Chapter 7 - Normal Probability Approximations

STAT 251

Lecture 22

Central Limit Theorem (CLT)

Dr. Lasantha Premarathna

Chapter 7 - Learning Outcomes

- Statistic and parameter
- Sampling distribution
- Central Limit Theorem (CLT)
- Normal Approximation to the Binomial distribution
- Normal Approximation to the Poisson distribution

Population and Sample

- **Population** is the entire collection of individuals we want to study
- **Sample** is a subset of individuals selected from the population

** Statistics techniques are used to make conclusions about the population based on the sample.

Random sample → unbiased, representative population

Statistic and Parameter

- A **statistic** is a numerical summery of the sample

  e.g. sample mean ($\bar{x}$)

  sample standard deviation (s)

- A **parameter** ia a numerical summary of a population

  e.g. population mean (μ) → unknown - how to estimate.

  population standard deviation (σ)

just use sample val.

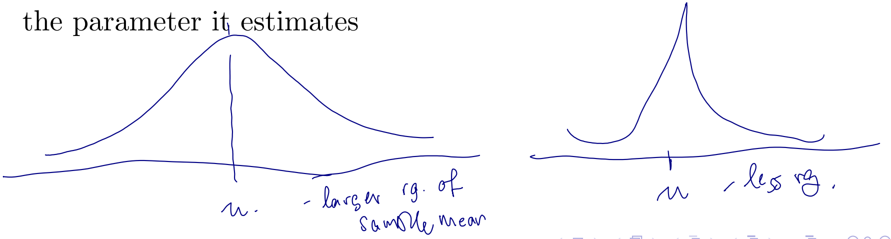
- Value of the parameter is unknown in practice.
- Due to sampling variability, a statistic takes on different values for different samples. → Many sample means / s.d.
- Parameters are estimated using sample data. We use statistics to estimate parameters.

Sampling Distributions

- The sampling distribution of a statistic is the probability distribution that specifies probabilities for the possible values the statistic can take

- Sampling distribution describes the variability that occurs from study to study using statistics to estimate population parameters.

- Sampling distributions help to predict how close a statistic falls to the parameter it estimates



Results from Ch 4 & 5: Linear combination of Normal rvs

IID: Idependent & Identical.

- If $X_1,\ X_2,\ \cdots,\ X_n$ is a random sample from a normal population with mean $= \mu$ and variance $= \sigma^2$ then,

$$\bar{X} = \frac{X_1\ +\ X_2\ +\ \cdots\ +\ X_n}{n}\ \sim\ N\left(\mu,\ \frac{\sigma^2}{n}\right)$$

- often, however, we have to deal with non-normal populations

Mean and Variance of the Sampling distribution of sample mean

Suppose a random sample of n observations is taken from a population with mean $= \mu$ and variance $= \sigma^2$ then,

- the mean of the of the Sampling distribution of sample mean is μ (same as the population mean)

- the standard deviation of the Sampling distribution of sample mean is $\frac{\sigma}{\sqrt{n}}$ (i.e. standard error $= \frac{\sigma}{\sqrt{n}}$) estimate.

** To distinguish the standard deviation of a sampling distribution from the standard deviation of an ordinary probability distribution, we refer it to as a standard error.

Central Limit Theorem (CLT)

- Let $X_1,\ X_2,\ \cdots,\ X_n$ be a random sample from an arbitrary population/distribution with mean μ and variance σ^2. When n is large (book says $n \geq$ 30 (good rule) 20), then $\bar{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ regardless of the actual shape of the population distribution

$$\bar{X} = \frac{X_1\ +\ X_2\ +\ \cdots\ +\ X_n}{n}\ \sim\ N\left(\mu,\ \frac{\sigma^2}{n}\right)\ ,\quad \textit{approximately}$$

** How the population mean and standard deviation are related to the mean and standard deviation of the sampling distribution of the sample mean

Explain this using applets

CLT for sum of rvs when n is large

- Consider a random sample $X_1,\ X_2,\ \cdots,\ X_n$ from a distribution with mean μ and variance σ^2.
  When n is large, by CLT
  $$\bar{X} = \frac{\sum_i X_i}{n} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right), \quad \textit{approximately}$$
- If a question is about sum instead an average, still we can use CLT
  $$T = X_1 + X_2 + \cdots + X_n = \sum_i^n X_i$$
  $$\text{then, } \bar{X} = \frac{\sum_i X_i}{n} \Rightarrow \bar{X} = \frac{T}{n}$$
  $$\Rightarrow \quad T = n\bar{X}$$
  $$E(T) = E(n\bar{X}) = nE(\bar{X}) = n\mu$$
  $$Var(T) = Var(n\bar{X}) = n^2 Var(\bar{X}) = n^2\,\frac{\sigma^2}{n} = n\sigma^2$$
- Therefore, $T \sim N\left(n\mu,\ n\sigma^2\right), \quad \textit{approximately}$

Online Resources to Learn CLT

- **Video: Sampling distribution of the mean** - UBC Flexible Learning Introductory Statistics Project

  This video explores the concept of a sampling distribution of the mean. It highlights how we can draw conclusions about a population mean based on a sample mean by understanding how sample means behave when we know the true values of the population.

  https://vimeo.com/196027417

Applet: Sampling from a non-Normally distributed population (CLT)

- **Applet: Sampling from a non-Normally distributed population (CLT)** - UBC StatSpace

  This web visualization explores the sampling distribution of the mean when the data do not necessarily follow a Normal distribution.

  http://www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm

- Also notice that when n is increasing, the standard deviation of the sampling distribution of sample mean is decreasing.

Applet: Sampling from a normally distributed population

- **Applet: Sampling from a normally distributed population** - UBC StatSpace

  This web visualization demonstrates the concept of a sampling distribution of an estimate, using the example of a mean of a Normally distributed variable. It also reinforces the idea of a histogram.

  http://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm

Before the next class ...

Visit the course website at canvas.ubc.ca

- Review Lecture 22 and related sections in the text book
- Topic of next class: **CLT Examples and Normal approximation to Binomial and Poisson distributions**