STATISTICS WITH APPLIED PROBABILITY

Custom eBook for STA258

Nishan Mudalige

Nurlana Alili

Bryan Su

CANADA

Statistics with Applied Probability
Custom eBook for STA258

Nishan Mudalige
Department of Mathematical and Computational Sciences
University of Toronto Mississauga

Nurlana Alili
University of Toronto Mississauga

Bryan Xu
University of Toronto Mississauga

© 2025 N. Mudalige, N. Alili, B. Xu
All rights reserved.

This work may not be copied, translated, reproduced or transmitted in any form or by any means — graphic, electronic or mechanical including but not limited to photocopying, scanning, recording, microfilming, electronic file sharing, web distribution or information storage systems — without the explicit written permission of the authors.

Every effort has been made to trace ownership of all copyright material and to secure permission from copyright holders. In the event of any question arising as to the use of copyright material, we will be pleased to make necessary corrections in future publications.

First edition: August 2025

Mudalige, M.; Alili, N.; Xu, B.
University of Guelph Bookstore Press

University of Toronto Mississauga,
Mississauga, Ontario, Canada

Contents

Chapter 0

Overview

Uncertainty is an inherent part of everyday life. We all face questions regarding uncertainty such as whether classes will go ahead as planned on any given day; will a flight leave on time; will a student pass a certain course? Uncertainties might also change depending on other factors, such as whether classes will still go ahead as planned when there is a snow warning in effect; if a flight is delayed can a person still manage to make their connection; will a student pass their course considering that the instructor is known to be a tough grader?

The ability to quantify uncertainty using rigorous mathematics is a powerful and useful tool. Calculating uncertainty on an intuitive level is something that is hard-wired in our DNA, such as the decision to fight or flight depending on a given set of circumstances. However we cannot always make such intuitive decisions based purely on hunches and gut feelings. Fortunes have been lost based on someone having a good feeling about something. If we have information available, we should make the best prediction possible using this information. For instance if we wanted to invest a lot of money in a company, we should use all available data such as past sales, market and industry trends, leadership ability of the CEO, forward looking statements etc. and with all this information we can then predict whether our investment will be profitable.

In order for companies to survive and remain competitive in todays environment it is essential to monitor industry trends and read markets properly. Companies that don't adapt and stick to an outdated business model tend to pay the price. At the other end of the spectrum, companies that understand the needs of the consumer, build their product around the consumer and keep evolving their product offerings based on consumer trends tend to perform well and remain competitive.

Statistics is the science of uncertainty and it is clearly a very useful subject for business. In this book you will be given an introduction to statistics and you will learn the framework as well as the language required at the introductory level. The material may be daunting at times, but the more you get familiar with the subject the more comfortable you will become with it. As business students, doing well in a statistics course will give you a competitive edge since the ability to interpret and perform quantitative analytics are skills that are highly desired by many employers.

Chapter 1

Introduction

1.1 Basics

Intuitively, statistics can be considered the science of uncertainty. Formally,

Definition 1.1 (Statistics).

Statistics is the science of collecting, classifying, summarizing, analyzing and interpreting data.

more information goes here
anything

Chapter 4

Normal Approximation to the Binomial Distribution

4.1 Introduction

Definition 4.1 (Statistic)**.**
A statistic is a function of the observable random variables in a sample and known constants. Since statistics are functions of the random variables observed in a sample, they themselves are random variables. As such, all statistics have a corresponding probability distribution, which we refer to as their sampling distribution.

Review from STA256

Bernoulli Distribution:
A Bernoulli trial is a single experiment with two outcomes:

- Success: $X = 1$ with probability p
- Failure: $X = 0$ with probability $1 - p$

$X = x$	0	1
$P(X = x)$	$1 - p$	p

The probability mass function (PMF) is:

$$f(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}$$

Binomial Distribution:
A binomial distribution arises from n independent Bernoulli trials. Let:

$$X = \text{number of successes in } n \text{ trials}$$

Then:

$$X \sim \text{Binomial}(n, p)$$

where:

- Each trial results in either success (with probability p) or failure (with probability $1 - p$)
- $X \in \{0, 1, \ldots, n\}$

The PMF is:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Moment Generating Function (MGF):
The moment generating function (MGF) of a random variable X is defined as:

$$M_X(t) = \mathbb{E}[e^{tX}]$$

The MGF uniquely characterizes the distribution of X (if it exists in an open interval around 0), and it can be used to compute moments such as the mean and variance.

4.2 Bernoulli Distribution

Bernoulli random variable is a discrete random variable that has exactly two possible outcomes which are either a **success** or a **failure**. An experiment in which there are exactly 2 outcomes (which are success or failure) is called a **Bernoulli trial**.

When $x = 1$ we have a success and when $x = 0$ we have a failure. The term success and failure are relative to the problem being studied.

TIP: "success" need not be something positive

We chose to label a person who refuses to administer the worst shock a "success" and all others as "failures". However, we could just as easily have reversed these labels. The mathematical framework we will build does not depend on which outcome is labeled a success and which a failure, as long as we are consistent.

Consider the random experiment of rolling a die once. Define the random variable:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th roll is a six,} \\ 0 & \text{otherwise} \end{cases}$$

Then $X_i \sim \text{Bernoulli}(p)$, where $p = P(\text{rolling a six})$.

Let $X \sim$ Bernoulli(p). The mass function of X is

$$P(X = x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

where p represents the probability of success.

Definition 4.2 (Mean and Variance of a Bernoulli Random Variable)**.**
Let $X \sim$ Bernoulli(p). The mean of X is

$$E(X) = \mu = p$$

and the variance of X is

$$Var(X) = \sigma^2 = p(1-p)$$

To support the earlier result, we now provide a derivation of the mean, variance, and standard deviation of a Bernoulli random variable.

