Distributions of Sampling Statistics

Mahbub Latif, PhD

February 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

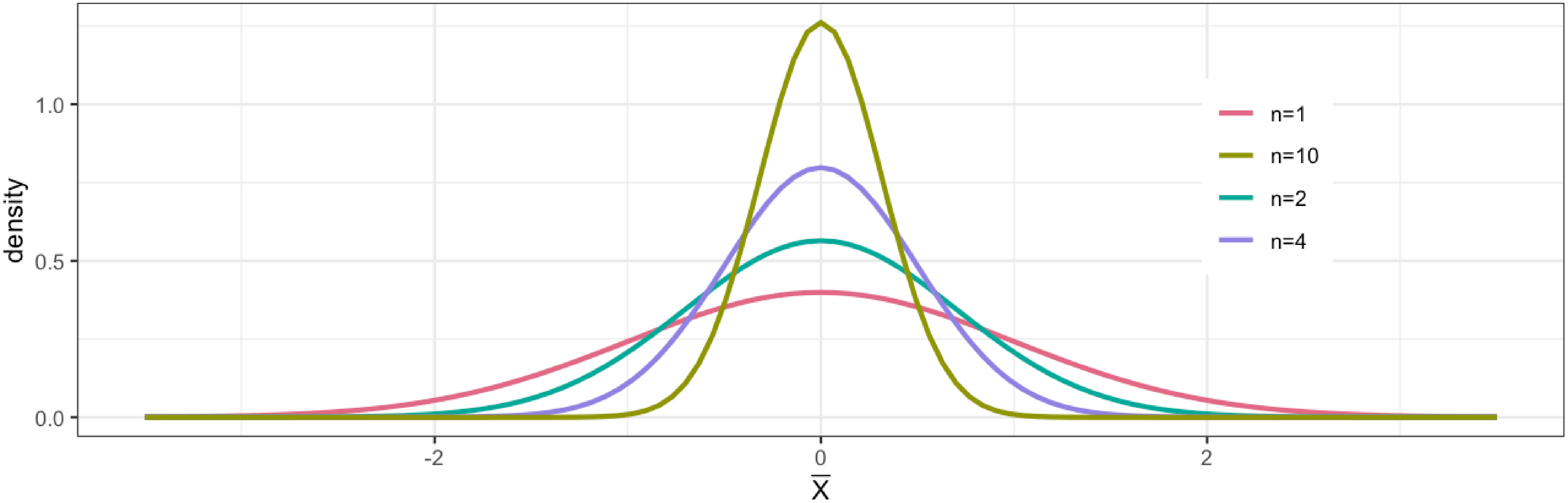
Properties of $\bar{X}$

- The variance

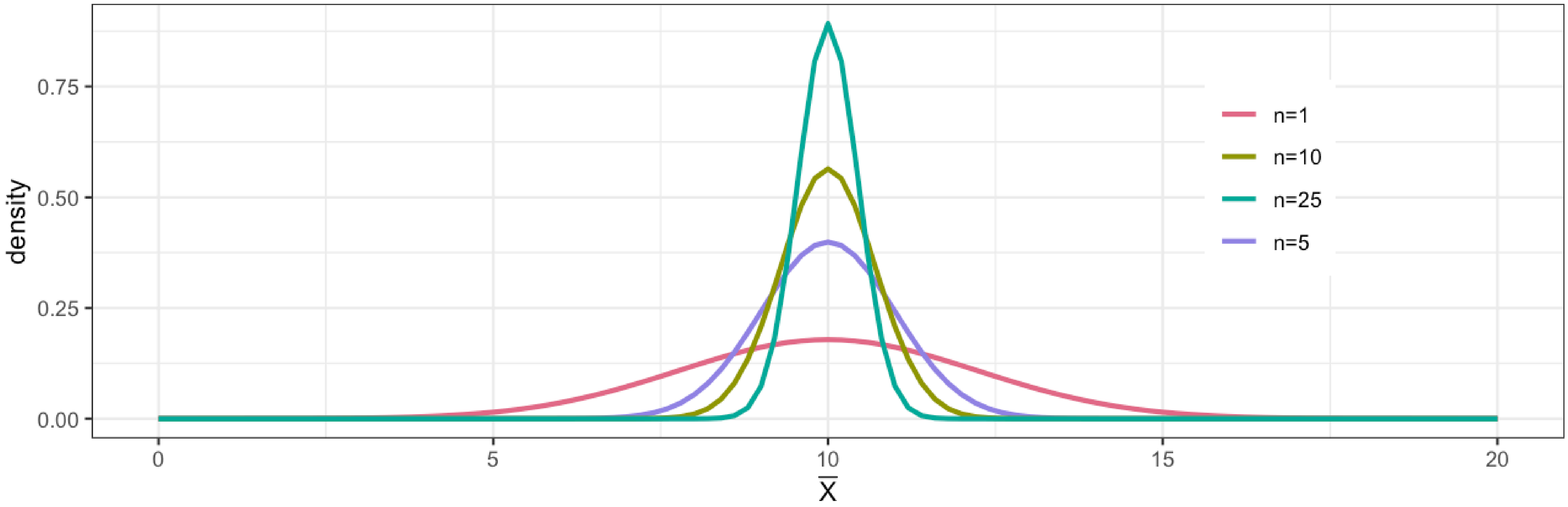
$$Var[\bar{X}] = Var\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

