Chapter 7. Sampling Distributions and the Central Limit Theorem

1 Introduction

2 Sampling Distributions related to the normal distribution

3 The central limit theorem

4 The normal approximation to binomial distribution

1. Introduction

Assume that $Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with the common distribution. Suppose one is interested in estimating a population mean μ from observed samples $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$.

Then one might want to use the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ to estimate μ. Given observed samples, $\bar{y}$ is just a single number. Then how one can know the goodness of this estimate $\bar{y}$ for μ?

Note that $\bar{y}$ is calculated from the formula

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

a function of the observable r.v.'s $Y_1, Y_2, \ldots, Y_n$ and the (constant) sample size n.

Since $\bar{Y}$ is also a r.v. itself, it has the probability distribution. If one knows the probability distribution of $\bar{Y}$, one can know the goodness of $\bar{Y}$ for μ.

Assumption

"$Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with probability mass function $p(y)$ or probability density function $f(y)$"

$\Leftrightarrow$ the random variable (r.v.)s $Y_1, Y_2, \ldots, Y_n$ are independent with common probability mass function $p(y)$ or common density function $f(y)$

$\Leftrightarrow$ $Y_1, \ldots, Y_n \overset{iid}{\sim} p(y)$ or $f(y)$

(Def 7.1)

A *statistic* is a function of the observable random variables in a sample and known constants for a parameter of our interest. A statistic itself is a random variable.

(e.g.) $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

(Def 7.2)

The *sampling distribution* of the statistic is the probability distribution for the statistic (or the distribution of the statistic for all possible samples of a given size).

(e.g.) Sampling distribution of $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$?

How one can obtain the sampling distribution of a statistic?

[M1] The sampling distribution of a statistic is the probability distribution, under repeated sampling of the population, of a given statistic.

(Example 1) The sample $\bar{Y}$ is to be calculated from a random sample of size 2 taken from a population consisting of ten values (2,3,4,5,6,7,8,9,10,11). Find the sampling distribution of $\bar{Y}$, based on a random sample of size 2.

There are 45 possible samples of two items selected from the ten items(see [Table 1]). Assuming each sample of size 2 is equally likely, [Table 2] shows the sampling distribution for $\bar{Y}$ based on $n = 2$ observations selected from the population (2,3,4,5,6,7,8,9,10,11).

(Example 2) Consider a large normal population. Assume we repeatedly take samples of a given size from the population and calculate the sample mean of the data values for each sample, $\bar{y}$. Different samples will lead to different sample means. The distribution of these means is the "sampling distribution of $\bar{Y}$" (for the given sample size).

[M2] One can mathematically derive the sampling distribution of a statistic if one knows the distribution of the random variables $Y_1, \ldots, Y_n$ (using Chapter 6)

[Table 1] for (Example 1)

Sample	$\bar{y}$	Sample	$\bar{y}$	Sample	$\bar{y}$	Sample	$\bar{y}$
2,3	2.5	2,4	3	2,5	3.5	2,6	4
2,7	4.5	2,8	5	2,9	5.5	2,10	6
2,11	6.5	3,4	3.5	3,5	4	3,6	4.5
3,7	5	3,8	5.5	3,9	6	3,10	6.5
3, 11	7	4,5	4.5	4,6	5	4,7	5.5
4,8	6	4,9	6.5	4,10	7	4,11	7.5
5,6	5.5	5,7	6	5,8	6.5	5,9	7
5,10	7.5	5,11	8	6,7	6.5	6,8	7
6,9	7.5	6,10	8	6,11	8.5	7,8	7.5
7,9	8	7,10	8.5	7,11	9	8,9	8.5
8,10	9	8,11	9.5	9,10	9.5	9,11	10
10,11	10.5						

[Table 2] for (Example 1)

$\bar{y}$	2.5	3	3.5	4	4.5	5	5.5	6	6.5
$p(\bar{y})$	$\frac{1}{45}$	$\frac{1}{45}$	$\frac{2}{45}$	$\frac{2}{45}$	$\frac{3}{45}$	$\frac{3}{45}$	$\frac{4}{45}$	$\frac{4}{45}$	$\frac{5}{45}$

$\bar{y}$	7	7.5	8	8.5	9	9.5	10	10.5
$p(\bar{y})$	$\frac{4}{45}$	$\frac{4}{45}$	$\frac{3}{45}$	$\frac{3}{45}$	$\frac{2}{45}$	$\frac{2}{45}$	$\frac{1}{45}$	$\frac{1}{45}$

This sampling distribution provides a way to make statistical inferences about $\bar{Y}$ in the example: calculate the following probability: $P(3.5 \leq \bar{Y} \leq 9.5) = 41/45$.

2. Sampling Distributions related to the normal distribution

In many applied problems it is reasonable to assume that the observed random variables in a random sample, $Y_1, Y_2, \ldots, Y_n$, are independent with a common, normal density function. In this section, we will develop the sampling distributions of various statistics calculated by using the observations in a random sample from a normal population(or independent random samples from two normal populations).