Let X be a Bernoulli random variable with the probability of a success as p. Then

$$E[X] = \mu = \sum_{i=1}^{n} x_i \cdot P(X = x_i) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = 0 \cdot (1-p) + 1 \cdot p = p$$

Similarly, the variance of X can be computed:

$$V(X) = \sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \cdot P(X = x_i) = (0-p)^2 \cdot P(X = 0) + (1-p)^2 \cdot P(X = 1) = p^2(1-p) + (1-p$$

The standard deviation is

$$\begin{aligned} \sigma &= \sqrt{\sigma^2} \\ &= \sqrt{p(1-p)} \end{aligned}$$

4.3 Sampling Distribution of the Sum and MGF Derivation

Consider determining the sampling distribution of the sample total:

$$T_n = X_1 + X_2 + \cdots + X_n$$

Suppose $X_i \overset{iid}{\sim}$ Bernoulli(p). Then the moment-generating function of T_n is:

$$\begin{aligned} M_{T_n}(t) &= \mathbb{E}[e^{tT_n}] \\ &= \mathbb{E}\left[e^{t(X_1 + X_2 + \cdots + X_n)}\right] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[e^{tX_1} e^{tX_2} \dots e^{tX_n}\right] \quad \text{(independence)} \\
&= \mathbb{E}[e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}] \cdots \mathbb{E}[e^{tX_n}] \\
&= M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t) \\
&= \left[pe^t + (1-p)\right]^n
\end{aligned}
$$

Since this is the MGF of a binomial random variable with parameters n and p, we conclude:

$$T_n \sim \text{Binomial}(n, p)$$

Example: Binomial Distribution from Die Rolls

We can think of rolling a die n times as an example of the binomial setting. Each roll gives either a six (a "success") or a number different from six (a "failure").
Knowing the outcome of one roll doesn't tell us anything about the others, so the n rolls are independent.
If we call a six a success, then:

- The probability of success on each trial is $p = P(\text{rolling a six}) = \frac{1}{6}$
- The probability of failure is $1 - p = \frac{5}{6}$

Let Y be the number of sixes rolled in n trials. Then $Y \sim \text{Binomial}(n, p)$, and the distribution of Y is called a **binomial distribution**.

4.4 Binomial Distribution

In section 4.2 we learnt about Bernoulli random variables in which we were interested in the outcome of just a single trial. A **binomial random variable** is a generalization of several independent Bernoulli trials. Instead of performing just a single Bernoulli trial and observing whether we have a success or not, we are now performing several Bernoulli trials and observing whether we have a certain number of successes and failures. The **binomial distribution** describes the probability of having exactly k successes in n independent Bernoulli trials with probability of a success p.

Let $X \sim Bin(n, p)$. The probability of observing x successes in these n independent trials is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

- n represents the number of trials,
- x represents the number of successes,

- p represents the probability of success on any given trial,

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{is the binomial coefficient.}$$

Definition 4.3 (Mean and Variance of a Binomial Random Variable)**.**
Let $X \sim Bin(n, p)$. The mean of X is

$$E(X) = \mu = np$$

and the variance of X is

$$Var(X) = \sigma^2 = np(1-p)$$

4.4.1 Visualizing the PMF of Binomial Distributions

R code:

```
## Pmf of Binomial with n=10 and p=1/6.

x <- seq(0, 10, by=1)
y <- dbinom(x, 10, 1/6)

plot(x, y, type="p", col="blue", pch=19)
```

Probability Mass Functions (PMFs) for increasing n:

The following plots display the probability mass functions (PMFs) for a binomial distribution with $p = \frac{1}{6}$ and increasing values of n. As n increases, the binomial distribution begins to resemble a normal distribution.

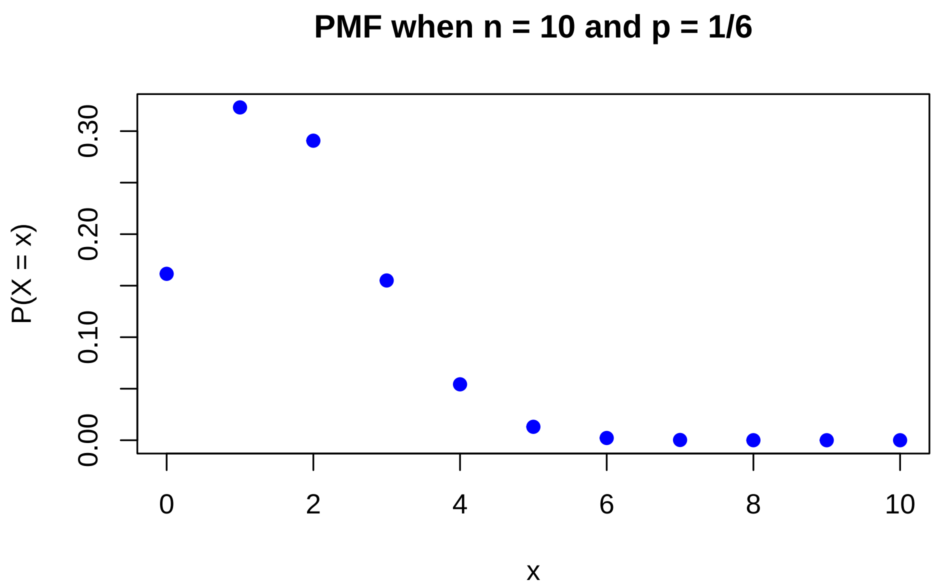


Figure 4.1: PMF of Binomial distribution with $n = 10$ and $p = \frac{1}{6}$.

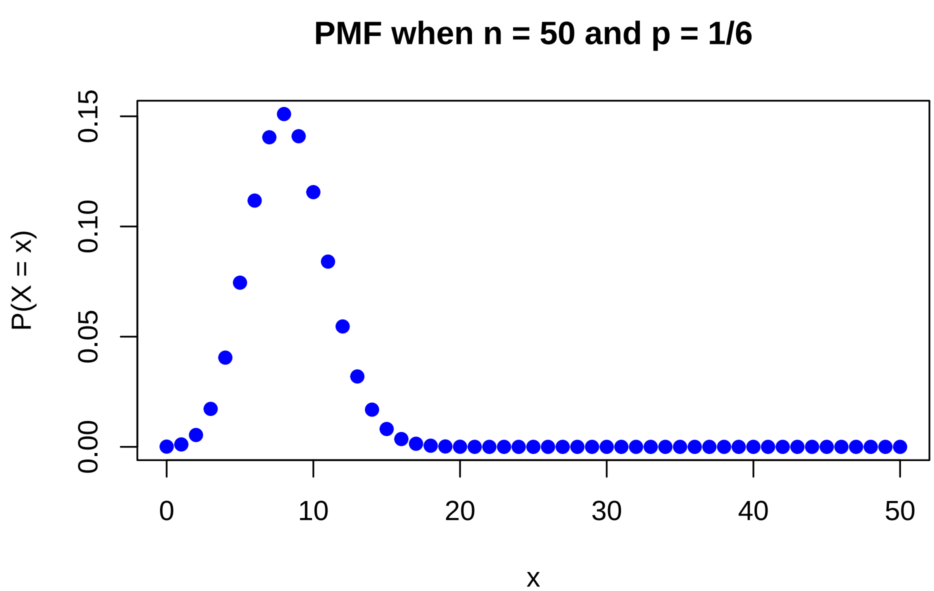


Figure 4.2: PMF of Binomial distribution with $n = 50$ and $p = \frac{1}{6}$.

