Distributions of Sampling Statistics

Mahbub Latif, PhD

January 2025

Plan

- Sampling distribution of sample mean
    - Central limit theorem
- Sampling distribution of sample variance

Introduction

- The science of statistics deals with drawing conclusions from observed data, which is often a sample from a population of interest
- To use sample data to make inferences about an entire population, it is necessary to make some assumptions between the two
  - There is an underlying probability distribution
  - The sample data are independent values drawn from this population

Introduction

- If $X_1, \ldots, X_n$ are independent random variables having a common distribution F, i.e.
  - $X_1, \ldots, X_n$ is a **random sample** from a distribution with distribution function F
- Two types of methods
  - F is specified up to some unknown parameters (parametric inference)
  - Nothing is known about F except the type of the associated variable (nonparametric inference)

Example 6.1a

- Suppose that a new process has just been installed to produce computer chips, and the successive chips produced by this new process will have lifetimes that are independent with a common unknown distribution F
- Physical reasons sometimes suggest the parametric form of the distribution F (e.g. F is a normal distribution, etc., i.e. parametric inference)
  - For normal distribution, only μ and σ^2 need to be estimated
- In other situations, there might not be any physical justification for supposing that F has any particular form (nonparametric inference)

The Sample Mean

The Sample Mean

- Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2
  - For any i, $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$
- The sample mean is defined as

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  - Sample mean $\bar{X}$ is a random variable because it is a function of random variables