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

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



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



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

Central Limit Theorem

Central Limit Theorem

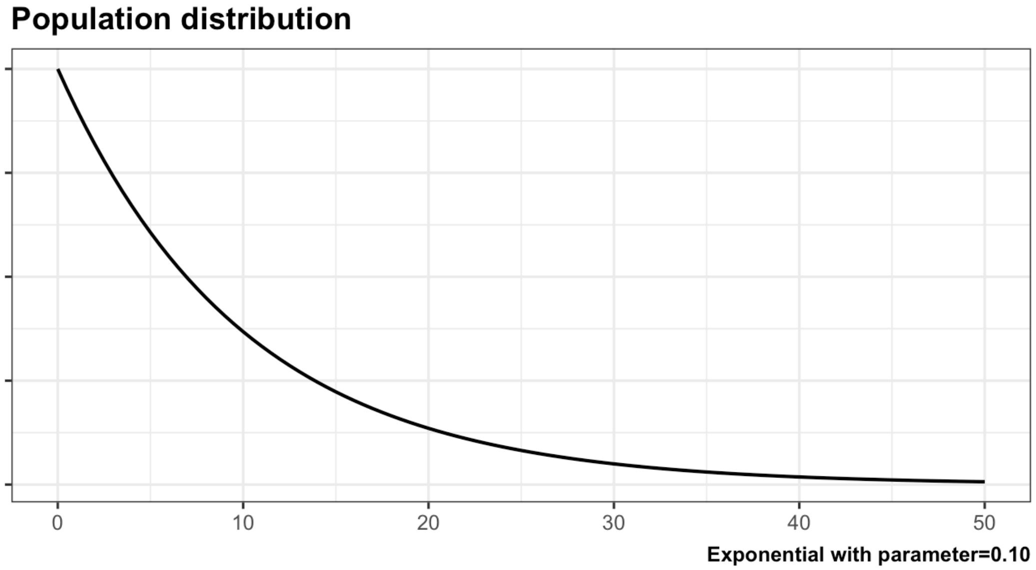
- Let $X_1, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2, **for a large n**