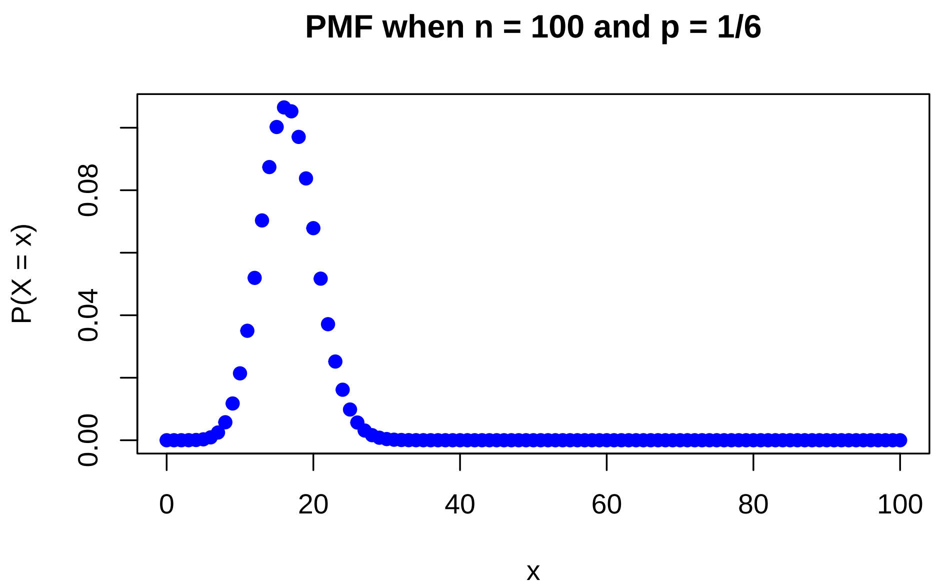


Figure 4.3: PMF of Binomial distribution with $n = 100$ and $p = \frac{1}{6}$.

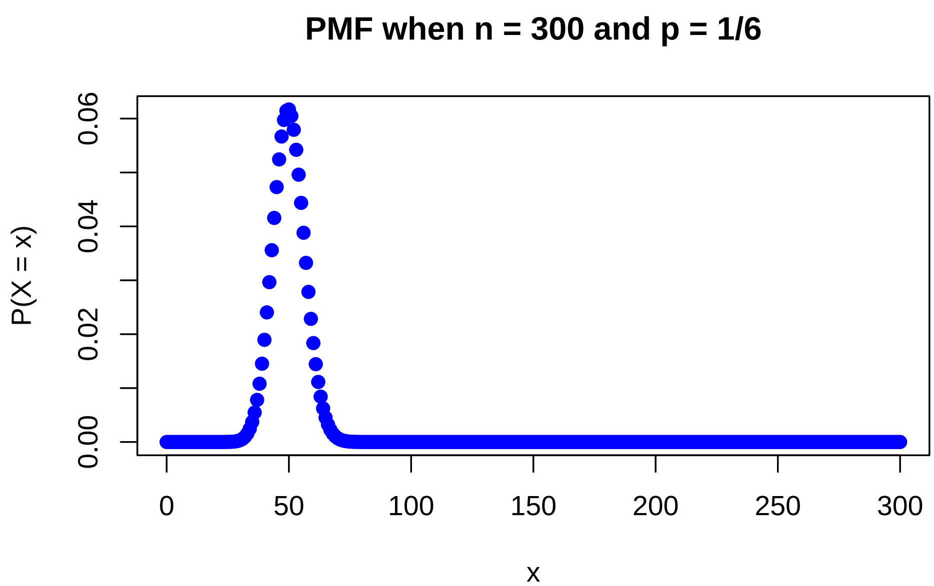


Figure 4.4: PMF of Binomial distribution with $n = 300$ and $p = \frac{1}{6}$.

4.5 Sampling Distribution of a Sample Proportion and the Normal Approximation

When studying categorical data, we are often interested not just in individual outcomes, but in the proportion of successes observed in a sample. Understanding how this proportion behaves across repeated samples is crucial for making inferences about a population. In this section, we explore the sampling distribution of a sample proportion and how it can be approximated by a normal distribution under certain conditions.

Draw a *Simple Random Sample (SRS)* of size n from a large population that contains proportion p of "successes". Let $\hat{p}$ be the ***sample proportion*** of successes:

$$\hat{p} = \frac{\text{number of successes in the sample}}{n}$$

Then:

- The **mean** of the sampling distribution of $\hat{p}$ is p.
- The **standard deviation** of the sampling distribution is $\sqrt{\frac{p(1-p)}{n}}$.

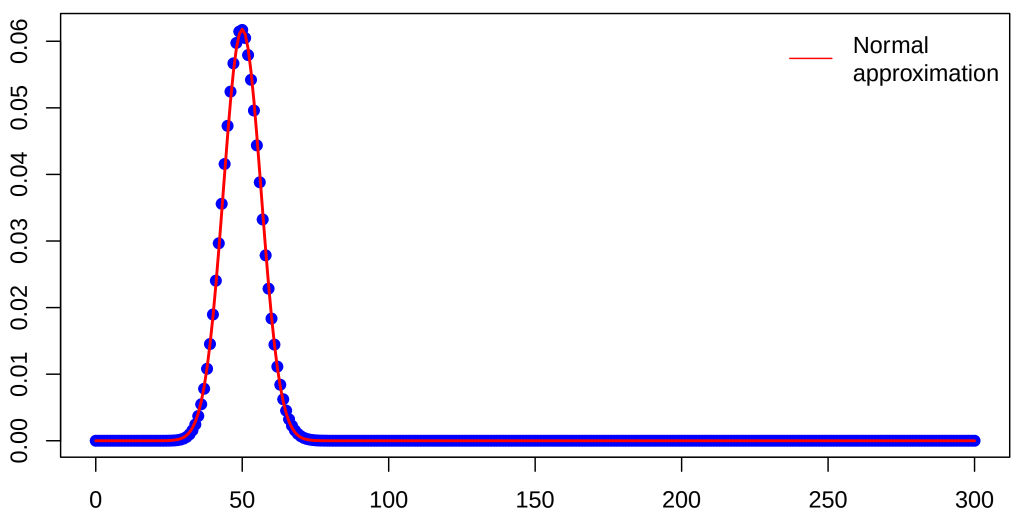


Figure 4.5: Binomial distribution with $n = 300$, $p = \frac{1}{6}$ and its Normal approximation.

According to the Central Limit Theorem (CLT), the sampling distribution of a sample proportion becomes approximately normal as the sample size increases.
That is:

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

This approximation is most accurate when both $np \geq 10$ and $n(1-p) \geq 10$.
These are called the **success-failure conditions**.

