It Can Be Shown

Notes on Statistical Theory

Benjamin Nutter

2018-09-27

Contents

Chapter 1

Introduction

There is one phrase that makes me cringe every time I see it. It's a phrase that embodies feelings of frustration, inadequacy, and failure to understand. That phrase:

> It can be shown

Everytime I read that phrase, I would look at the subsequent result and think "Really? It can?"

This book is a collection of notes that I've put together to avoid having to feel that way in the future. It is, essentially, a collection of definitions and proofs that have helped me understand and apply mathematical and statistical theory. Most imporantly, it spells even the smallest steps along each development so that I don't have to worry about solving it again in the future.

You won't find much in the way of application. There are no exercises. There is only minimal explanation. My intent is to show development of statistical theory and nothing else.

Chapter 2

Analysis of Variance

2.1 One-Way Design

2.1.1 Decomposition of Sums of Squares

$$
\begin{aligned}
SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+}+\bar{x}_{i+}-x_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(x_{ij}-\bar{x}i+)+(\bar{x}_{i+}-\bar{x}_{++})\right]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(\bar{x}_{i+}-\bar{x}_{++})+(x_{ij}-\bar{x}i+)\right]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(\bar{x}_{i+}-\bar{x}_{++})^2+2(\bar{x}_{i+}-\bar{x}_{++})(x_{ij}-\bar{x}_{i+})+(x_{ij}-\bar{x}_{i+})^2\right] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{x}_{i+}-\bar{x}_{++})^2+\sum_{i=1}^{a}\sum_{j=1}^{n_i}2(\bar{x}_{i+}-\bar{x}_{++})(x_{ij}-\bar{x}_{i+})+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+2\sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+2\sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})\left(\sum_{j=1}^{n_i}x_{ij}-\sum_{j=1}^{n_i}\bar{x}_{i+}\right)+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+2\sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})(x_{i+}-n_i\bar{x}_{i+})+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+2\sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})\big(x_{i+}-n_i\frac{x_{i+}}{n_i})+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+2\sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})(x_{i+}-x_{i+})+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+2\sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})\cdot 0+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+0+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+}-\bar{x}_{++})^2+\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2
\end{aligned}
$$

The components are commonly referred to as

$$SS_{Factor} = \sum_{i=1}^{a} n_i(\bar{x}_{i+} - \bar{x}_{++})^2$$

and

$$SS_{Error} = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2$$

Notice that SS_{Factor} compares the factor means to the overall mean, and it can be said that SS_{Factor} measures the variation *between* factors. SS_{Error} compares each observation to the overall mean, and can be said to describe the variation *within* factors.

When $n_1 = n_2 = \cdots n_i = n$, the design is said to be balanced.

See (Montgomery 2004, 66)

2.2 Computational Formulas

SS_{Total} and SS_{Factor} can be simplified for convenient computation.

$$\begin{aligned} SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{++})^2 \\ {}^{[1]} &= \sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij}^2 - x_{++}\sum_{j=1}^{n_i}\frac{1}{n_i} \end{aligned}$$

1. See Theorem 42.3.1

$$\begin{aligned} SS_{Factor} &= \sum_{i=1}^{a} n_i(\bar{x}_{i+} - \bar{x}_{++})^2 \\ {}^{[1]} &= \sum_{i=1}^{a}\frac{\bar{x}_{i+}^2}{n_i} - \bar{x}_{++}\sum_{i=1}^{a}\frac{1}{n_i} \end{aligned}$$

1. See Theorem 42.3.1

SS_{Error} does not simplify to a convenient form, but

$$\begin{aligned} SS_{Total} &= SS_{Factor} + SS_{Error} \\ \Rightarrow SS_{Error} &= SS_{Total} - SS_{Factor} \end{aligned}$$

2.3 Randomized Complete Block Design

Blocking in ANOVA is a method of eliminate the effect of a controllable nuisance variable. To implement this design, suppose we have a treatments we want to compare, and b blocks. We may analyze the data by use of the sums of squares, similar to the one-way design.

