Some Important Sampling Distribution (II)

Huining Kang

HuKang@salud.unm.edu

August 17, 2020

Outline

- Content
  - Review
    - concepts of population distributions and sampling distributions
    - Sampling distribution of a sample mean (sampling from a finite population)
  - Sampling distribution of a sample mean (cont.)
    - Some useful properties of the random variables
    - Mean and variance of a sample mean
    - Sampling distribution of a sample mean (sampling from a normally distributed population)
    - Sampling distribution of a sample mean (sampling from a non-normal distribution)
      - The Central Limit Theorem (CTL)
      - Sampling distribution of a sample mean (from a non-normal distribution)

Review – Population Distribution and Sampling Distribution

- Population distribution
  - Suppose X is a random sample of size one drawn from a population, then the probability distribution of X is defined as the distribution of the population
- Sampling Distribution
  - A statistic T (e.g. sample mean $\bar{X}$, sample variance S^2) is a random variable. The probability distribution of T is called the sampling distribution (of T).
  - Example
    - The population of heights of adult male in the US follows a normal distribution with with mean $\mu = 70.9$ and variance $\sigma^2 = 2.75^2$. Let X be the height of a randomly selected male adult in the US. Then $X \sim N(\mu, \sigma^2)$.
    - Suppose $X_1, X_2, \cdots, X_n$ is a simple random sample of size n drawn from the above population. Later we will see that $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ also follows a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, i.e., the sampling distribution of $\bar{X}$ is a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$

Review – Population Distribution and Sampling Distribution

- Population distribution
  - Suppose X is a random sample of size one drawn from a population, then the probability distribution of X is defined as the distribution of the population
- Sampling Distribution
  - A statistic T (e.g. sample mean $\bar{X}$, sample variance S^2) is a random variable. The probability distribution of T is called the sampling distribution (of T).
  - Definition from the text book
    - The distribution of all possible values that can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population, is called the sampling distribution of that statistic.

Review - Distribution of sample mean

- Example 5.3.1 (page 136, 10th Ed)
  - A population of size $N = 5$

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

  - Population distribution

X	x				
	6	8	10	12	14
$p(x) = \mathrm{P(X = x)}$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

  - Population Mean
    - $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{6+8+10+12+14}{5} = 10,$
    - $\mu = E(X) = \sum_{i=1}^{N} x_i p(x_i) = 6 \times \frac{1}{5} + 8 \times \frac{1}{5} + 10 \times \frac{1}{5} + 12 \times \frac{1}{5} + 14 \times \frac{1}{5} = 10$
  - Population variance
    - $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{(6-10)^2+(8-10)^2+(10-10)^2+(12-10)^2+(14-10)^2}{5} = 8$
    - $\sigma^2 = Var(X) = \sum_{i=1}^{N}(x_i - \mu)^2 p(x_i) = 8$

Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

  - Problem
    - Draw a sample of size $n = 2$ (with replacement) X_1, X_2 and find the distribution of sample mean $\bar{X} = (X_1 + X_2)/2$.
  - Solution

All possible samples of size $n = 2$

		Second Draw (X_2)				
		6	**8**	**10**	**12**	**14**
First Draw	**6**	6, 6	6, 8	6, 10	6, 12	6, 14
		(6)	(7)	(8)	(9)	(10)
	8	8, 6	8, 8	8, 10	8, 12	8, 14
(X_1)		(7)	(8)	(9)	(10)	(11)
	10	10, 6	10, 8	10, 10	10, 12	10, 14
		(8)	(9)	(10)	(11)	(12)
	12	12, 6	12, 8	12, 10	12, 12	12, 14
		(9)	(10)	(11)	(12)	(13)
	14	14, 6	14, 8	14, 10	14, 12	14, 14
		(10)	(11)	(12)	(13)	(14)

Sampling distribution of $\bar{X}$

$\bar{x}$	**Frequency**	**Relative Frequency**
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
Total	25	25/25

Distribution of sample mean

- Definition the sampling distribution from the text book
  - The distribution of all possible values that can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population, is called the sampling distribution of that statistic.