Key Point: When the success-failure conditions are met, the normal approximation to the sampling distribution of $\hat{p}$ can be used for probability calculations.

Conditions for Using the Normal Approximation

Suppose $X \sim \text{Binomial}(n, p)$. Then:

$$\mu = np, \quad \sigma^2 = np(1-p)$$

Binomial probabilities can be approximated by the normal distribution:

$$X \approx \mathcal{N}(np,\, np(1-p))$$

This approximation is *useful for large n* and valid under the following conditions:

Standard Conditions

The binomial setting holds (i.e., independent trials, fixed n, same probability p) and

$$np \geq 10 \quad \text{and} \quad np(1-p) \geq 10$$

Alternatively, a more conservative criterion for using the normal approximation is:

$$n > 9 \cdot \left(\frac{\max(p,\, 1-p)}{\min(p,\, 1-p)} \right)$$

These ensure that the binomial distribution is sufficiently symmetric and smooth to approximate with the normal distribution.

We derive the sampling distribution of $\hat{p}$ using properties of the Bernoulli distribution.

Bernoulli Distribution (Binomial with $n = 1$)

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th roll is a six} \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = \mathbb{E}(X_i) = p, \quad \sigma^2 = \text{Var}(X_i) = p(1-p)$$

Let $\hat{p}$ be our estimate of p. Note that $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$. Let $\hat{p} = \frac{\#\text{ successes } (X)}{\text{sample size } (n)}$
Recall that for $X \sim \text{Binomial}(n, p)$:

$$X \mathrel{\dot\sim} \mathcal{N}(np, np(1-p))$$

Let $\hat{p} = \frac{X}{n}$

Mean of $\hat{p}$:

$$\mathbb{E}(\hat{p}) = \mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot \mathbb{E}(X) = \frac{1}{n} \cdot np = p$$

Variance of $\hat{p}$:

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot \mathrm{Var}(X) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n}$$

By the Central Limit Theorem (CLT), for sufficiently large n:

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

Standardization of $\hat{p}$:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

If n is large, then by the Central Limit Theorem:

$$\bar{X} \approx \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \Rightarrow \quad \hat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Example 4.1.

[Normal Approximation for Proportions] In the last election, a state representative received 52% of the votes cast. One year after the election, the representative organized a survey that asked a random sample of 300 people whether they would vote for him in the next election. If we assume that his popularity has not changed, what is the probability that more than half the sample would vote for him?

Solution 1 (using Normal Approximation)

We want to determine the probability that the sample proportion is greater than 50%. In other words, we want to find $P(\hat{p} > 0.50)$.
We know that the sample proportion $\hat{p}$ is roughly Normally distributed with mean $p = 0.52$ and standard deviation

$$\sqrt{p(1-p)/n} = \sqrt{(0.52)(0.48)/300} = 0.0288. Thus, we calculate P(\hat{p} > 0.50) = P\left(\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} > \frac{0.50 -}{0.028}\right.$$

If we assume that the level of support remains at 52%, the probability that more than half the sample of 300 people would vote for the representative is 0.7549.

R code (Normal Approximation):

```
1 - pnorm(0.50, mean = 0.52, sd = 0.0288)
## [1] 0.7562982
```

Recall that, `pnorm` will give you the area to the left of 0.50, for a Normal distribution with mean 0.52 and standard deviation 0.0288.

Solution 2 (using Binomial)

We want to determine the probability that the sample proportion is greater than 50%. In other words, we want to find $P(\hat{p} > 0.50)$. We know that $n = 300$ and $p = 0.52$.
Thus, we calculate

$$\begin{aligned}
P(\hat{p} > 0.50) &= P\left(\frac{\sum_{i=1}^{n} x_i}{n} > 0.50\right) \\
&= P\left(\sum_{i=1}^{300} x_i > 150\right) \\
&= 1 - P\left(\sum_{i=1}^{300} x_i \leq 150\right) \\
&\text{(it can be shown that } Y = \sum_{i=1}^{300} x_i \text{ has a Binomial distribution with} \\
&n = 300 \text{ and } p = 0.52) \\
&= 1 - F_Y(150)
\end{aligned}$$

R code (using Binomial distribution):

```
1- pbinom(150, size = 300, prob = 0.52);
## [1] 0.7375949
```

Recall that, `pbinom` will give you the CDF at 150, for a Binomial distribution with $n = 300$ and $p = 0.52$.

Solution 3 (using continuity correction)

We have that $n = 300$ and $p = 0.52$. Thus, we calculate

$$\begin{aligned}
P(\hat{p} > 0.50) &= P\left(\frac{\sum_{i=1}^{n} x_i}{n} > 0.50\right) \\
&= P\left(\sum_{i=1}^{300} x_i > 150\right) \\
&= 1 - P\left(\sum_{i=1}^{300} x_i \leq 150\right) \\
&\text{(it can be shown that } Y = \sum_{i=1}^{300} x_i \text{ has a Binomial distribution with} \\
&n = 300 \text{ and } p = 0.52). \\
&\approx 1 - P\left(\sum_{i=1}^{300} x_i \leq 150.5\right) \quad \text{(continuity correction)}
\end{aligned}$$

$$
\begin{aligned}
&= 1 - P\left(\frac{\sum_{i=1}^{300} x_i}{n} \le \frac{150.5}{300}\right) \\
&= 1 - P(\hat{p} \le 0.5017) \\
&= 1 - P\left(Z \le -0.6354\right) \quad \text{(Why?)}
\end{aligned}
$$

R code (Normal approximation with continuity correction):

```
1 - pnorm(0.5017, mean = 0.52, sd = 0.0288)
## [1] 0.7374216
```

Recall that, `pnorm` will give you the area to the left of 0.5017, for a Normal distribution with mean 0.52 and standard deviation 0.0288.

4.6 Normal Approximation to Binomial

Let $X = \sum_{i=1}^{n} Y_i$ where $Y_1, Y_2, \ldots, Y_n$ are iid Bernoulli random variables. Note that $X = n\hat{p}$.

1. $n\hat{p}$ is approximately Normally distributed provided that $np \ge 10$ and $n(1-p) \ge 10$.
2. Another criterion is that the Normal approximation is adequate if

$$
n > 9\left(\frac{\text{larger of } p \text{ and } q}{\text{smaller of } p \text{ and } q}\right)
$$

3. The expected value: $E(\hat{p}) = np$.
4. The variance: $V(\hat{p}) = np(1-p) = npq$.

4.7 Continuity Correction

The normal distribution is continuous, while the binomial distribution is discrete. When we approximate a binomial probability using the normal distribution, this mismatch can lead to inaccuracy—especially near the boundaries of discrete values. A continuity correction improves the approximation by adjusting for this difference. In this section, we explore how and why this correction is applied.