2.3.1 Decomposition of Sums of Squares

$$\begin{aligned}
SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} + \bar{x}_{i+} - \bar{x}_{i+} + \bar{x}_{+j} - \bar{x}_{+j} + \bar{x}_{++} - \bar{x}_{++} - \bar{x}_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++}) + (\bar{x}_{+j} - \bar{x}_{++}) + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})\big]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++})^2 + 2(\bar{x}_{i+} - \bar{x}_{++})(\bar{x}_{+j} - \bar{x}_{++}) \\
&\quad + 2(\bar{x}_{i+} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++}) + (\bar{x}_{+j} - \bar{x}_{++})^2 \\
&\quad + 2(\bar{x}_{+j} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++}) \\
&\quad + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2\big] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++})^2 + (\bar{x}_{+j} - \bar{x}_{++})^2 + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2 \\
&\quad + 2(\bar{x}_{i+} - \bar{x}_{++})(\bar{x}_{+j} - \bar{x}_{++}) + 2(\bar{x}_{i+} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++}) \\
&\quad + 2(\bar{x}_{+j} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})\big] \\
{}^{[1]} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++})^2 + (\bar{x}_{+j} - \bar{x}_{++})^2 + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2 + 0 + 0 + 0\big] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2 \\
&= b\sum_{i=1}^{a}(\bar{x}_{i+} - \bar{x}_{++})^2 + a\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2
\end{aligned}$$

1. It is shown that the cross products are equal to zero in Section 38.3.3

These terms are commonly referred to as

$$\begin{aligned}
SS_{Factor} &= b\sum_{i=1}^{a}(\bar{x}_{i+} - \bar{x}_{++})^2 \\
SS_{Block} &= a\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 \\
SS_{Error} &= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2
\end{aligned}$$

See (Montgomery 2004, 126)

2.3.2 Computational Formulae

SS_{Total}, SS_{Factor}, and SS_{Block} can all be simplified for convenient computation.

$$
\begin{aligned}
SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{++})^2 \\
[1] &= \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}^2 - \frac{x_{++}}{ab}
\end{aligned}
$$

$$
\begin{aligned}
SS_{Factor} &= b\sum_{i=1}^{a}(\bar{x}_{i+} - \bar{x}_{++})^2 \\
[1] &= \frac{1}{b}\sum_{i=1}^{a} x_{i+}^2 - \frac{x_{++}^2}{ab}
\end{aligned}
$$

$$
\begin{aligned}
SS_{Block} &= a\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 \\
[1] &= \frac{1}{a}\sum_{j=1}^{b} x_{+j}^2 - \frac{x_{++}^2}{ab}
\end{aligned}
$$

1. See Theorem 42.3.1

SS_{Error} does not simplify to any convenient form, but may be calculated from the other terms as $SS_{Error} = SS_{Total} - SS_{Factor} - SS_{Block}$

2.3.3 RCBD Cross Products

The cross products of the RCBD design

$$
\begin{aligned}
&2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+} - \bar{x}_{++})(\bar{x}_{+j} - \bar{x}{++}) \\
&+2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})(x_{ij} + \bar{x}_{i+} + \bar{x}_{+j} - \bar{x}_{++}) \\
&+2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+i} - \bar{x}_{++})(x_{ij} + \bar{x}_{i+} + \bar{x}_{+j} - \bar{x}_{++}) = 0
\end{aligned}
$$

Proof:

$$
\begin{aligned}
&2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+}-\bar{x}_{++})(\bar{x}_{+j}-\bar{x}{++}) + 2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&\quad + 2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+i}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&= 2\Bigg(\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+}-\bar{x}_{++})(\bar{x}_{+j}-\bar{x}{++}) + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&\quad + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+i}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++})\Bigg) \\
&= 2\sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+}-\bar{x}_{++})(\bar{x}_{+j}-\bar{x}{++}) + (\bar{x}_{+j}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&\quad + (\bar{x}_{+i}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++})\big] \\
&= 2\sum_{i=1}^{a}\sum_{j=1}^{b}\big[\bar{x}_{i+}\bar{x}_{+j}-\bar{x}_{i+}\bar{x}_{++}-\bar{x}_{+j}\bar{x}_{++}+\bar{x}_{++}^2 \\
&\quad + x_{ij}\bar{x}_{+j}-\bar{x}_{i+}\bar{x}_{+j}-\bar{x}_{+j}^2+\bar{x}_{+j}\bar{x}_{++}-x_{ij}\bar{x}_{++}+\bar{x}_{i+}\bar{x}_{++}+\bar{x}_{+j}\bar{x}_{++}-\bar{x}_{++}^2 \\
&\quad + x_{ij}\bar{x}_{+j}-\bar{x}_{i+}^2-\bar{x}_{i+}\bar{x}_{+j}+\bar{x}_{+j}\bar{x}_{++}-x_{ij}\bar{x}_{++}+\bar{x}_{i+}\bar{x}_{++}+\bar{x}_{+j}\bar{x}_{++}-\bar{x}_{++}^2\big] \\
&= 2\sum_{i=1}^{a}\sum_{j=1}^{b}(-\bar{x}_{++}^2-\bar{x}_{i+}^2-\bar{x}_{+j}^2+x_{ij}\bar{x}_{i+}+x_{ij}\bar{x}_{+j}-2x_{ij}\bar{x}_{++}-\bar{x}_{i+}\bar{x}_{+j} \\
&\quad + 2\bar{x}_{i+}\bar{x}_{++}+2\bar{x}_{+j}\bar{x}_{++}) \\
&= 2\Bigg(-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{++}^2-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}^2-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{i+}+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{+j} \\
&\quad -\sum_{i=1}^{a}\sum_{j=1}^{b}2x_{ij}\bar{x}_{++}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \\
&= 2\Bigg(\frac{ab\bar{x}_{++}^2}{a^2b^2}-\frac{b}{b^2}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{a}{a^2}\sum_{j=1}^{b}\bar{x}_{+j}^2+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{i+}+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{+j} \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \\
{}^{[1]}&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{+j} \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \\
{}^{[2]}&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \\
{}^{[3]}&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\frac{\bar{x}_{++}}{ab}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \\
{}^{[4]}&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2
\end{aligned}
$$

$$-\frac{2\bar{x}{++}^2}{ab}-\frac{\bar{x}_{++}}{ab}+\frac{2\bar{x}_{++}^2}{ab}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \quad {}^{[5]} = 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 \quad -\frac{2\bar{x}{++}^2}{ab}-\frac{\bar{x}_{++}}{ab}+$$

1. See Summation Theorem 37.1.7
2. See Summation Theorem 37.1.8
3. See Summation Theorem 37.1.4
4. See Summation Theorem 37.1.5
5. See Summation Theorem 37.1.6

Using the theorems in Chapter @ref{summation-chapter} it is can be shown that each of the three cross products is equal to zero. However, the physical tedium of reducing each cross product is much greater than the approach taken above.

Chapter 3

Bernoulli Distribution

3.1 Probability Mass Function

A random variable is said to have a Bernoulli Distribution with parameter p if its probability mass function is:

$$\Pr(X = x) = \begin{cases} \pi^x (1-\pi)^{1-x}, & x = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

Where π is the probability of a success.

3.2 Cumulative Mass Function

$$\Pr(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - \pi & x = 0 \\ 1 & 1 \leq x \end{cases}$$