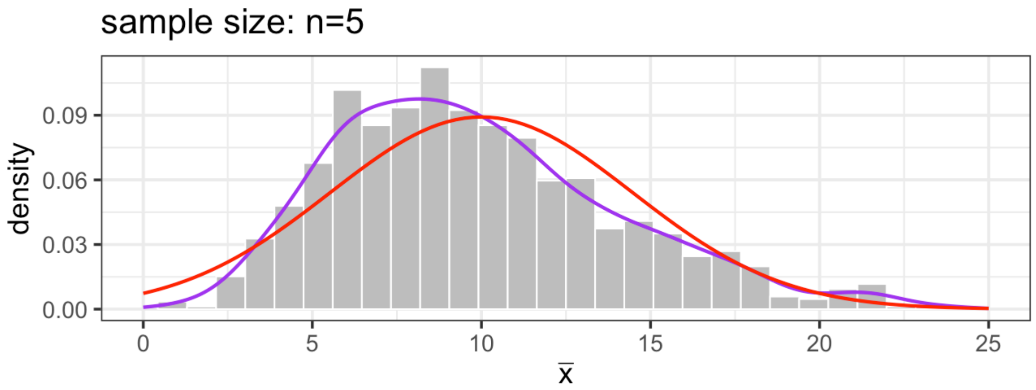
$$Y = (X_1 + \cdots + X_n) \sim N(n\mu, n\sigma^2)$$

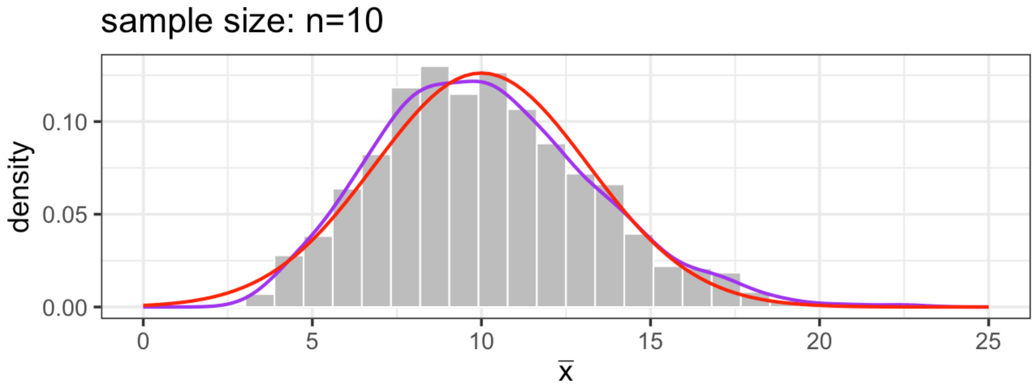
$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