Continuity Correction Table

Binomial Probability	Continuity Correction	Normal Approximation
$P(X = x)$	$P(x - 0.5 \leq X \leq x + 0.5)$	$P\left(\frac{x - 0.5 - \mu}{\sigma} \leq Z \leq \frac{x + 0.5 - \mu}{\sigma}\right)$
$P(X \leq x)$	$P(X \leq x + 0.5)$	$P\left(Z \leq \frac{x + 0.5 - \mu}{\sigma}\right)$
$P(X < x)$	$P(X \leq x - 0.5)$	$P\left(Z \leq \frac{x - 0.5 - \mu}{\sigma}\right)$
$P(X \geq x)$	$P(X \geq x - 0.5)$	$P\left(Z \geq \frac{x - 0.5 - \mu}{\sigma}\right)$
$P(X > x)$	$P(X \geq x + 0.5)$	$P\left(Z \geq \frac{x + 0.5 - \mu}{\sigma}\right)$
$P(a \leq X \leq b)$	$P(a - 0.5 \leq X \leq b + 0.5)$	$P\left(\frac{a - 0.5 - \mu}{\sigma} \leq Z \leq \frac{b + 0.5 - \mu}{\sigma}\right)$

Suppose that Y has a Binomial distribution with $n = 20$ and $p = 0.4$. We will find the exact probabilities that $Y \leq y$ and compare these to the corresponding values found by using two Normal approximations. One of them, when X is Normally distributed with $\mu_X = np$ and $\sigma_X = \sqrt{np(1-p)}$. The other one, W, a shifted version of X.

For example,

$$P(Y \leq 8) = 0.5955987$$

As previously stated, we can think of Y as having approximately the same distribution as X.

$$P(Y \leq 8) \approx P(X \leq 8) = P\left[\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{8 - 8}{\sqrt{20(0.4)(0.6)}}\right] = P(Z \leq 0) = 0.5$$

$$P(Y \leq 8) \approx P(W \leq 8.5) = P\left[\frac{W - np}{\sqrt{np(1-p)}} \leq \frac{8.5 - 8}{\sqrt{20(0.4)(0.6)}}\right] = P(Z \leq 0.2282) = 0.5902615$$

Example 4.2.

Fifty-one percent of adults in the U. S. whose New Year's resolution was to exercise more achieved their resolution. You randomly select 65 adults in the U. S. whose resolution was to exercise more and ask each if he or she achieved that resolution. What is the probability that exactly forty of them respond yes?

We are given that $p = 0.51$, $n = 65$, and we want to find $P(X = 40)$ where $X \sim Binomial(n = 65, p = 0.51)$.

Use Normal Approximation We use normal approximation to the binomial. First, compute the mean and standard deviation:

$$\begin{aligned} \mu &= np = 65 \times 0.51 = 33.15 \\ \sigma^2 &= np(1-p) = 65 \times 0.51 \times 0.49 = 16.485 \\ \sigma &= \sqrt{16.485} \approx 4.06 \end{aligned}$$

We apply continuity correction:

$$P(X = 40) = P(39.5 \leq X \leq 40.5)$$

$$= P\left(\frac{39.5 - 33.15}{4.06} \leq Z \leq \frac{40.5 - 33.15}{4.06}\right) = P(1.56 \leq Z \leq 1.81)$$

From the standard normal table:

$$= P(Z \leq 1.81) - P(Z \leq 1.56) = 0.0594 - 0.0352 = 0.0242$$

So the approximate probability is:

$$P(X = 40) \approx 0.0242$$

Chapter 5

Law of Large Numbers

5.1 Convergence in Probability

Definition 5.1 (Convergence in Probability).
The sequence of random variables $X_1, X_2, X_3, \ldots, X_n, \ldots$ is said to ***converge in probability*** *to the constant c, if for every $\epsilon > 0$,*

$$\lim_{n \to \infty} P\left(|X_n - c| \leq \epsilon\right) = 1$$

or equivalently,

$$\lim_{n \to \infty} P\left(|X_n - c| > \epsilon\right) = 0$$

Notation: $X_n \xrightarrow{P} c$

This concept plays a key role in the Law of Large Numbers, where the sample mean of independent and identically distributed random variables converges in probability to the population mean as the sample size grows.

Definition 5.2 (Chebyshev's Inequality).
Let X be a random variable with finite mean μ and variance σ^2. Then, for any $k > 0$,

$$P\left(|X - \mu| \geq k\right) \leq \frac{\sigma^2}{k^2}$$

Using complements:

$$P\left(|X - \mu| < k\right) \geq 1 - \frac{\sigma^2}{k^2}$$

5.2 Weak Law of Large Numbers (WLLN)

Definition 5.3 (Weak Law of Large Numbers (WLLN)).
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E(X_i) = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2$. Then, for any $\epsilon > 0$,

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \quad \textit{as } n \to \infty$$

Notation: $\bar{X}_n \xrightarrow{P} \mu$

Proof of the Weak Law of Large Numbers (WLLN)

We aim to show that for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) = 0$$

where $\bar{X}_n$ is the sample mean of n independent and identically distributed (i.i.d.) random variables with

$$E(X_i) = \mu, \quad \text{and} \quad \mathrm{Var}(X_i) = \sigma^2.$$

Let

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

By the Central Limit Theorem (CLT), we know that

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Now, applying **Chebyshev's Inequality**, which states that for any random variable X with mean μ and variance σ^2,

$$P\left(|X - \mu| > k\right) \leq \frac{\sigma^2}{k^2} \quad \text{for } k > 0,$$

to $\bar{X}_n$, we set $k = \epsilon$, and obtain:

$$P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Taking the limit as $n \to \infty$, we have:

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

Since probabilities are always non-negative, we conclude:

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) = 0.$$

By the definition of convergence in probability,

$$\bar{X}_n \xrightarrow{P} \mu.$$

$\square$

Example 5.1.

