Lecture 15

Normal Distribiutions II

Text: Chapter 4

STAT 8010 Statistical Methods I

September 23, 2019

Whitney Huang
Clemson University

Agenda

1. Sums of Normal Random Variables
2. Normal approximation of Binomial Distribution
3. Sampling Distribution and Central Limit Theorem (CLT)

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^{n} X_i$ then $S_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$

Sums of Normal Random Variables

If X_i $1 \le i \le n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^n X_i$ then $S_n \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^{n} X_i$ then $S_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$
- This can be applied for any integer n

Example

Let X_1, X_2, and X_3 be mutually independent, Normal random variables. Let their means and standard deviations be $3k$ and k for $k = 1$, 2, and 3 respectively. Find the following distributions:

1. $\sum_{i=1}^{3} X_i$
2. $X_1 + 2X_2 - 3X_3$
3. $X_1 + 5X_3$

Example Cont'd

1. $\sum_{i=1}^{3} X_i$

   $\sum_{i=1}^{3} X_i \sim N(\mu = 3+6+9 = 15, \sigma^2 = 1^2+2^2+3^2 = 14)$

Example Cont'd

1. $\sum_{i=1}^{3} X_i$

   $\sum_{i=1}^{3} X_i \sim N(\mu = 3 + 6 + 9 = 15, \sigma^2 = 1^2 + 2^2 + 3^2 = 14)$

Example Cont'd

1. $\sum_{i=1}^{3} X_i$

   $\sum_{i=1}^{3} X_i \sim N(\mu = 3 + 6 + 9 = 15, \sigma^2 = 1^2 + 2^2 + 3^2 = 14)$

2. $X_1 + 2X_2 - 3X_3$

   $X_1 + 2X_2 - 3X_3 \sim N(\mu = 3 + 12 - 27 = -12, \sigma^2 = 1^2 + 4 \times 2^2 + 9 \times 3^2 = 98)$

Example Cont'd

1. $\sum_{i=1}^{3} X_i$

   $\sum_{i=1}^{3} X_i \sim N(\mu = 3 + 6 + 9 = 15, \sigma^2 = 1^2 + 2^2 + 3^2 = 14)$

2. $X_1 + 2X_2 - 3X_3$

   $X_1 + 2X_2 - 3X_3 \sim N(\mu = 3 + 12 - 27 = -12, \sigma^2 = 1^2 + 4 \times 2^2 + 9 \times 3^2 = 98)$

CLEMSON UNIVERSITY

Example Cont'd

1. $\sum_{i=1}^{3} X_i$

   $\sum_{i=1}^{3} X_i \sim N(\mu = 3+6+9 = 15, \sigma^2 = 1^2 + 2^2 + 3^2 = 14)$

2. $X_1 + 2X_2 - 3X_3$

   $X_1 + 2X_2 - 3X_3 \sim N(\mu = 3 + 12 - 27 = -12, \sigma^2 = 1^2 + 4 \times 2^2 + 9 \times 3^2 = 98)$

3. $X_1 + 5X_3$

   $X_1 + 5X_3 \sim N(\mu = 3 + 45 = 48, \sigma^2 = 1^2 + 25 \times 3^2 = 226)$

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

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 $n(1-p) > 5$

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 $X^* \sim \text{N}(\mu = np, \sigma^2 = np(1-p))$ 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 $X^* \sim \text{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathbb{P}(X^* = x) = 0 \; \forall x$

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 $X^* \sim \text{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathbb{P}(X^* = x) = 0 \ \forall x$

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 $X^* \sim \text{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathbb{P}(X^* = x) = 0 \; \forall x$
- **Continuity correction**: we use $\mathbb{P}(x - 0.5 \leq X^* \leq x + 0.5)$ to approximate $\mathbb{P}(X = x)$

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 (if an approximation appropriate)

Sampling Distribution & Central Limit Theorem (CLT)

Sampling distribution: the probability distribution of a given **random-sample-based** statistic

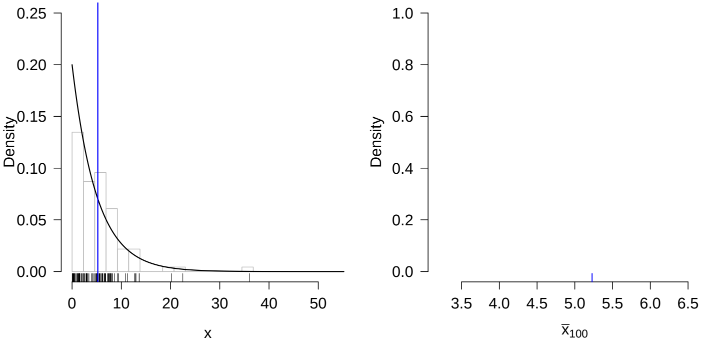
CLT

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!**

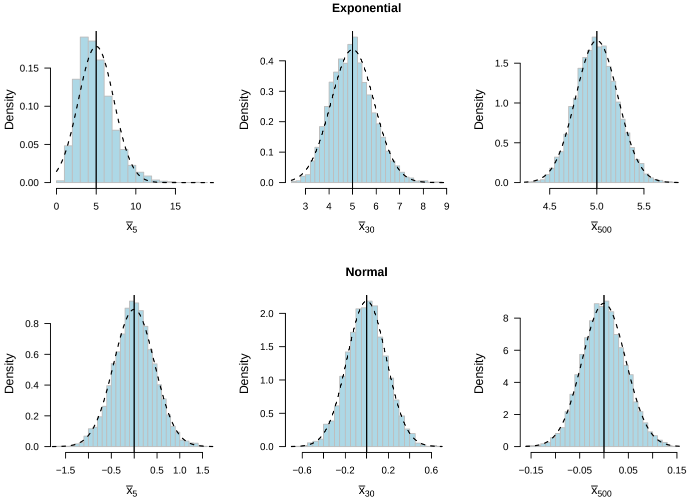
Let $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} F$ with $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}[X_i]$.
Then $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{d} \mathrm{N}(\mu, \frac{\sigma^2}{n})$ as $n \to \infty$.

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 $\Rightarrow \bar{X}_n \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$
- Applications: Hypothesis testing, confidence interval