Properties of $\bar{X}$

- The expected value

$$E[\bar{X}] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu$$

  - $\mu \rightarrow$ population mean

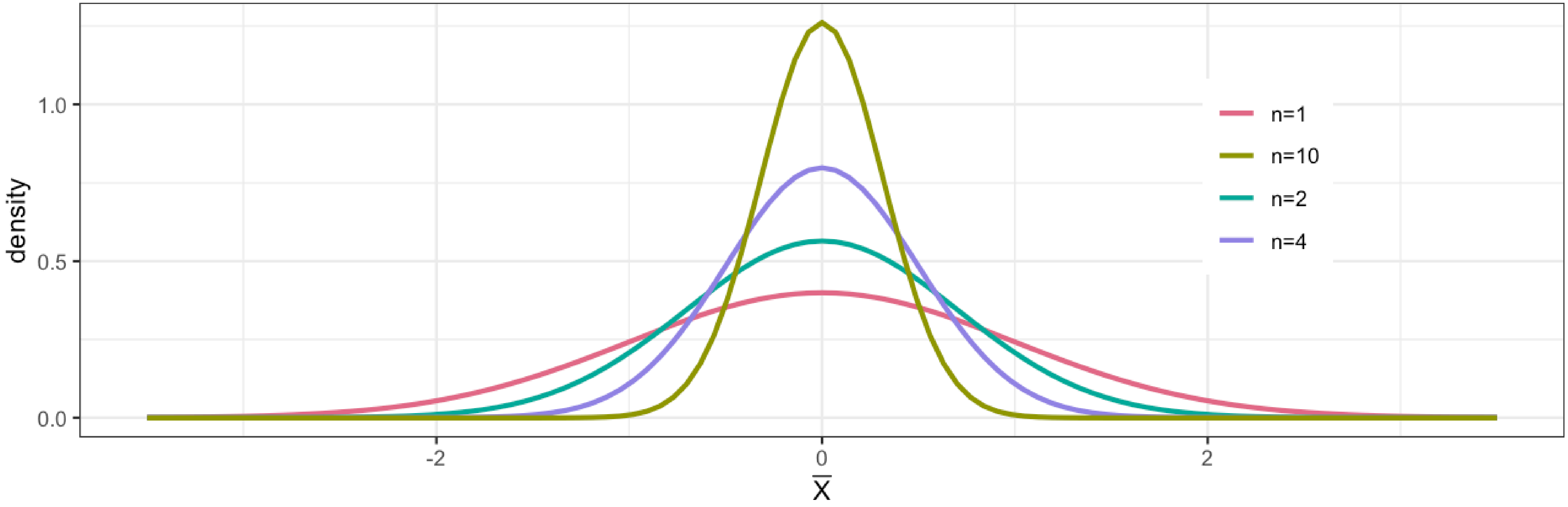
Properties of $\bar{X}$

- The variance

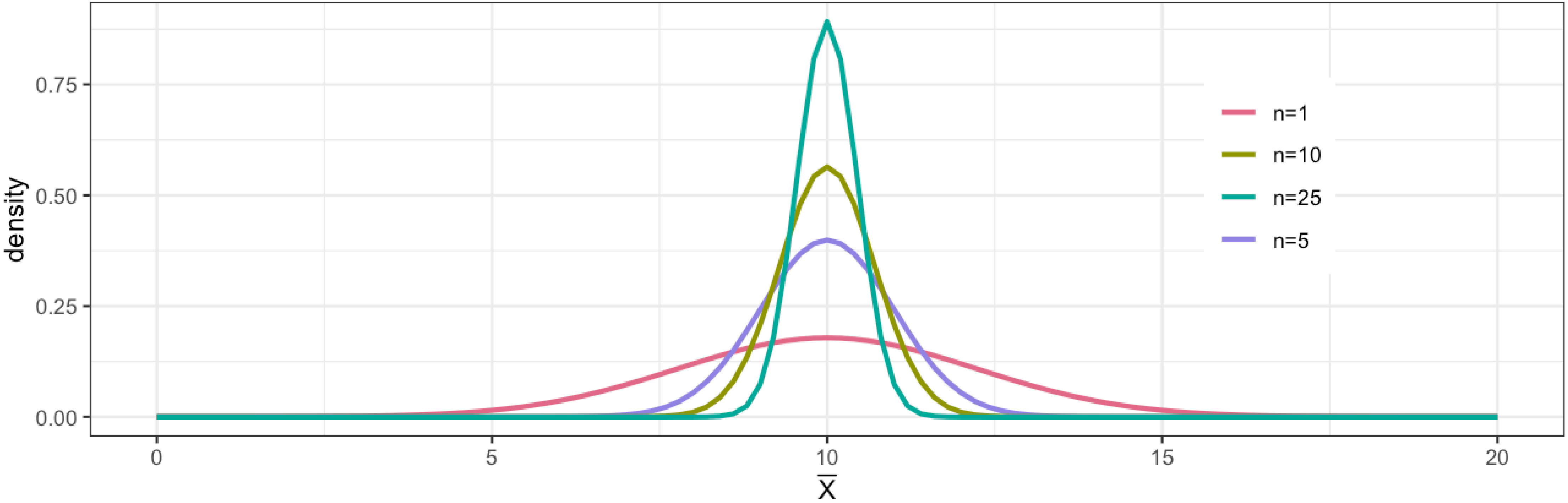
$$Var[\bar{X}] = Var\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

  - $\sigma^2 \rightarrow$ population variance
  - $n \rightarrow$ sample size

- $X_1, \ldots, X_n$ is a random sample from $N(0, 1)$
- $\bar{X} \sim N\left(0, \frac{1}{n}\right)$



- Suppose $X_1, \ldots, X_n$ is a random sample from $N(10, 5)$
  - $\bar{X} \sim N(10, 1)$ when $n = 25$



- What would be the distribution of $\bar{X}$ when the population is not normal?

Central Limit Theorem

Central Limit Theorem

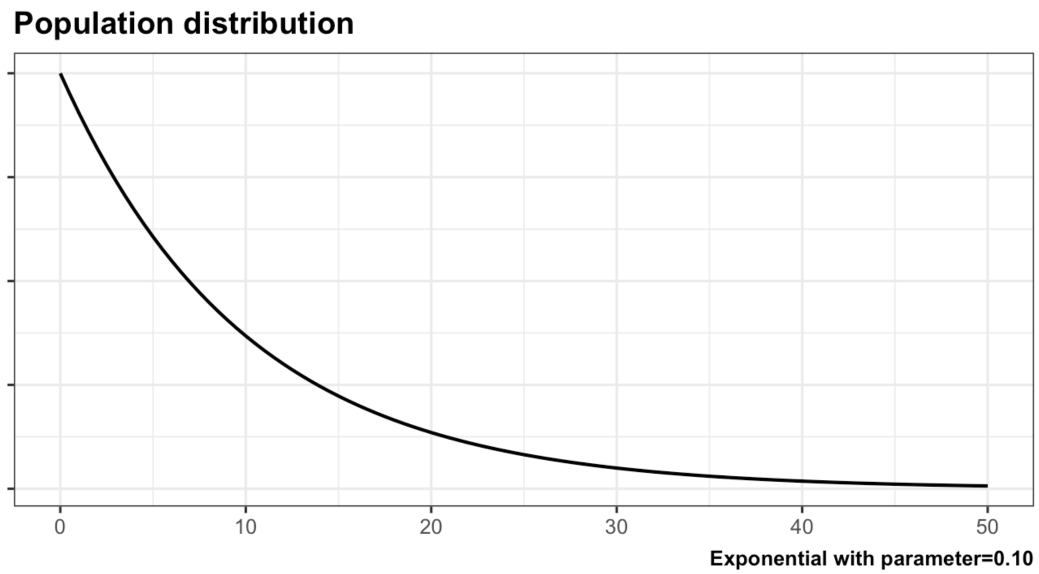
- Let $X_1, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2, **for a large n**