All possible samples of size $n = 2$

		Second Draw (X_2)				
		6	**8**	**10**	**12**	**14**
First Draw	**6**	6, 6	6, 8	6, 10	6, 12	6, 14
		(6)	(7)	(8)	(9)	(10)
	8	8, 6	8, 8	8, 10	8, 12	8, 14
		(7)	(8)	(9)	(10)	(11)
(X_1)	**10**	10, 6	10, 8	10, 10	10, 12	10, 14
		(8)	(9)	(10)	(11)	(12)
	12	12, 6	12, 8	12, 10	12, 12	12, 14
		(9)	(10)	(11)	(12)	(13)
	14	14, 6	14, 8	14, 10	14, 12	14, 14
		(10)	(11)	(12)	(13)	(14)

Sampling distribution of $\bar{X}$

$\bar{x}$	**Frequency**	**Relative Frequency**
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
Total	25	25/25

Mean and variance of a sample mean

- Suppose
  - Population mean and variance are μ and σ^2.
  - Simple random sample (SRS): $X_1, X_2, \cdots, X_n$
  - Sample mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Sampling with replacement
  - $\mu_{\bar{X}} = \mu, \qquad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$
- Sampling without replacement (from a population of size N)
  - $\mu_{\bar{X}} = \mu, \qquad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}.$
- Sampling without replacement (from an infinite population)
  - $\mu_{\bar{X}} = \mu, \qquad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$

Some useful properties of the random variables

- Suppose X is a random variables, and α is a constant, then we have
  - $\mu_{\alpha \cdot X} = E(\alpha \cdot X) = \alpha \cdot E(X) = \alpha \cdot \mu_X$
  - $\sigma^2_{\alpha \cdot X} = Var(\alpha \cdot X) = \alpha^2 \cdot Var(X) = \alpha^2 \cdot \sigma^2_X$
- Why?

$\boldsymbol{X}$	x_1	x_2	x_3	x_4	x_5
$p(x) = P(X = x)$	p_1	p_2	p_3	p_4	p_5

$\boldsymbol{\alpha \cdot X}$	$\alpha \cdot x_1$	$\alpha \cdot x_2$	$\alpha \cdot x_3$	$\alpha \cdot x_4$	$\alpha \cdot x_5$
$p(x) = P(\alpha \cdot X = x)$	p_1	p_2	p_3	p_4	p_5

  - $\mu_X = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + x_4 \cdot p_4 + x_5 \cdot p_5$
  - $\mu_{\alpha \cdot X} = \alpha \cdot x_1 \cdot p_1 + \alpha \cdot x_2 \cdot p_2 + \alpha \cdot x_3 \cdot p_3 + \alpha \cdot x_4 \cdot p_4 + \alpha \cdot x_5 \cdot p_5$

$= \alpha \cdot (x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + x_4 \cdot p_4 + x_5 \cdot p_5)$

$= \alpha \cdot \mu_X$

Some useful properties of the random variables

- Suppose X is a random variables, and α is a constant, then we have
  - $\mu_{\alpha \cdot X} = E(\alpha \cdot X) = \alpha \cdot E(X) = \alpha \cdot \mu_X$
  - $\sigma^2_{\alpha \cdot X} = Var(\alpha \cdot X) = \alpha^2 Var(X) = \alpha^2 \cdot \sigma^2_X$
- Why

X	x_1	x_2	x_3	x_4	x_5
$p(x) = P(X = x)$	p_1	p_2	p_3	p_4	p_5

$\alpha \cdot X$	$\alpha \cdot x_1$	$\alpha \cdot x_2$	$\alpha \cdot x_3$	$\alpha \cdot x_4$	$\alpha \cdot x_5$
$p(x) = P(\alpha \cdot X = x)$	p_1	p_2	p_3	p_4	p_5

  - $\sigma^2_X = (x_1 - \mu_X)^2 \cdot p_1 + (x_2 - \mu_X)^2 \cdot p_2 + \cdots + (x_5 - \mu_X)^2 \cdot p_5$

  - $\sigma^2_{\alpha \cdot X} = (\alpha \cdot x_1 - \mu_{\alpha \cdot X})^2 \cdot p_1 + (\alpha \cdot x_2 - \mu_{\alpha \cdot X})^2 \cdot p_2 + \cdots + (\alpha \cdot x_5 - \mu_{\alpha \cdot X})^2 \cdot p_5$

$= \alpha^2 \cdot (x_1 - \mu_X)^2 \cdot p_1 + \alpha^2 \cdot (x_2 - \mu_X)^2 \cdot p_2 + \cdots + \alpha^2 \cdot (x_5 - \mu_X)^2 \cdot p_5$

$= \alpha^2 \cdot [(x_1 - \mu_X)^2 \cdot p_1 + (x_2 - \mu_X)^2 \cdot p_2 + \cdots + (x_5 - \mu_X)^2 \cdot p_5]$

$= \alpha^2 \cdot \sigma^2_X$

Replace $\mu_{\alpha \cdot X}$ with $\alpha \cdot \mu_X$

Some useful properties of the random variables

- Suppose X and Y are random variables, then we have
  - $\mu_{X+Y} = E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$
- If we further assume that X and Y independent of each other, then we have
  - $\sigma_{X+Y}^2 = Var(X+Y) = Var(X) + Var(Y) = \sigma_X^2 + \sigma_Y^2$