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

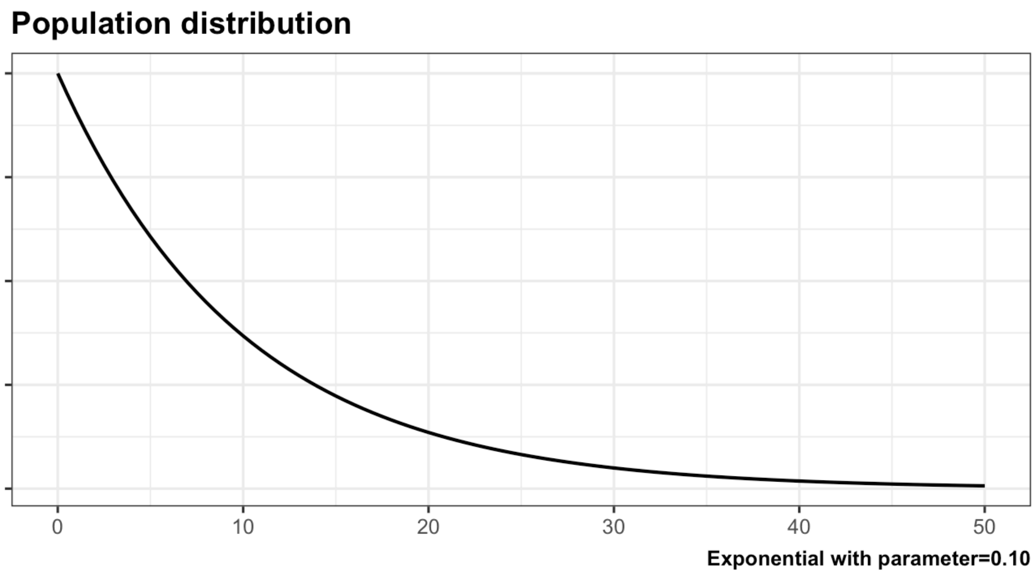
Sampling distribution of a sample mean

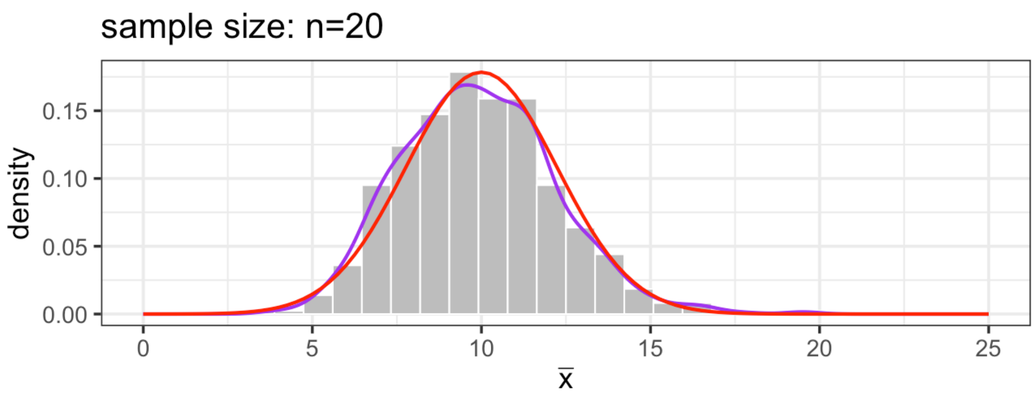


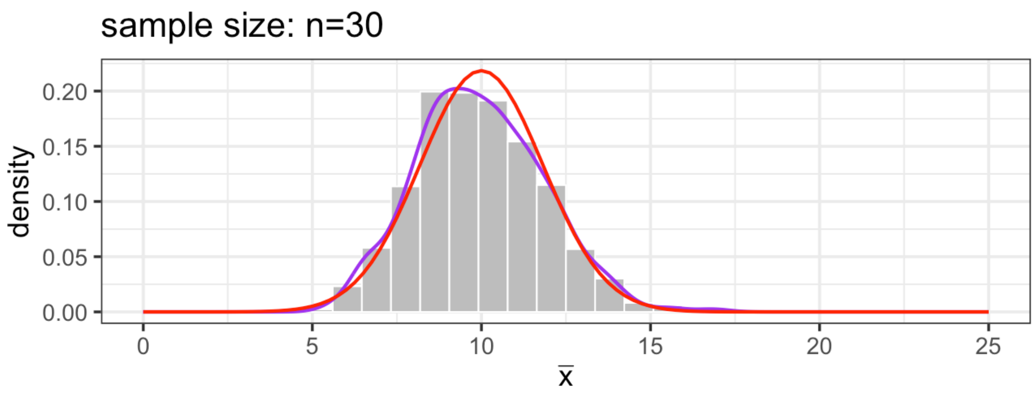




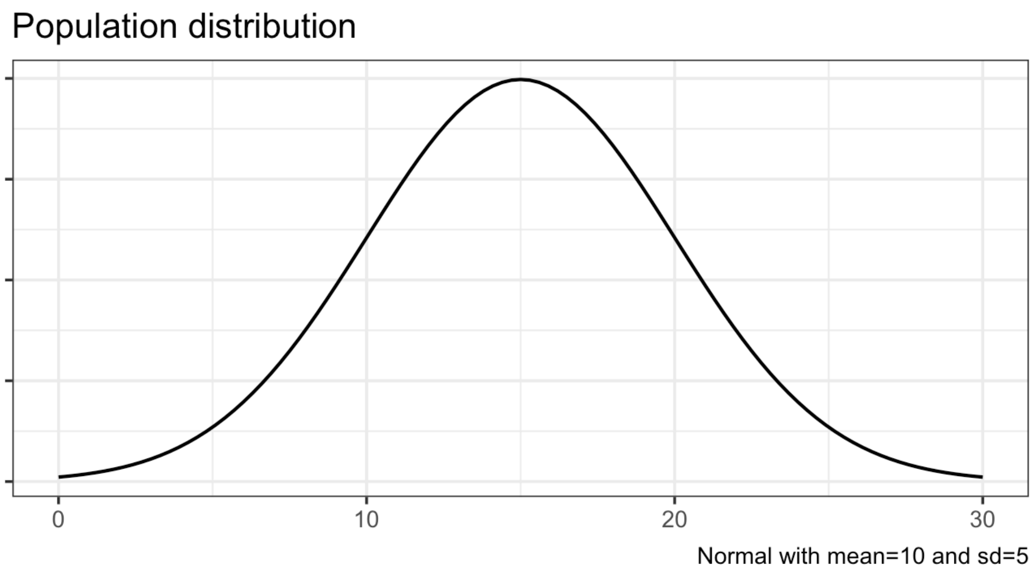
Sampling distribution of a sample mean

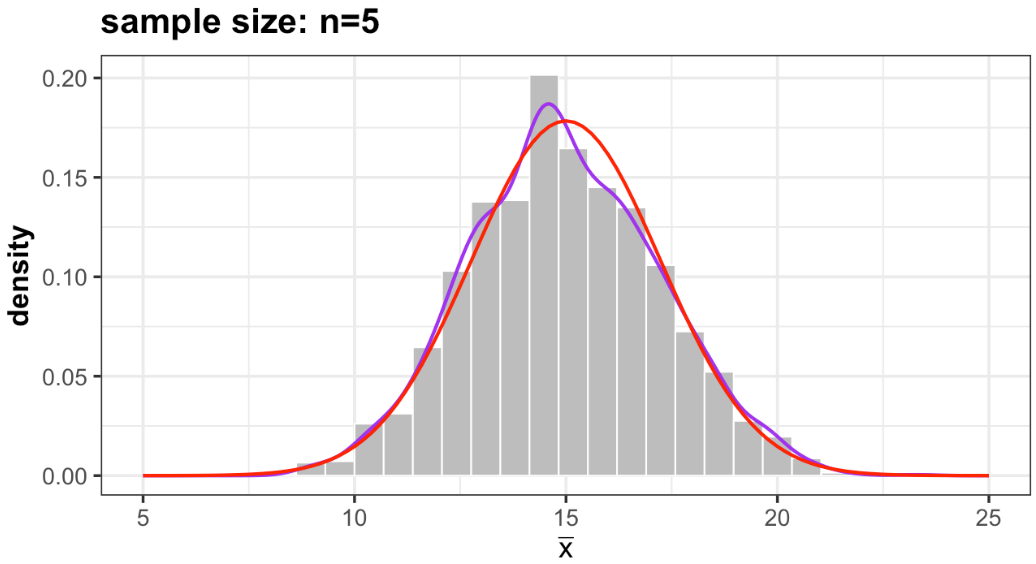






Normal distribution





Summary of central limit theorem

- Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2, and the corresponding sample mean is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  - If the population is normal then for any n

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

  - If the population is non-normal 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.

Sample variance

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

- Sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Sample standard deviation (SD): $S = \sqrt{S^2}$

- It can be shown that $E(S^2) = \sigma^2$

Sampling distribution of sample variance

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

$$\bar{X} \sim N\left(\mu, \sigma^2/n\right) \quad and \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

  - $\chi^2_{n-1} \rightarrow$ Chi-square distribution with $(n-1)$ degrees of freedom

Chi-square distribution

- Let $Z_1, \ldots, Z_n$ be n independent standard normal random variables
- The distribution of $X = \sum_{i=1}^{n} Z_i^2$ follows a chi-square distribution with n degrees of freedom, i.e.