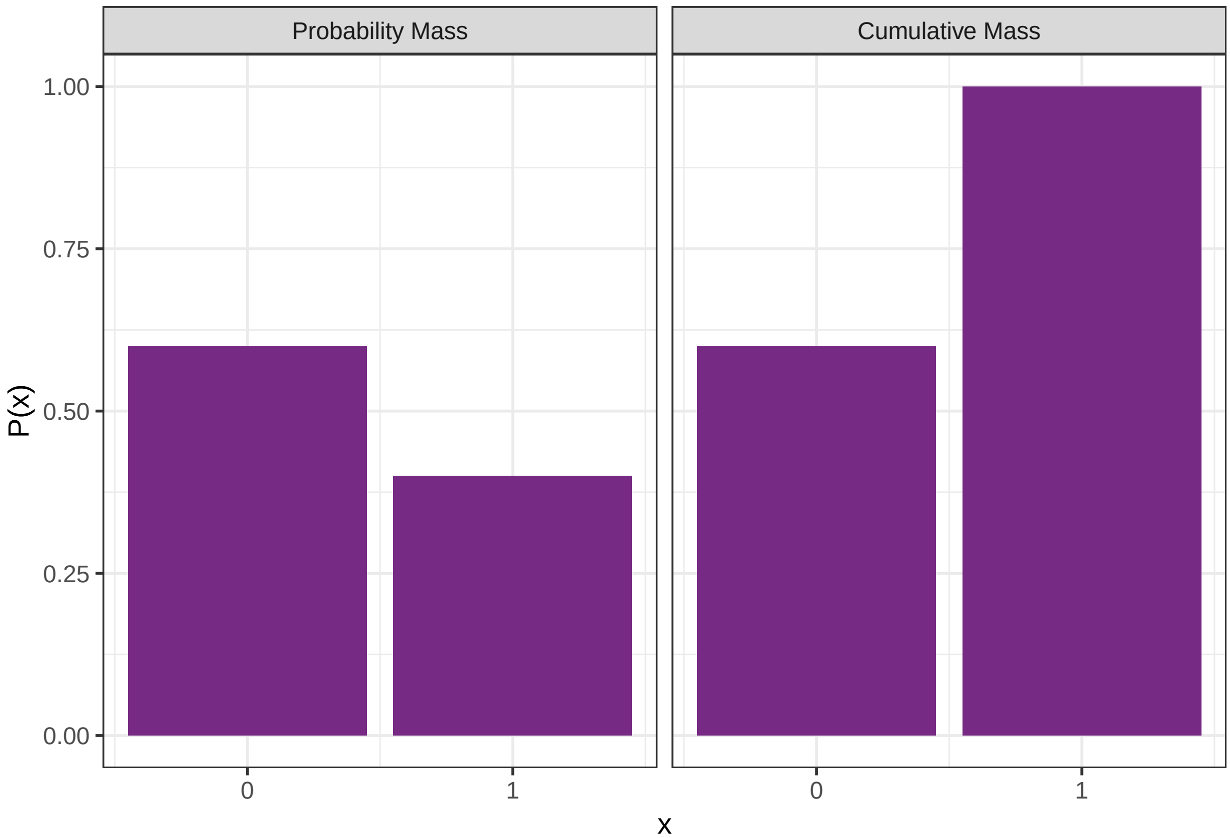

Figure 3.1: (#fig:Bernoulli_Distribution)The graphs on the left and right show a Bernoulli Probability Distribution and Cumulative Distribution Function, respectively, with $\pi = .4$. Note that this is identical to a Binomial Distribution with parameters $n = 1$ and $\pi = .4$.

3.3 Expected Values

$$
\begin{aligned}
E(X) &= \sum_{i=0}^{1} x \cdot \Pr(X = x) \\
&= \sum_{i=0}^{1} x \cdot \pi^x (1-\pi)^{1-x} \\
&= 0 \cdot \pi^0 (1-\pi)^{1-0} + 1 \cdot \pi^1 (1-\pi)^{1-1} \\
&= 0 + \pi(1-\pi)^0 \\
&= \pi
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \sum_{i=0}^{1} x^2 \cdot \Pr(X = x) \\
&= \sum_{i=0}^{1} x^2 \cdot \pi^x (1-\pi)^{1-x} \\
&= \sum_{i=0}^{1} 0^2 \cdot \pi^0 (1-\pi)^{1-0} + 1^2 \cdot \pi^1 (1-\pi)^{1-1} \\
&= 0 \cdot 1 \cdot 1 + 1 \cdot \pi \cdot 1 \\
&= 0 + \pi \\
&= \pi
\end{aligned}
$$

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

$$
\begin{aligned}
\sigma^2 &= E(X^2) - E(X)^2 \\
&= \pi - \pi^2 \\
&= \pi(1-\pi)
\end{aligned}
$$

3.4 Moment Generating Function

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{i=0}^{1} e^{tx} p(x) \\
&= \sum_{i=0}^{1} e^{tx} \pi^x (1-\pi)^{1-x} \\
&= e^{t0} \pi^0 (1-\pi)^{1-0} + e^t \pi^t (1-\pi)^{1-1} \\
&= (1-\pi) + e^t \pi \\
&= \pi e^t + (1-\pi)
\end{aligned}$$

$$M_X^{(1)}(t) = \pi e^t$$

$$M_X^{(2)}(t) = \pi e^t$$

$$\begin{aligned}
E(X) &= M_X^{(1)}(0) \\
&= \pi e^0 \\
&= \pi e^0 \\
&= \pi
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= M_X^{(2)}(0) \\
&= \pi e^0 \\
&= \pi
\end{aligned}$$

$$\begin{aligned}
\mu &= E(X) \\
&= \pi
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= E(X^2) - E(X)^2 \\
&= \pi - \pi^2 \\
&= \pi(1-\pi)
\end{aligned}$$