[Poisson Convergence via WLLN]
Let X_i, for $i = 1, 2, 3, \ldots$, be independent Poisson random variables with rate parameter $\lambda = 3$. Prove that:

$$\bar{X}_n \xrightarrow{P} 3$$

Properties of Poisson Distribution:

$$E(X_i) = \lambda, \quad \mathrm{Var}(X_i) = \lambda$$

In this case, $\lambda = 3$, so:

$$E(X_i) = \mathrm{Var}(X_i) = 3$$

Proof:
We know:

$$E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = 3, \quad \text{and} \quad \mathrm{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{3}{n}$$

Applying Chebyshev's Inequality:

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - 3\right| \geq \epsilon\right) \leq \frac{3}{n\epsilon^2}$$

Taking the limit as $n \to \infty$:

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - 3\right| \geq \epsilon\right) \to 0$$

Conclusion:

$$\bar{X}_n \xrightarrow{P} 3$$

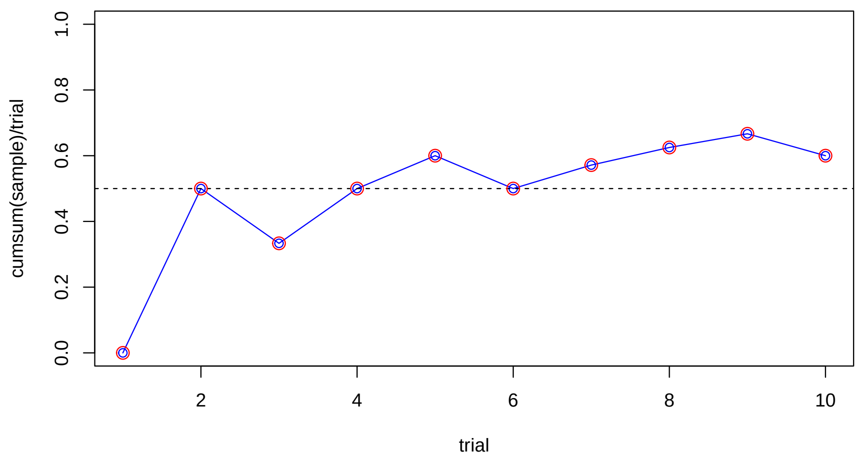


Figure 5.1: Simulation of running sample mean of Bernoulli($p = 0.5$) trials over time.

R Simulation Code (Single Sample Path):

```
n = 10
trial = seq(1, n, by = 1)
sample = rbinom(n, 1, 1/2)

plot(trial, cumsum(sample)/trial, type = "l", ylim = c(0,1), col = "blue")
points(trial, cumsum(sample)/trial, col = "red")
abline(h = 0.5, lty = 2, col = "black")
```

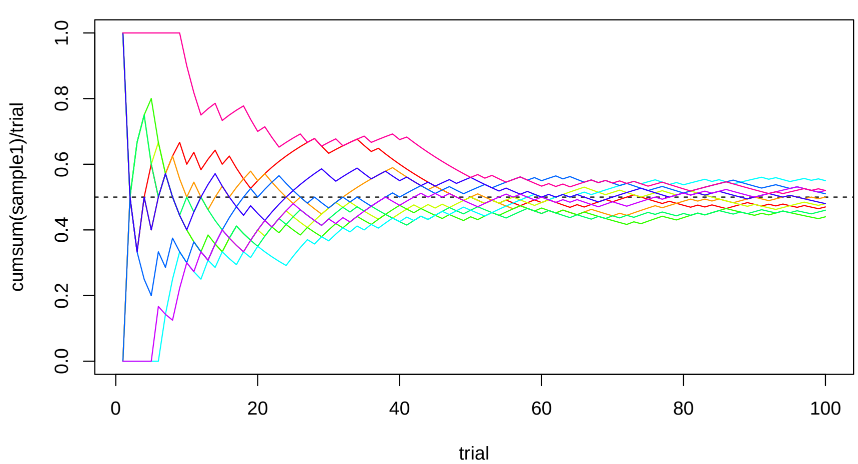


Figure 5.2: Simulation of 10 running sample means of Bernoulli($p = 0.5$) trials converging over 100 trials.

R Simulation Code (Multiple Sample Paths):

```
n = 100
trial = seq(1, 100, by = 1)

sample1 = rbinom(n, 1, 1/2)
sample2 = rbinom(n, 1, 1/2)
sample3 = rbinom(n, 1, 1/2)
sample4 = rbinom(n, 1, 1/2)
sample5 = rbinom(n, 1, 1/2)
sample6 = rbinom(n, 1, 1/2)
sample7 = rbinom(n, 1, 1/2)
sample8 = rbinom(n, 1, 1/2)

colors = rainbow(8)


plot(trial, cumsum(sample1)/trial, type = "l", col = colors[1], ylim = c(0,1))
lines(trial, cumsum(sample2)/trial, col = colors[2])
lines(trial, cumsum(sample3)/trial, col = colors[3])
lines(trial, cumsum(sample4)/trial, col = colors[4])
lines(trial, cumsum(sample5)/trial, col = colors[5])
lines(trial, cumsum(sample6)/trial, col = colors[6])
lines(trial, cumsum(sample7)/trial, col = colors[7])
lines(trial, cumsum(sample8)/trial, col = colors[8])
abline(h = 0.5, lty = 2, col = "black")
```

Empirical Probability Insight

The Law of Large Numbers gives us empirical probabilities. Consider tossing a fair coin. Define the random variable X as:

$$X = \begin{cases} 1 & \text{heads up} \\ 0 & \text{tails up} \end{cases}$$

Then as we sample more and more values of X, the sample mean $\bar{X}_n$ converges in probability to $P(\text{heads up})$, that is:

$$\bar{X}_n \xrightarrow{P} P(\text{heads up})$$

Chapter 6

One Sample Confidence Intervals on a Mean When the Population Variance is Known

6.1 Introduction

Statistical inference is concerned primarily with understanding the quality of parameter estimates. For example, a classic inferential question is, "How sure are we that the estimated mean, $\bar{x}$, is near the true population mean, μ?" While the equations and details change depending on the setting, the foundations for inference are the same throughout all of statistics. We introduce these common themes by discussing inference about the population mean, μ, and set the stage for other parameters and scenarios. Some advanced considerations are discussed. Understanding this chapter will make the rest of this book, and indeed the rest of statistics, seem much more familiar.

Definition 6.1 (Key Terms).

Population: *A group of interest (typically large).*
Sample: *A subset of a population.*
Parameter (of population): *A numerical characteristic of a population. These are usually unknown in real-life settings.*

μ: *population mean*
σ^2: *population variance*
σ: *population standard deviation*

Note: Different from a parameter of a distribution.
Statistic (of sample): *A numerical characteristic of a sample, which is calculated and known (i.e., a function of the data).*

$\bar{x}$: *sample mean*
s^2: *sample variance*
s: *sample standard deviation*

Statistical Inference: *Use statistics (known) to make conclusions on parameters (un-*

known) and quantify the degree of certainty of statements made.

The sample mean, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, is a number we use to estimate the population mean, μ. This is called a **point estimate**.

But, we know it's not equal to μ. Then, we'd rather estimate the population mean using an **interval estimate** that gives a *range of real numbers* that we hope contains the population mean, μ.

Example 6.1.