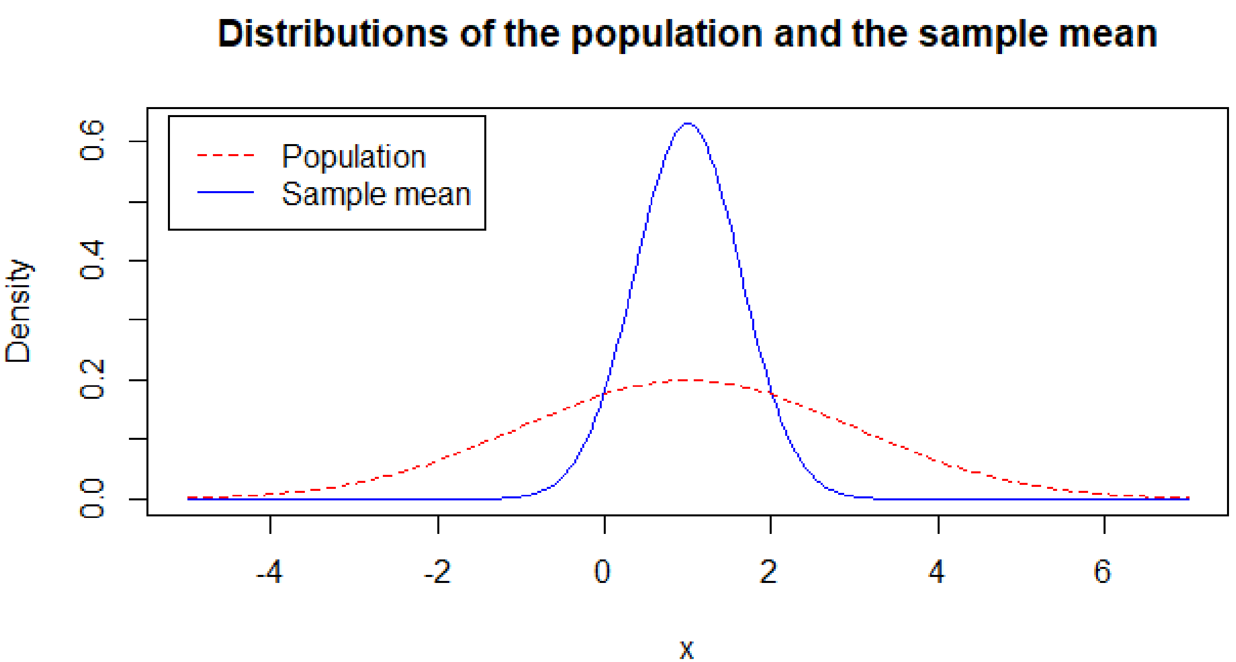
- Suppose random variables X and Y both follow the normal distribution. Then the linear combination $\alpha X + \beta Y$ follows the normal distribution, too, where α and β are constants.
- Example
  - Suppose $X_1, X_2, \cdots, X_n$ is a simple random sample of size n drawn from a normal distributed population. Then the sampling distribution of sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is also the normal distribution.

Mean and variance of a sample mean

- Application to the sampling distribution
  - Population mean and variance are μ and σ^2.
  - Simple random sample (SRS): $X_1, X_2, \cdots, X_n$
  - Sample mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - Assuming the sampling from a infinite population (or sampling with replacement from a finite population), then individual sample unit X_i has the same mean and variance as that of the population, and $X_1, X_2, \cdots, X_n$ are independent from each other
- Mean of a sample mean
  - $\mu_{\bar{X}} = E\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right) = \frac{1}{n}E(X_1 + X_2 + \cdots + X_n)$
  - $\frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right) = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu$
- Variance of a sample mean
  - $\sigma_{\bar{X}}^2 = Var\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right) = \left(\frac{1}{n}\right)^2 Var(X_1 + X_2 + \cdots + X_n)$
  - $= \frac{1}{n^2}(Var(X_1) + Var(X_2) + \cdots + Var(X_n))$
  - $= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \cdots + \sigma^2\right) = \frac{1}{n^2} \cdot n\sigma^2 = \sigma^2/n$

Sampling distribution of the sample mean $(\bar{X})$: Sampling from normally distributed populations

- Suppose $X_1, X_2, \cdots, X_n$ is a simple random sample of size n drawn from a normal distributed population with mean μ and variance σ^2. Then the sample mean $\bar{X} \sim N(\mu, \sigma^2)$
  - The distribution of $\bar{X}$ will be normal
  - The mean, $\mu_{\bar{X}}$, of the distribution of $\bar{X}$ will be equal to the mean of the population from which the sample was drawm.
  - The variance, $\sigma_{\bar{X}}^2$, of the distribution of $\bar{X}$ will be equal to the variance of the population divided by the sample size.