$$X = Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2$$

Chi-square distribution

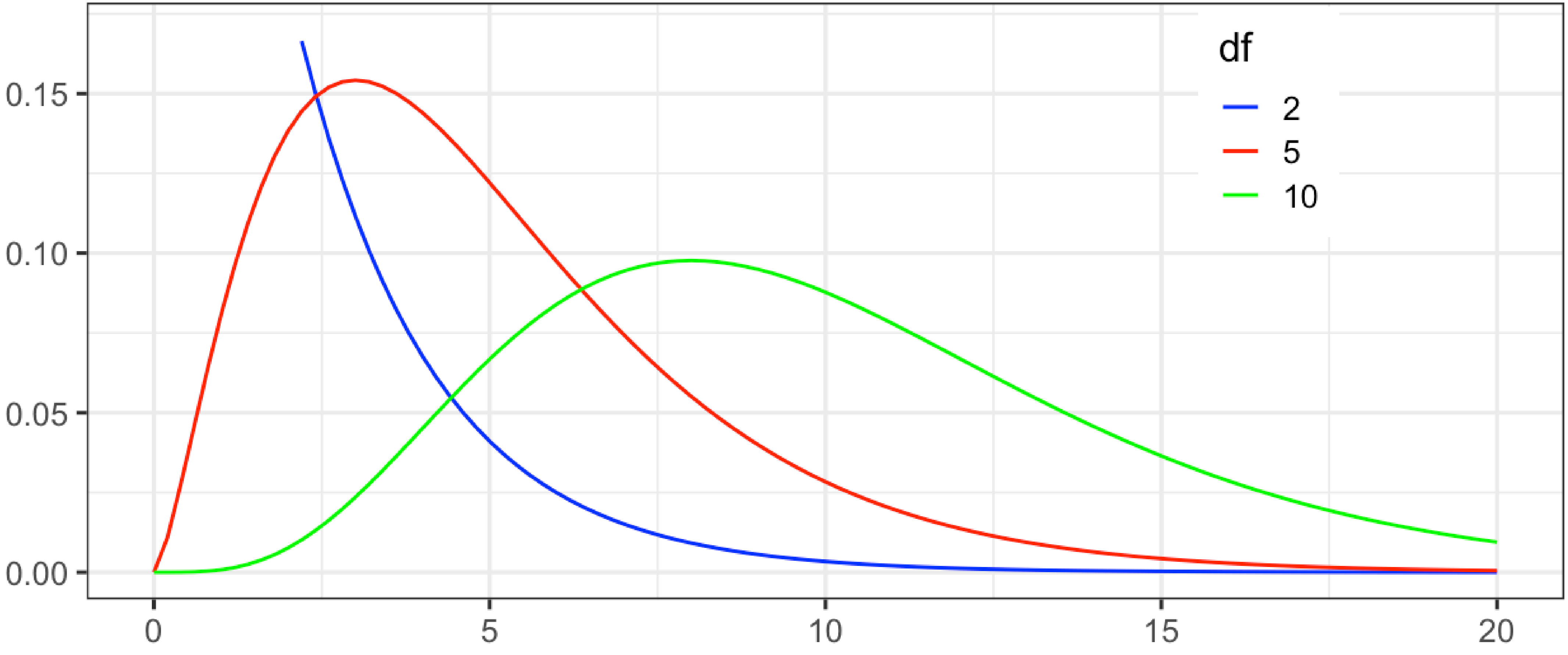


TABLE A2 *Values of* $x^2_{\alpha,n}$

n	$\alpha = .995$	$\alpha = .99$	$\alpha = .975$	$\alpha = .95$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	.0000393	.000157	.000982	.00393	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.103	5.991	7.378	9.210	10.597
3	.0717	.115	.216	.352	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	11.070	12.832	13.086	16.750
6	.676	.872	1.237	1.635	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801

The t-distribution

- Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2
- For a large n, $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$

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

The t-distribution

- If σ is unknown, it is replaced by sample standard deviation s in Z statistic
- The resulting statistic follows a t-distribution with $(n-1)$ degrees of freedom

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Comparison between t and standard normal distributions

