Chapter 37: Central Limit Theorem

Meike Niederhausen and Nicky Wakim

2023-11-29

Table of contents

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

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

converges in distribution as $n \to \infty$

X_is do NOT need to be normal

Do NOT need to know the distribution of X_i

when n is large, we can use a Normal approximation

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

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

variance

$$SE = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Example of Corollary in use

$\mu = \underline{74} \quad \sigma = 13 \quad n = 36$ in sample

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.

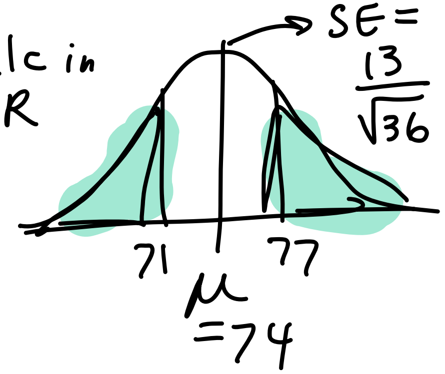
① $P(\bar{X} < 71 \ \underline{\text{or}} \ \bar{X} > 77)$

$$\bar{X} \sim N\Big(\underset{\mu}{\overset{\parallel}{74}},\ \overset{sd}{\frac{\sigma}{\sqrt{n}}} = \frac{13}{\sqrt{36}}\Big)$$

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

$$= 2P(\bar{X} < 71) \quad \text{calc in R}$$

$$= 0.166$$



Rule of thumb for unknown dist'n?

Example of CLT for exponential distribution

Example 2

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

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

$$\left.\mu_{X_i} = \frac{1}{\lambda} \qquad \sigma_{X_i}^2 = \frac{1}{\lambda^2}\right] \text{ b/c exp}$$

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

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

CLT for Discrete RVs

1. **Binomial rv's**: Let $X \sim \text{Bin}(n, p)$

- $X = \sum_{i=1}^{n} X_i$, where X_i are iid Bernoulli(p)

$\hookrightarrow \mu_{X_i} = p$

$\sigma^2_{X_i} = p(1-p)$

$p(q)$

Rule of Thumb: $npq \geq 10$ to use Normal approx

$$\sum_{i=1}^{n} X_i \longrightarrow N(n\mu_{X_i}, n\sigma^2_{X_i})$$

$$\longrightarrow N(np, \overset{\text{var}}{npq})$$

ALSO $\hat{p} = \frac{\Sigma x_i}{n} \rightarrow N(p, \underset{\text{var}}{\frac{pq}{n}})$

2. **Poisson rv's**: Let $X \sim \text{Poisson}(\lambda)$

- $X = \sum_{i=1}^{\lambda} X_i$, where X_i are iid Poiss(1)
- Recall from Chapter 18 that if $X_i \sim \text{Poiss}(\lambda_i)$ and X_i independent, then $\sum_{i=1}^{n} X_i \sim \text{Poiss}(\sum_{i=1}^{n} \lambda_i)$

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

Rule of Thumb: $\lambda \geq 10$ to use Normal approx.

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!