Population (red)
- Mean, $\mu = 1$
- Variance, $\sigma^2 = 4$

Sample mean (blue)
- Mean, $\mu_{\bar{X}} = 1$
- Variance, $\sigma_{\bar{X}}^2 = \frac{4}{10}$
- Sample size, $n = 10$

Sampling distribution of the sample mean $(\bar{X})$: Sampling from non-normally distributed populations

- The Central Limit Theorem
  - Suppose $X_1, X_2, \cdots, X_n$ is a simple random sample of size n drawn from a population with mean μ and variance σ^2. Then the sampling distribution of sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ will
    - have mean μ and variance $\frac{\sigma^2}{n}$,
    - be approximately normally distributed when sample size is large,
    - i.e.,

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right), \text{ when sample size } n \text{ is large,}$$

    - or equivalently

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1), \text{ when sample size } n \text{ is large.}$$

- How large does the sample have to be?
  - There is no one answer.
  - One rule of thumb: $n \geq 30$.

Sampling distribution of the sample mean $(\bar{X})$

- Summary
  - The sample mean is at least approximately normally distributed under the following conditions
    - The population distribution is normally distributed,
    - The population distribution is non-normally distributed and sample size is large,
    - The population distribution has an unknown functional form and the sample size is large.

Distribution of the Sample mean $\bar{X}$

- Example 5.3.2 (page 142)
  - In certain large human population cranial length is approximately normally distributed with a mean of 185.6(mm) and a standard deviation of 12.7(mm). What is the probability that a random sample of size 10 from this population will have a mean greater than 190?
- Solution
  - $\mu = 185.6$, $\sigma = 12.7$, $n = 10$, population distribution is normal
  - Sampling distribution is normal with the mean and SD being $\mu_{\bar{X}} = \mu = 185.6$ and $\sigma_{\bar{X}} = \sigma / \sqrt{n} = 12.7 / \sqrt{10} = 4.0161$
    i.e. $\bar{X} \sim N(85.6, 4.0161^2)$
  - Want to find $P(\bar{X} > 190) = 1 - P(\bar{X} \leq 190) = 1 - \Phi\left((190 - 185.6) / 4.0161\right)$
  - Stata command: disp 1 - normal((190-85.6)/4.0161)
  - $P(\bar{X} > 190) = .1366$

Note: Command for calculating $12.7 / \sqrt{10}$ is: dsip 12.7/sqrt(10)

Distribution of the Sample mean $\bar{X}$

- Example 5.3.3 (page 143)
  - If the mean and standard deviation of serum iron values for healthy women are 100 and 20 micrograms per deciliter (μg/dL), respectively, what is the probability that a random sample of 50 normal women will yield a mean between 95 and 110 μg/dL.
- Solution
  - $\mu = 100$, $\sigma = 20$, $n = 50$, population distribution is not specified. But sample size is large.
  - Sampling distribution of $\bar{X}$ is approximately normal with the mean $\mu_{\bar{X}} = 100$ and SD $\sigma_{\bar{X}} = \frac{20}{\sqrt{50}} = 2.8284$.
  - The probability we see is

$$P(95 \le \bar{X} \le 110) = P(\bar{X} \le 110) - P(\bar{X} \le 95) = \Phi\left(\frac{110-100}{2.8284}\right) - \Phi\left(\frac{95-100}{2.8284}\right)$$
$$= .9612$$

  - Stata Command
    - disp normal((110-100)/(20/sqrt(50))) - normal((95-100)/(20/sqrt(50)))
    - .9612

Summary

- Some useful properties of the random variables
  - $\mu_{\alpha \cdot X} = E(\alpha \cdot X) = \alpha \cdot E(X) = \alpha \cdot \mu_X$
  - $\sigma^2_{\alpha \cdot X} = Var(\alpha \cdot X) = \alpha^2 Var(X) = \alpha^2 \cdot \sigma^2_X$
  - $\mu_{X+Y} = \mu_X + \mu_Y,$
  - $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$, if X and Y are independent
  - $\alpha X + \beta Y \sim normal$, if $X \sim normal$ and $Y \sim normal$
  - The Central Limit Theorem
- Mean and variance of sample mean $\bar{X}$
  - $\mu_{\bar{X}} = \mu, \quad \sigma^2_{\bar{X}} = \sigma^2/n$
- Distribution of sample mean $\bar{X}$
  - $\bar{X} \sim N\left(\mu, \ \frac{\sigma^2}{n}\right)$ holds under the three conditions:
    - Population is normally distributed
    - Population is non-normally distributed and the sample size is large
    - Population has a unknown functional distribution form and the sample size is large

Homework 09

- Due date: August 31.
- Exercises 5.3.2 and 5.3.3 (Page 144, 10^{th} Ed)