Lecture 11

Sampling Distribution & Central Limit Theorem

Text: Chapters 4 & 5

STAT 8010 Statistical Methods I
September 24, 2020

Whitney Huang
Clemson University

Agenda

1. **Normal approximation of Binomial Distribution**
2. **Sampling Distribution**
3. **Central Limit Theorem (CLT)**

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large

- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1 - p) > 5$

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large
- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$
- If $X \sim \text{Bin}(n, p)$ with $np > 5$ and $n(1-p) > 5$ then we can use $X^* \sim \text{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large
- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$
- If $X \sim \text{Bin}(n, p)$ with $np > 5$ and $n(1-p) > 5$ then we can use $X^* \sim \mathrm{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathrm{P}(X^* = x) = 0 \ \forall x$

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large
- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$
- If $X \sim \text{Bin}(n, p)$ with $np > 5$ and $n(1-p) > 5$ then we can use $X^* \sim \mathrm{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathrm{P}(X^* = x) = 0 \; \forall x$
- **Continuity correction**: we use $\mathrm{P}(x - 0.5 \le X^* \le x + 0.5)$ to approximate $\mathrm{P}(X = x)$

Example

Suppose a class has 400 students (to begin with), that each student drops independently of any other student with a probability of .07. Let X be the number of students that finish this course

1. Find the probability that X is between 370 and 373 inclusive
2. Is an approximation appropriate for the number of students that finish the course?
3. If so, what is this distribution and what are the parameter(s)?
4. Find the probability that is between 370 and 373 inclusive by using the approximation

Sampling Distribution

- Independent random variables $X_1, X_2, \cdots, X_n$ with the same distribution are called a random sample

Sampling Distribution

- Independent random variables $X_1, X_2, \cdots, X_n$ with the same distribution are called a random sample

- A statistic is a function of a random sample

  Example:

Sampling Distribution

- Independent random variables $X_1, X_2, \cdots, X_n$ with the same distribution are called a random sample

- A statistic is a function of a random sample

  Example:

  - Sample mean: $\bar{X}_n = \sum_{i=1}^{n} X_i / n$

Sampling Distribution

- Independent random variables $X_1, X_2, \cdots, X_n$ with the same distribution are called a random sample
- A statistic is a function of a random sample

Example:

- Sample mean: $\bar{X}_n = \sum_{i=1}^n X_i / n$
- Sample variance: $\sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n-1)$

Sampling Distribution

- Independent random variables $X_1, X_2, \cdots, X_n$ with the same distribution are called a random sample

- A statistic is a function of a random sample

Example:

- Sample mean: $\bar{X}_n = \sum_{i=1}^n X_i / n$

- Sample variance: $\sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n-1)$

- Sample maximum: $\max_{i=1}^n X_i$

Sampling Distribution

- Independent random variables $X_1, X_2, \cdots, X_n$ with the same distribution are called a random sample
- A statistic is a function of a random sample

  Example:
  - Sample mean: $\bar{X}_n = \sum_{i=1}^{n} X_i / n$
  - Sample variance: $\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 / (n-1)$
  - Sample maximum: $\max_{i=1}^{n} X_i$
- The probability distribution of a statistic is called its sampling distribution

Example

Suppose $X_1, X_2, \cdots, X_n$ is a random sample from a $\mathrm{N}(\mu, \sigma^2)$ population, Find the sampling distribution of sample mean.

Example

Suppose $X_1, X_2, \cdots, X_n$ is a random sample from a $\mathrm{N}(\mu, \sigma^2)$ population, Find the sampling distribution of sample mean.

$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} = \sum_{i=1}^n \frac{1}{n} X_i$. From last lecture we know that sum of normal r.v.s is still a normal r.v. Hence we only need to figure its mean and variance.

$$\mathrm{E}[\bar{X}_n] = \sum_{i=1}^n \frac{1}{n}\mu = \mu$$

$$\mathrm{Var}[\bar{X}_n] = \sum_{i=1}^n \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

Therefore, we have $\bar{X}_n \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$

Central Limit Theorem (CLT)

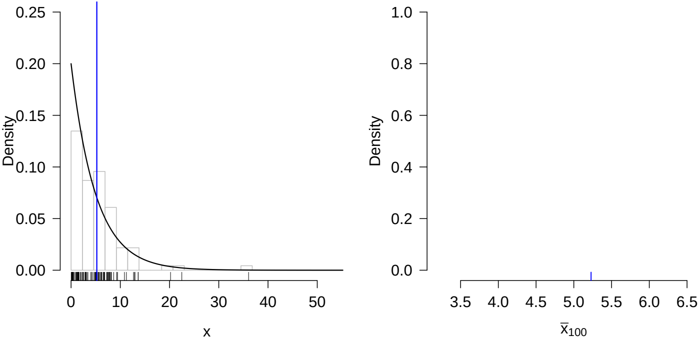
CLT

The **sampling distribution** of the **mean** will become approximately **normally distributed** as the **sample size becomes larger**, **irrespective of the shape of the population distribution!**

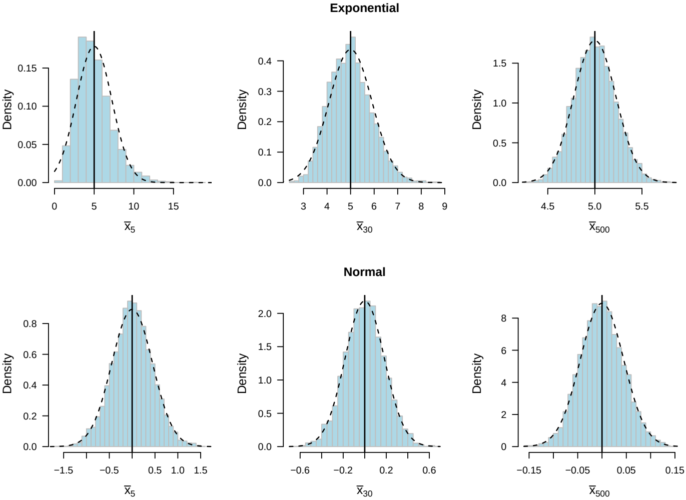
Let $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} F$ with $\mu = \mathrm{E}[X_i]$ and $\sigma^2 = \mathrm{Var}[X_i]$.
Then $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{d} \mathrm{N}(\mu, \frac{\sigma^2}{n})$ as $n \to \infty$.

CLT In Action

1. Generate 100 (n) random numbers from an Exponential distribution (population distribution)
2. Compute the sample mean of these 100 random numbers
3. Repeat this process 120 times



CLT: Sample Size (n) and the Normal Approximation



Why CLT is important?

- In many cases, we would like to make statistical inference about the population mean μ
  - The sample mean $\bar{X}_n$ is a sensible estimator for the population mean
  - CLT tells us the **distribution** of our estimator $\Rightarrow \bar{X}_n \approx \mathrm{N}(\mu, \frac{\sigma^2}{n})$
- Applications: Confidence Interval, Hypothesis Testing