Inference about μ of a normal population with *known variance* σ^2

(Theorem 7.1)
Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N\left(\mu, \sigma^2/n\right).$$

$$Z \equiv \frac{\bar{Y} - \mu_{\bar{Y}}}{\sigma^2_{\bar{Y}}} = \sqrt{n}\left(\frac{\bar{Y} - \mu}{\sigma}\right) \sim N(0, 1).$$

(Example 7.2) A bottling machine can be regulated so that it discharges an average of μ ounces per bottle. It has been observed that the amount of fill dispensed by the machine is normally distributed with $\sigma = 1.0$ ounce. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine a given day and the ounces of fill machined for each. Find the probability that the sample mean will be within 0.3 ounce of the true mean μ for the particular setting.

(Example 7.3) In Example 7.2, how many observations should be included in the sample if we wish $\bar{Y}$ to be within 0.3 ounce of μ with (at least) probability 0.95?

3. The central limit theorem

By Theorem 5.12, $E(\bar{Y}) = \mu$ and $V(\bar{Y}) = \sigma^2/n$ if $Y_1, Y_2, \ldots, Y_n$ represents a random sample from *any* distribution with mean μ and variance σ^2.

If one samples from a normal distribution, $\bar{Y}$ has a normal distribution(Theorem 7.1).

[Question]
But what can we say about the sampling distribution of $\bar{Y}$ if the variables Y_i are not normally distributed?

[Answer]
Under some conditions, $\bar{Y}$ will have a sampling distribution that is approximately normal as long as the sample size is large.
In this section we will develop an approximation for the sampling distribution of $\bar{Y}$ that can used **regardless of the distribution of the population from which the sample is taken** : the *Central Limit Theorem*.

(Theorem 7.4)

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables with $E(Y_i) = \mu < \infty$ and $V(Y_i) = \sigma^2 < \infty$. Define

$$U_n \equiv \frac{\bar{Y} - E(\bar{Y})}{\sqrt{Var(\bar{Y})}} = \sqrt{n}\left(\frac{\bar{Y} - \mu}{\sigma}\right)$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Then the distribution function of U_n converges to a standard normal distribution function as $n \to \infty$.

Note that

- $P(a \leq \bar{Y} - \mu \leq b) = P\left(\frac{a}{\sigma/\sqrt{n}} \leq U_n \leq \frac{b}{\sigma/\sqrt{n}}\right)$ $\approx P\left(\frac{a}{\sigma/\sqrt{n}} \leq Z \leq \frac{b}{\sigma/\sqrt{n}}\right)$ for large n where $Z \sim N(0, 1)$.

- The central limit theorem can be applied to a random sample $Y_1, Y_2, \ldots, Y_n$ from *any* distributions, so long as $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$ are both finite and the sample size is large.

(Example 7.8) Achievement test scores of all high school seniors in a state have mean 60 and variance 64. A random sample of $n = 100$ students from one large high school had a mean score of 58. Is there evidence to suggest that this high school is inferior?(Calculate the probability that the sample mean is at most 58 when $n = 100$)

(Example 7.9) The service time for customers through a checkout counter in a retail store are independent random variable with mean 1.5 minutes and variance 1.0. Approximate the probability that 100 customers can be served in less than 2 hours of total service time.

4. The normal approximation to binomial distribution

The central limit theorem also can be used to approximate probabilities for some discrete random variables when the exact probabilities are tedious to calculate. One useful example involves the binomial distribution for large values of the number of trials, n.

Suppose that Y has a binomial distribution with n trials and probability of success on any one trial denoted by p. How we can obtain $P(Y \leq b)$?

[M1]. $P(Y \leq b) = \sum_{i=0}^{b} P(Y = b)$ where $Y \sim b(n, p)$. For some values of the sample size n, tables are available, but direct calculation is tedious for large values of n for which tables may be not available.

[M2]. We can use the central limit theorem for large values of n : we can think Y, the number of successes in n trials, as a sum of a sample consisting of 0s and 1s; that is

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

where

$$X_i = \begin{cases} 1 & \text{if ith trial results in success,} \\ 0 & \text{otherwise.} \end{cases}$$

The X_i for $i = 1, 2, \ldots, n$ are independent Bernoulli random variables, and X_i has $E(X_i) = p$ and $V(X_i) = p(1-p)$ for $i = 1, 2, \ldots, n$. Consequently, when n is large, the sample fraction of successes,

$$\frac{Y}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

possesses an approximately normal sampling distribution with mean $E(Y/n) = E(X_i) = p$ and variance $V(Y/n) = V(X_i)/n = p(1-p)/n$.

Thus, by the central limit theorem, we can think that if $Y \sim b(n, p)$ and n *is large*, then Y/n has an approximately normal sampling distribution with mean $E(Y/n) = p$ and variance $V(Y/n) = p(1-p)/n$ (in other words, Y possesses an approximately normal sampling distribution with mean $E(Y) = np$ and variance $V(Y) = np(1-p)$)

(Example 7.10)Candidate A believes that she can win a city election if she can earn at least 55% of the votes in precinct I. She also believes that about 50% of the city's voters favor her. If $n = 100$ voters show up to vote at precinct I, what is the probability that candidate A will receive at least 55% of their votes?

In this approximation,
(1) One tries to approximate a discrete distribution represented by a histogram with a continuous density function.

(2) Slight adjustment on the boundaries (called 0.5 continuity correction) can lead to substantial improvement in the approximation.

(Example) Suppose $Y \sim B(6, 5)$. Calculate $P(2 \leq Y \leq 4)$

(Example 7.10 revisited)