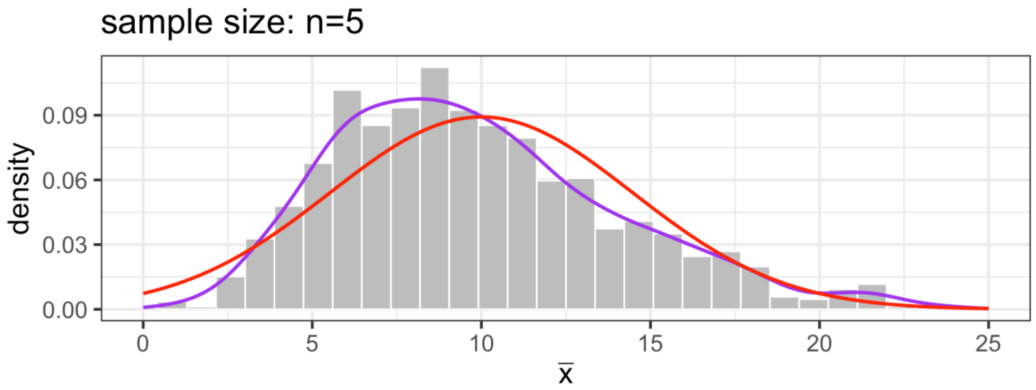
$$Y = (X_1 + \cdots + X_n) \sim N(n\mu, n\sigma^2)$$

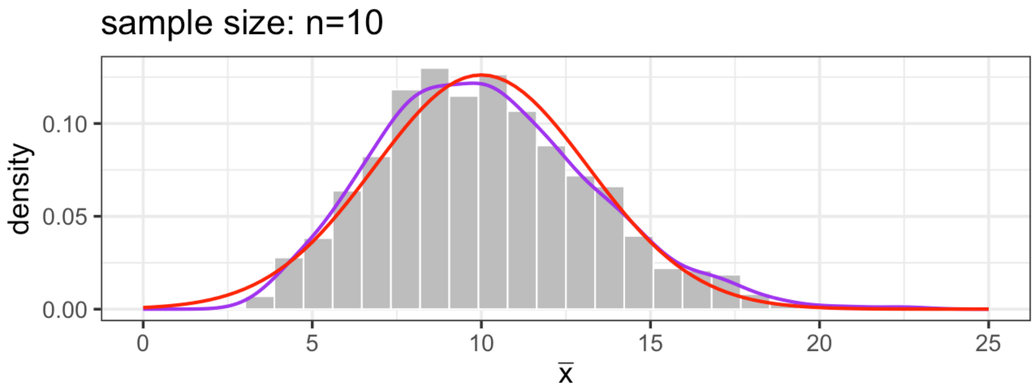
$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