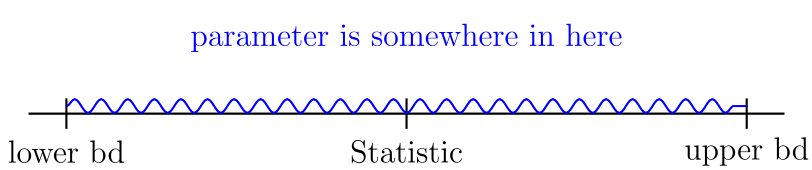
- $\bar{x}$ is a point estimate of μ
- s^2 is a point estimate of σ^2
- s is a point estimate of σ

(All calculated with data from a sample)

Due to the nature of randomness and calculating based on a subset, statistics are not guaranteed to be exactly equal to parameters.

Therefore, we create <u>intervals</u> around statistics which we believe capture the parameter.

Definition 6.2 (Confidence Interval)**.**
A confidence interval is a plausible range of values that captures a parameter with a quantified degree of confidence.



Suppose we are interested in the average mark for `STA258` for the current semester. We are 100% confident that the average mark is between 0 and 100; however, this is not useful information as we already know that the average mark must lie between 0 and 100. Using the marks of previous years, we can construct a 95% interval for the average mark. If it is determined that the average mark lies within 70% and 80%, this is much more meaningful as we can state with a high degree of certainty that the average mark is going to lie within a substantially narrow range.

In this course, all confidence intervals have the same basic skeleton:

$$\textit{estimator} \pm \underbrace{(\textit{value from reference distribution}) \times (\textit{standard error of estimate})}_{\textit{margin of error}}$$

The value from the reference distribution in the skeleton above will be either a value from the standard normal distribution or the Student t-distribution. The margin of error (MOE) can be considered as the distance around our estimator in which the true value of the parameter of interest will be found, with a specified level of confidence.

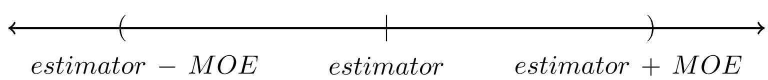


Figure 6.1: Visualization of a confidence interval on the real number line. The margin of error is abbreviated as MOE. The estimator is the centre of the interval. The confidence interval consists of all values between the estimator$-MOE$ and the estimator$+MOE$.

6.2 Interpretation

We use very specific language when we interpret a confidence interval.

Suppose we construct a C% confidence interval for some parameter such that C is between 0 and 100. In repeated sampling, we are C% confident that approximately C% of the intervals will capture the true value of the parameter.

By this we mean that if we constructed several C% confidence intervals using different samples (with or without replacing the units), then we should expect approximately C% of these intervals to capture the parameter of interest. For example suppose we construct 1000 95% confidence intervals for the population mean μ. We would expect approximately 95% of these 1000 intervals (i.e. $95\% \times 1000 = 950$) to actually capture μ.

Note 6.1.

A more intuitive but equivalent interpretation is to state that we are C% confident that our target parameter is inside the interval constructed.

It is incorrect to state that there is a $C\%$ probability that the interval we constructed contains the parameter of interest. We assume that the value of a parameter is fixed. Therefore when we construct a confidence interval, the interval either contains the parameter or it does not.

6.3 Confidence Interval for μ (Known Variance)

When we know the population standard deviation σ, we can construct a confidence interval for μ in the following manner.

Confidence Interval 6.1 (Confidence Interval on μ when σ is Known)

A $(100-\alpha)\%$ *confidence interval on* μ *when* σ *is known is*

$$\bar{x} \,\pm\, z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

The $z_{\alpha/2}$ value is obtained from standard normal tables. The standard error is $\frac{\sigma}{\sqrt{n}}$ and the margin of error is $z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$.

Let $X_1, X_2, ..., X_n$ be iid $N(\mu, \sigma^2)$, where μ is unknown and σ is known.
We know that:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

and

$$P(-1.96 < Z < 1.96) = 0.95$$

Therefore:

$$P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95 \Rightarrow P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Interpretation of Confidence Interval:

- This is a random interval $\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$
- The interval is random since $\bar{X}$ is random due to sampling.
- The population mean μ is a fixed, but unknown, number.
- The probability μ is inside the random interval is 0.95 (success rate of the method).
- 95% of all samples give an interval that captures μ, and 5% do not.

Once we observe our sample:

- This is **not** a random interval $\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$
- The probability μ is inside this interval is either 1 or 0

Confidence Interval Isn't Always Right:
Not all CIs contain the true value of the parameter. This can be illustrated by plotting many intervals simultaneously and observing.
—

R Output:

```
## Step 1. Generate random samples;
set.seed(2017)
m = 50;          # m = number of samples;
n = 25;          # n = number of obs in sample;
mu.i = 0;        # mu.i = mean of obs;
sigma.i = 5;     # sigma.i = std. dev. of obs;

mu.total = n * mu.i;            # mean of Total;
sigma.total = sqrt(n) * sigma.i;  # std. dev. of Total;
```

R Output:

```
## Step 2. Construct CIs;
xbar = rnorm(m, mu.total, sigma.total) / n;
SE = sigma.i / sqrt(n);

alpha = 0.10;
z.star = qnorm(1 - alpha / 2);
```

R Output:

```
## Step 3. Graph CIs;
matplot(rbind(xbar - z.star * SE, xbar + z.star * SE),
        rbind(1:m, 1:m),
        type = "l", lty = 1,
        xlab = "", ylab = "");
abline(v = 0, lty = 2);
```

Confidence Interval for the Mean of a Normal Population

Draw an SRS (Simple Random Sample) of size n from a Normal population having unknown mean μ and **known** standard deviation σ. A level C confidence interval for μ is:

$$\bar{x} \pm z_{*} \cdot \frac{\sigma}{\sqrt{n}}$$

The critical value z_{*} is illustrated in a Figure below and depends on C.

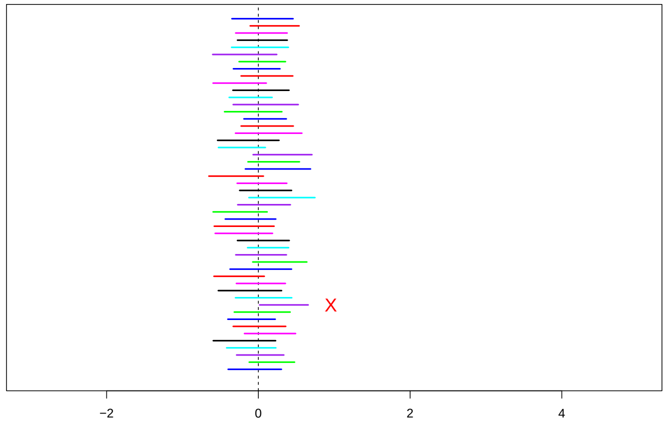


Figure 6.2: Simulated 95% confidence intervals for the population mean. Red "X" marks indicate intervals that do not contain the true mean ($\mu = 0$).