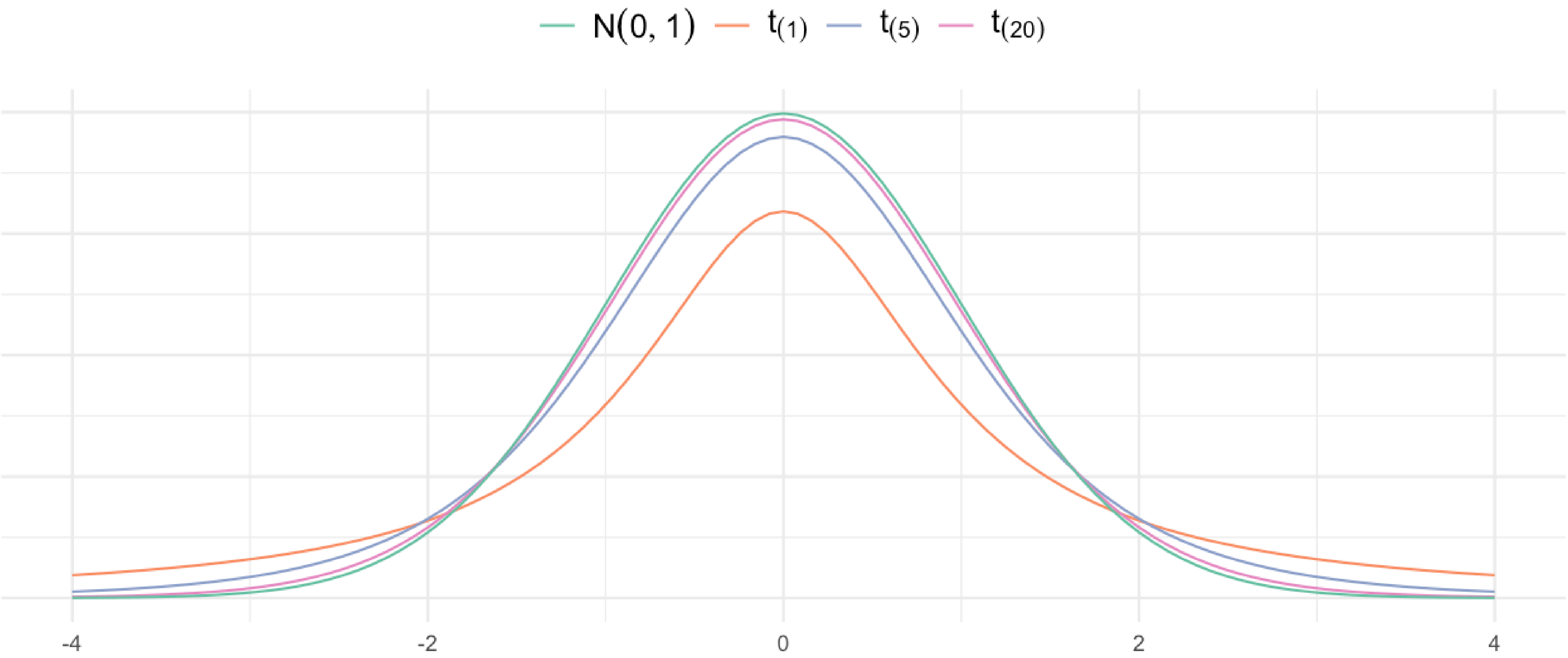


TABLE A3 *Values of $t_{\alpha,n}$*

n	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.474	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

Problems

- 1, 2, 5, 8, 9, 10, 11, 12, 13, 14, 18

Problem 1

- Plot the probability mass function of the sample mean of $X_1, \ldots, X_n$ when (i) $n = 2$ and (ii) $n = 3$.

- Assume

$$P(X = 0) = .2, \quad P(X = 1) = 0.3, \quad P(X = 3) = 0.5$$

- Calculate $E(X) = \mu$ and $V(X) = \sigma^2$

- In both cases, determine $E(\bar{X})$ and $V(\bar{X})$

Problem 1 (for $n = 2$)

X_1	X_2	$\bar{X}$	$P(\bar{X} = x)$
0	0	0.0	0.04
0	1	0.5	0.06
0	3	1.5	0.10
1	0	0.5	0.06
1	1	1.0	0.09
1	3	2.0	0.15
3	0	1.5	0.10
3	1	2.0	0.15
3	3	3.0	0.25

Probability distribution of sample mean

$\bar{x}$	$P(\bar{X} = \bar{x})$
0.0	0.04
0.5	0.12
1.0	0.09
1.5	0.20
2.0	0.30
3.0	0.25

- Obtain $E(\bar{X})$ and $Var(\bar{X})$