Chapter 37: Central Limit Theorem

Meike Niederhausen and Nicky Wakim

2024-12-01

Learning Objectives

1. Calculate probability of a sample mean using a population mean and variance with unknown distribution
2. Use the Central Limit Theorem to construct the Normal approximation of the Binomial and Poisson distributions

The Central Limit Theorem

Theorem 1: Central Limit Theorem (CLT)

Let X_i be iid rv's with common mean μ and variance σ^2, for $i = 1, 2, \ldots, n$. Then

$$\sum_{i=1}^{n} X_i \to \mathrm{N}(n\mu, n\sigma^2)$$

Converges in distribution as $n \to \infty$

- X_i's do NOT need to be normally distrib.
- do NOT need to know dist of X
- when n is large, we can use the Normal approx

Extension of the CLT

Corollary 1

Let X_i be iid rv's with common mean μ and variance σ^2, for $i = 1, 2, \ldots, n$. Then

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \to \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Example of Corollary in use

$n \geq 30$, then can Normal approx

$\mu = 74 \quad \sigma = 13 \quad n = 36 \quad \bar{x} = 77$ or 71

Example 1

According to a large US study, the mean resting heart rate of adult women is about 74 beats per minutes (bpm), with standard deviation 13 bpm (NHANES 2003-2004).

1. Find the probability that the average resting heart rate for a random sample of 36 adult women is more than 3 bpm away from the mean.
2. Repeat the previous question for a single adult woman.

① $P(\bar{X} < 71 \text{ or } \bar{X} > 77)$

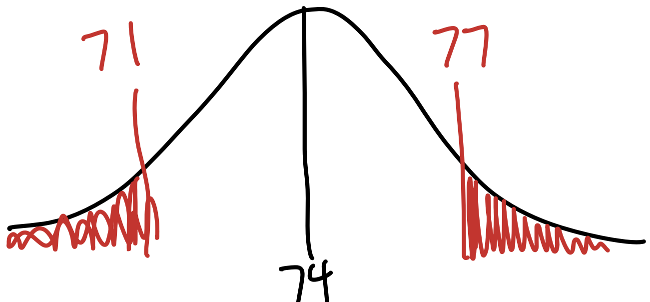
$\left(\bar{X} \sim N\left(74, sd = \frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{36}}\right) \right)$

$P(\bar{X} < 71) + P(\bar{X} > 77)$

$= 2P(\bar{X} < 71)$

$= 2 \cdot \text{pnorm}(x = 71, \text{mean} = 74, sd = 13/\text{sqrt}(36))$

$= 0.166$



② $n = 1 < 30$, cannot normal approx

Example of CLT for exponential distribution

Example 2 $E(X_i) = \frac{1}{\lambda} = \mu$

Let $X_i \sim Exp(\lambda)$ be iid RVs for $i = 1, 2, \ldots, n$. Then $Var(X_i) = \frac{1}{\lambda^2} = \sigma^2$

$$\sum_{i=1}^{n} X_i \rightarrow$$

$$\sum_{i=1}^{n} X_i \rightarrow N(n\mu, n\sigma^2)$$

converges in distribution as $n \to \infty$

$$n\mu = \frac{n}{\lambda}$$

$$n\sigma^2 = \frac{n}{\lambda^2}$$

$$\sum_{i=1}^{n} X_i \rightarrow N\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)$$

CLT for Discrete RVs

1. Binomial rv's: Let $X \sim Bin(n, p)$ ✓

- $X = \sum_{i=1}^{n} X_i$, where X_i are iid $\boxed{\text{Bernoulli}(p)}$
- Rule of thumb: $np \geq 10$ and $n(1-p) \geq 10$ to use Normal approximation

$$\sum_{i=1}^{n} X_i \longrightarrow N(np, np(1-p))$$

$$\hat{p} \longrightarrow N\left(p, \frac{p(1-p)}{n}\right)$$

2. Poisson rv's: Let $X \sim Poisson(\lambda)$ ✓

- $X = \sum_{i=1}^{n} X_i$, where X_i are iid $\text{Poiss}(1)$
- Recall from Chapter 18 that if $X_i \sim Poiss(\lambda_i)$ and X_i independent, then $\sum_{i=1}^{n} X_i \sim Poiss(\sum_{i=1}^{n} \lambda_i)$
- Rule of thumb: $\lambda \geq 10$ to use Normal approximation

$$\sum_{i=1}^{n} X_i \rightarrow N(\lambda, \sigma = \lambda)$$

$$\downarrow$$

$$\text{var} = \lambda^2$$

At home example

Example 3

Suppose that the probability of developing a specific type of breast cancer in women aged 40-49 is 0.001. Assume the occurrences of cancer are independent. Suppose you have data from a random sample of 20,000 women aged 40-49.

1. How many of the 20,000 women would you expect to develop this type of breast cancer, and what is the standard deviation?
2. Find the **exact** probability that more than 15 of the 20,000 women will develop this type of breast cancer.
3. Use the CLT to find the **approximate** probability that more than 15 of the 20,000 women will develop this type of breast cancer.
4. Use the CLT to approximate the following probabilities, where X is the number of women that will develop this type of breast cancer.
    a. $\mathbb{P}(15 \leq X \leq 22)$
    b. $\mathbb{P}(X > 20)$
    c. $\mathbb{P}(X < 20)$
5. Find the **approximate** probability that more than 15 of the 20,000 women will develop this type of breast cancer - not using the CLT!
6. Use the CLT to approximate the approximate probability in the previous question!