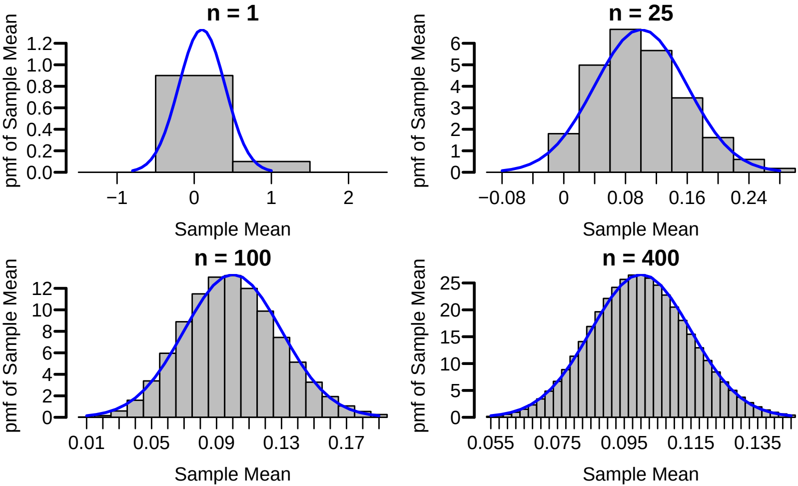
$$N(\mu_Y = np,\ \sigma_Y^2 = np(1-p)).$$

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

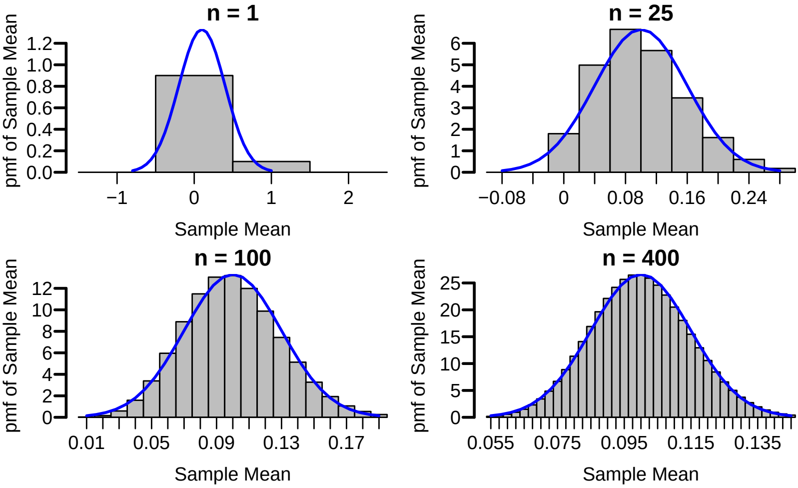
When $X_1, \ldots, X_n \sim$ Bernoulli$(p = 0.5)$, the sampling distribution of $\overline{X}$ is



For $X_1, \ldots, X_n \sim \text{Bernoulli}(p = 0.1)$, the sampling distribution of $\overline{X}$ is



For $X_1, \ldots, X_n \sim$ Bernoulli$(p = 0.1)$, the sampling distribution of $\overline{X}$ is



If the population distribution is skewed, so is the sampling distribution of the sample mean, though the skewness diminishes as the number of draws goes up.

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

$$E(X) = np = 3800(18/38) = 1800$$

$$SD(X) = \sqrt{np(1-p)} = \sqrt{3800(18/38)(20/38)} \approx 30.78$$

By CLT, X is approximately $N(\mu = 1800, \sigma^2 = (30.78)^2)$. Thus,

$$P(X \geq 1890) \approx P\left(\frac{X - 1800}{30.78} \geq \frac{1890 - 1800}{30.78}\right) \approx P(Z \geq 2.92) \approx 0.00173$$

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

Example 3: Roulette Calibration

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

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

found using R as follows.

```
sum(dbinom(1890:3800, size=3800, p = 18/38))
[1] 0.00183
```

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

Example 3: Roulette Calibration

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

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

found using R as follows.

```
sum(dbinom(1890:3800, size=3800, p = 18/38))
[1] 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 $X \sim \text{Bin}(n = 3800,\ p = 18/38)$, the exact probability of $X \geq 1890$ is

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

found using R as follows.

```r
sum(dbinom(1890:3800, size=3800, p = 18/38))
[1] 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

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