$$\Rightarrow \quad Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

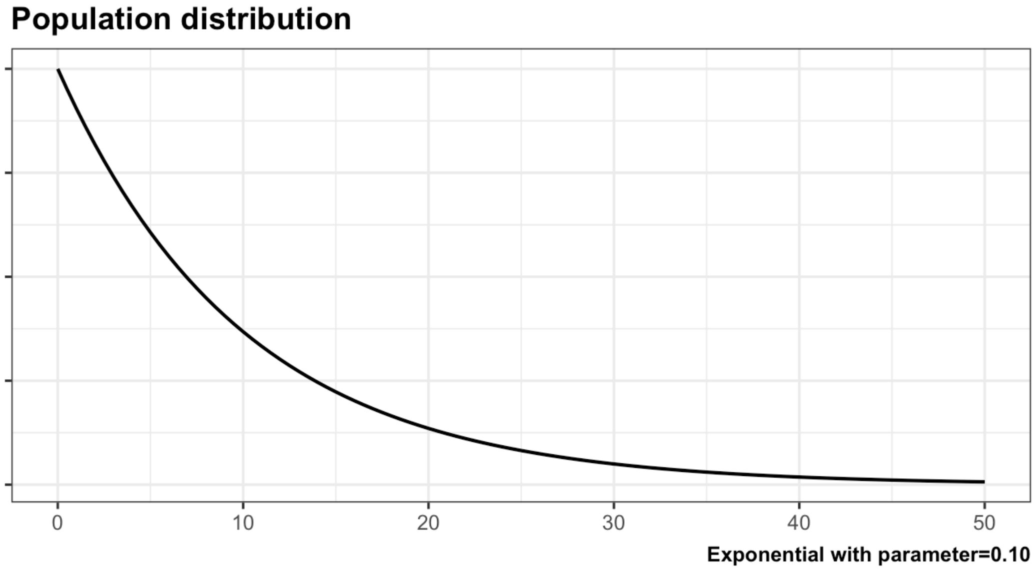
Sampling distribution of a sample mean

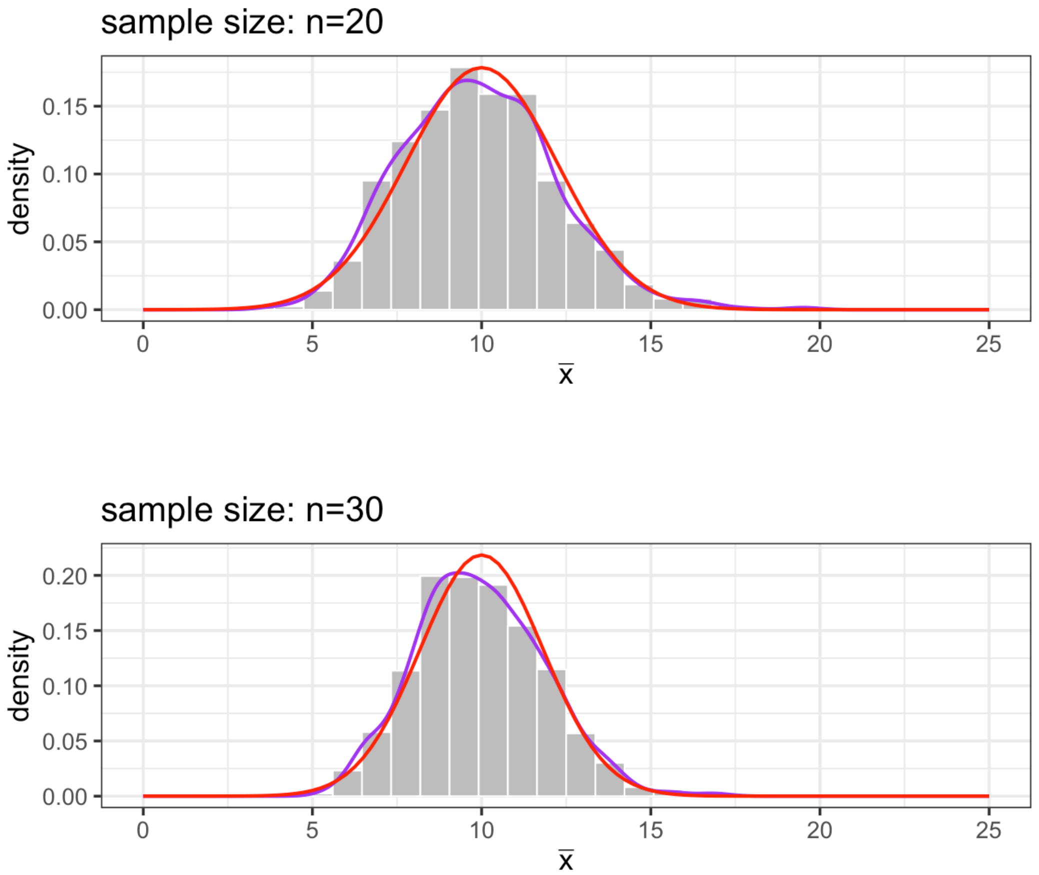




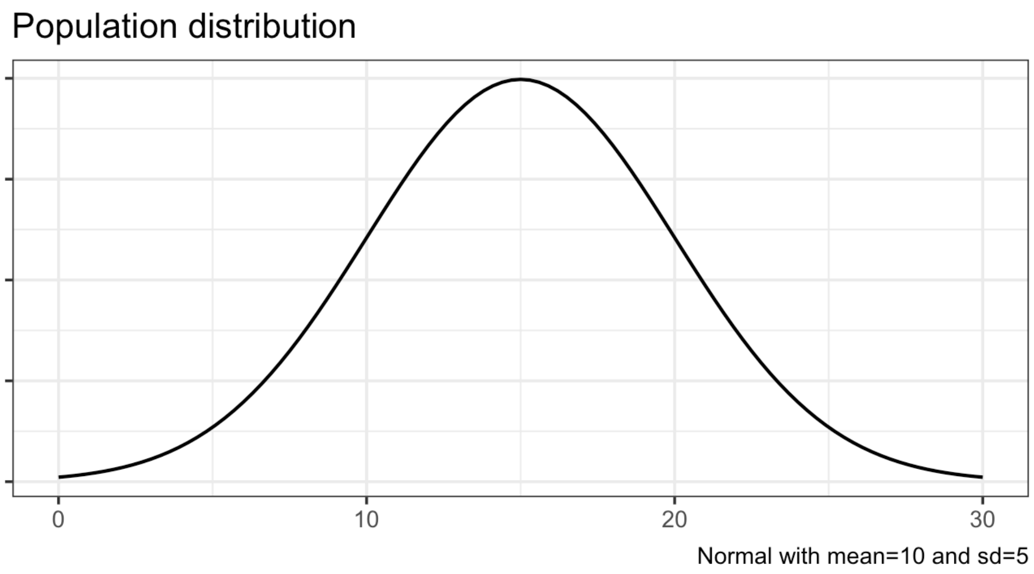


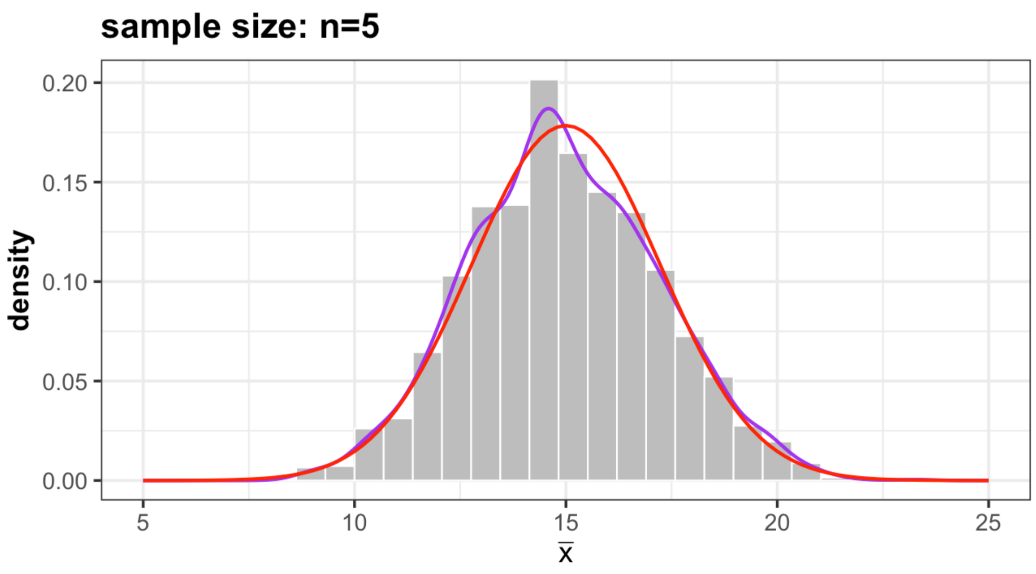
Sampling distribution of a sample mean





Normal distribution





Summary of central limit theorem

- Let $X_1, \ldots, X_n$ be a random sample from a population and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean
    - If the population is normal with mean μ and variance σ^2 then for any n

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

    - If the population is non-normal with mean μ and variance σ^2 then only for a large n

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Application of central limit theorem to binomial distribution

- Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p
- Define $X = X_1 + \cdots + X_n$ and $X \sim B(n, p)$
- Using central limit theorem, for a large n

$$X \sim N\big(np, np(1-p)\big)$$

  - $E(X) = np$ and $Var(X) = np(1-p)$

Example 6.3c

- The ideal size of a first-year class at a particular college is 150 students.
- From the past experience college knows that, on the average, only 30 percent of those accepted for admission will actually attend
- The college uses a policy of approving the applications of 450 students.
- Compute the probability that more than 150 first-year students attend this college.

Example 6.3c

- X denotes the number of students that attend and $X \sim B(450, .3)$
- Using binomial formula

$$P(X > 150) = \sum_{i=151}^{450} \binom{450}{i} (.3)^i (1 - .3)^{450-i}$$

- Using normal approximation

$$P(X > 150) = P(X > 150.5) = 1 - \Phi\left(\frac{150.5 - (450)(.3)}{(450)(.3)(1 - .3)} \right)$$
$$= 1 - \Phi(1.59) = 1 - 0.9441$$

Example 6.3d

- The weights of a population of workers have mean 167 and standard deviation 27.0
  - If a sample of 36 workers is chosen, approximate the probability that the sample mean of their weights lies between 163 and 170.
  - Repeat the above question when the sample is of size 144.