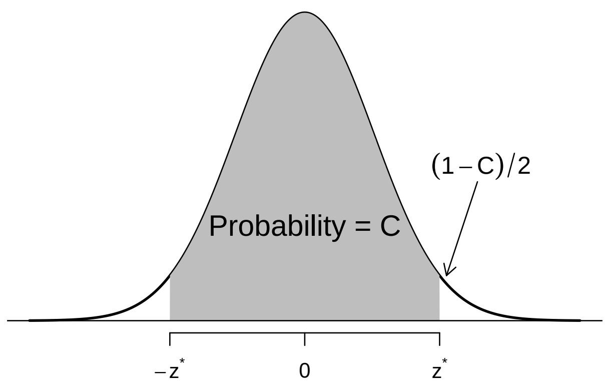


Figure 6.3: The central area under the standard normal curve with confidence level C.

Large Sample CI for μ (Normal data)

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Valid if:

- n large
- random sample from a Normal distribution
- independent observations

Some definitions:

- $1 - \alpha$ is the confidence coefficient
- $100(1 - \alpha)\%$ is the confidence level

One Sample CI on the Population Mean μ

- When population standard deviation σ is **known**
- Formula: $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
- Margin of error comes from standard normal and standard error

How to find $z_{\alpha/2}$?
Example: Find $z_{\alpha/2}$ for a 95% CI on μ:

$$1 - \alpha = 0.95, \quad \alpha = 0.05, \quad \alpha/2 = 0.025$$

$$z_{\alpha/2} = 1.96 \quad \text{(from table or R: } \texttt{qnorm(0.975)}\text{)}$$

Table of Common z-values

Confidence coefficient	Confidence level	z
0.90	90%	1.645
0.95	95%	1.96
0.99	99%	2.576

Example 6.2.
Playbill magazine reported that the mean annual household income of its readers is $119,155. Assume this estimate is based on a sample of 80 households, and that the population standard deviation is known to be $\sigma = 30{,}000$.

- $\bar{x} = 119{,}155$
- $n = 80$
- $\sigma = 30{,}000$

Tasks:

(a) Develop a 90% confidence interval estimate of the population mean.

(b) Develop a 95% confidence interval estimate of the population mean.

(c) Develop a 99% confidence interval estimate of the population mean.

90% CI Calculation

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 119{,}155 \pm 1.645 \cdot \frac{30{,}000}{\sqrt{80}}$$

$$= 119{,}155 \pm 5{,}500.73$$

$$= (113{,}654.27,\ 124{,}655.73)$$

95% CI Calculation

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 119{,}155 \pm 1.96 \cdot \frac{30{,}000}{\sqrt{80}}$$

$$= 119{,}155 \pm 6{,}574.04$$

$$= (112{,}580.96,\ 125{,}729.04)$$

99% CI Calculation

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 119{,}155 \pm 2.576 \cdot \frac{30{,}000}{\sqrt{80}}$$

$$= 119{,}155 \pm 8{,}620.04$$

$$= (110{,}534.96,\ 127{,}775.04)$$

Interpretation

We are 99% confident the mean household income of magazine readers is between \$110,534.96 and \$127,775.04.

Example 6.3.

Scenario:

The number of cars sold annually by used car salespeople is known to be **normally distributed**, with a population standard deviation of $\sigma = 15$. A random sample of $n = 15$ salespeople was taken, and the number of cars each sold is recorded below. Construct a **95% confidence interval** for the population mean number of cars sold, and provide an interpretation.

Raw data:

79	43	58	66	101
63	79	33	58	71
60	101	74	55	88

The sample mean is:

$$\bar{x} = \frac{79 + 43 + \cdots + 55 + 88}{15} = 68.6$$

R function:

```
simple.z.test = function(x, sigma, conf.level = 0.95) {
  n = length(x);
  xbar = mean(x);
  alpha = 1 - conf.level;
  zstar = qnorm(1 - alpha/2);
  SE = sigma / sqrt(n);
  xbar + c(-zstar * SE, zstar * SE);
}
```

R output:

```
# Step 1. Entering data;
cars = c(79, 43, 58, 66, 101, 63, 79,
         33, 58, 71, 60, 101, 74, 55, 88)

# Step 2. Finding CI;
simple.z.test(cars, 15)

## [1] 61.00909 76.19091
```

Interpretation: We estimate that the mean number of cars sold annually by all used car salespeople lies between 61 and 76, approximately. This type of estimate is correct 95% of the time.

Cases Where Valid

- Large samples where population is **normal**.
- Large samples where population is **not normal** (By CLT).
- Small samples where population is **normal**.

Note: A sample is considered large if $n \geq 30$.

Example 6.4.

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius) 102.5, 101.7, 103.1, 100.9, 100.5, and 102.2 on 6 different samples of the liquid. He calculates the sample mean to be 101.82. If he knows that the distribution of boiling points is Normal, with standard deviation 1.2 degrees, what is the confidence interval for the population mean at a 95% confidence level?

A confidence interval uses sample data to estimate an unknown population parameter with an indication of how accurate the estimate is and of how confident we are that the result is correct.

The **interval** often has the form
estimate ± margin of error

The **confidence level** is the success rate of the method that produces the interval. A level C **confidence interval for the mean** μ of a Normal population with **known** standard deviation σ, based on an SRS of size n, is given by

$$\bar{x} \pm z^{\star} \frac{\sigma}{\sqrt{n}}$$

The **critical value** $z^{\star}$ is chosen so that the standard Normal curve has area C between $-z^{\star}$ and $z^{\star}$.

Other things being equal, the **margin of error** of a confidence interval gets smaller as

- the confidence level C decreases,
- the population standard deviation σ decreases, and
- the sample size n increases.

6.4 APPENDIX

Interval estimators are commonly called **confidence intervals**. The upper and lower endpoints of a confidence interval are called the **upper** and **lower confidence limits**, respectively. The probability that a (random) confidence interval will enclose θ (a fixed quantity) is called the **confidence coefficient**.

Suppose that $\hat{\theta}_L$ and $\hat{\theta}_U$ are the (random) lower and upper confidence limits, respectively, for a parameter θ. Then, if

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha,$$

the probability $(1 - \alpha)$ is the **confidence coefficient**.

Pivotal quantities

One very useful method for finding confidence intervals is called the **pivotal method**. This method depends on finding a pivotal quantity that possesses two characteristics:

- It is a function of the sample measurements and the unknown parameter θ, where θ is the **only** unknown quantity.
- Its probability distribution does not depend on the parameter θ.

Index