3.5 Theorems for the Bernoulli Distribution

3.5.1 Validity of the Distribution

$$\sum_{x=0}^{1} \pi^x (1-\pi)^{1-x} = 1$$

Proof:

$$\begin{aligned}
\sum_{x=0}^{1} \pi^x (1-\pi)^{1-x} &= \pi^0 (1-\pi)^1 + \pi^1 (1-\pi)^0 \\
&= (1-\pi) + \pi \\
&= 1
\end{aligned}$$

3.5.2 Sum of Bernoulli Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from a Bernoulli distribution with parameter p. Let $Y = \sum_{i=1}^{n} X_i$.\ Then $Y \sim \text{Binomial}(n, \pi)$

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{tX_1} e^{tX_2} \cdots e^{tX_n}) \\
&= E(e^{tX_1}) E(e^{tX_2}) \cdots E(e^{tX_n}) \\
&= (\pi e^t + (1-\pi))(\pi e^t + (1-\pi)) \cdots (\pi e^t + (1-\pi)) \\
&= (\pi e^t + (1-\pi))^n
\end{aligned}$$

Which is the moment generating function of a Binomial random variable with parameters n and π. Thus, $Y \sim \text{Binomial}(n, \pi)$.

Chapter 4

Binomial Distribution

4.1 Probability Mass Function

A random variable is said to follow a Binomial distribution with parameters n and π if its probability mass function is:

$$\Pr(X = x) = \begin{cases} \binom{n}{x}\pi^x(1-\pi)^{n-x}, & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Where n is the number of trials performed and π is the probability of a success on each individual trial.

4.2 Cumulative Mass Function

$$\Pr(X \le x) = \begin{cases} 0 & x < 0 \\ \sum\limits_{i=0}^{x} \binom{n}{i}\pi^i(1-\pi)^{n-i} & 0 \le x = 0, 1, 2, \dots, n \\ 1 & n \le x \end{cases}$$

A recursive form of the cdf can be derived and has some usefulness in computer applications. With it, one need only initiate the first value and additional cumulative probabilities can be calculated. It is derived as follows:

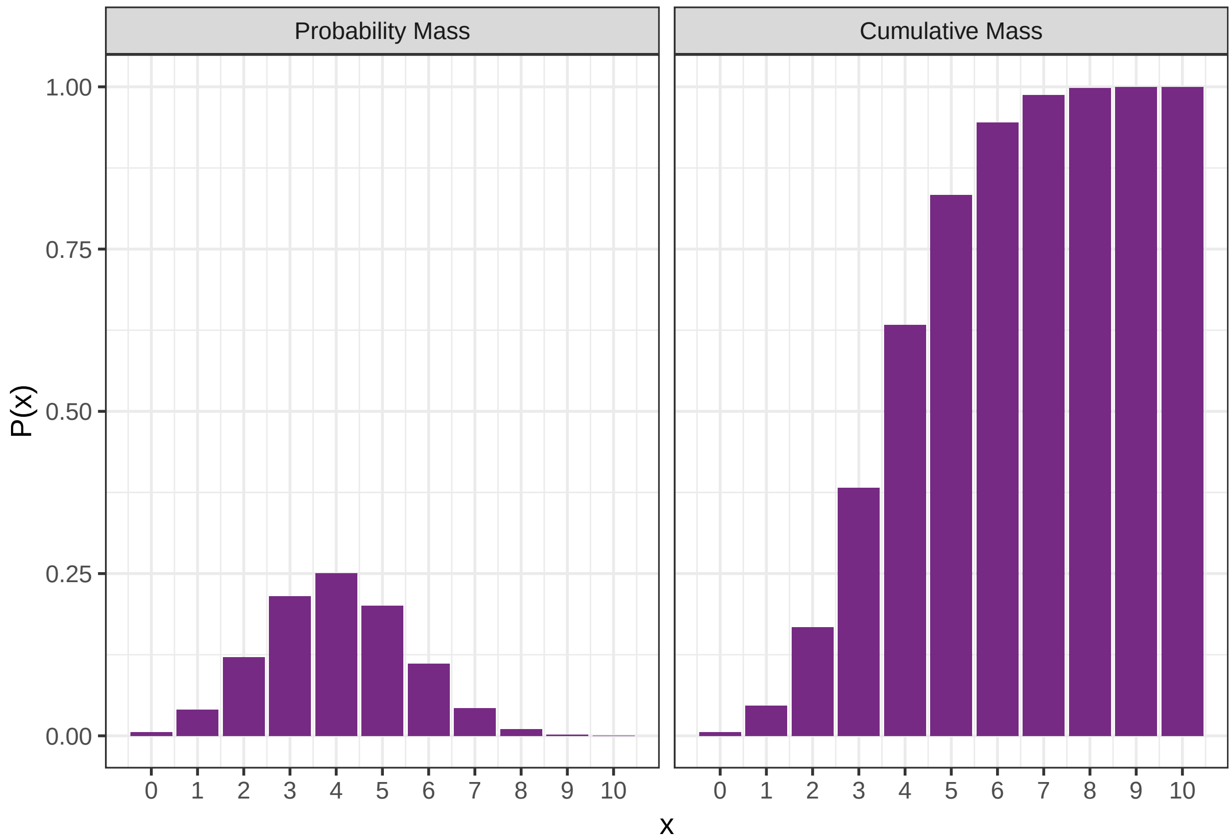

Figure 4.1: (#fig:Binomial_Distribution)The graphs on the left and right show a Binomial Probability Distribution and Cumulative Distribution Function, respectively, with $n = 10$ and $\pi = .4$.

$$
\begin{aligned}
F(x+1) &= \binom{n}{x+1}\pi^{x+1}(1-\pi)^{n-(x+1)} \\
&= \frac{n!}{(x+1)!(n-(x+1))!}\pi^{x+1}(1-\pi)^{n-(x+1)} \\
&= \frac{n!}{(x+1)!(n-x-1)!}\pi^{x+1}(1-\pi)^{n-x-1} \\
&= \frac{(n-x)n!}{(x+1)x!(n-x)(n-x-1)!}\pi\cdot\pi^{x}\frac{(1-\pi)^{n-x}}{(1-\pi)} \\
&= \frac{(n-x)n!}{(x+1)x!(n-x)!}\cdot\frac{\pi}{1-\pi}\pi^{x}(1-\pi)^{n-x} \\
&= \frac{\pi}{1-\pi}\cdot\frac{n-x}{x+1}\cdot\frac{n!}{x!(n-x)!}\pi^{x}(1-\pi)^{n-x} \\
&= \frac{\pi}{1-\pi}\cdot\frac{n-x}{x+1}\cdot\binom{n}{x}\pi^{x}(1-\pi)^{n-x} \\
&= \frac{\pi}{1-\pi}\cdot\frac{n-x}{x+1}\cdot F(x)
\end{aligned}
$$

4.3 Expected Values

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \cdot \Pr(X = x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} \pi^x (1-\pi)^{n-x} \\
{}^{[1]} &= \sum_{x=0}^{n} x \binom{n}{x} \pi^x q^{n-x} \\
&= 0 \cdot \binom{n}{0} \pi^0 q^n + 1 \cdot \binom{n}{1} \pi^1 q^{n-1} + \cdots + n \binom{n}{n} \pi^n q^{n-n} \\
&= 0 + 1 \binom{n}{1} \pi^1 q^{n-1} + 2 \binom{n}{2} \pi^2 q^{n-2} + \cdots + n \binom{n}{n} \pi^n q^{n-n} \\
&= n\pi^1 q^{n-1} + n(n-1)\pi^2 q^{n-2} + \cdots + n(n-1)\pi^{n-1} q^{n-(n-1)} + n\pi^n \\
&= n\pi[q^{n-1} + (n-1)\pi q^{n-2} + \cdots + \pi^{n-1}] \\
&= n\pi\Big[\binom{n-1}{0} \pi^0 q^{n-1} + \binom{n-1}{1} \pi^1 q^{(n-1)-1} + \cdots + \binom{n-1}{n-1} \pi^{n-1} q^{(n-1)-(n-1)}\Big] \\
&= n\pi(\sum_{x=0}^{n-1} \binom{n-1}{x} \pi^x q^{(n-1)-x}) \\
{}^{[2]} &= n\pi(\pi + q)^{n-1} \\
{}^{[1]} &= n\pi(\pi + (1-\pi))^{n-1} \\
&= n\pi(\pi + 1 - \pi)^{n-1} \\
&= n\pi(1)^{n-1} \\
&= n\pi(1) \\
&= n\pi
\end{aligned}
$$

1. Let $q = (1 - \pi)$
2. By the Binomial Theorem (6.1.2), $\sum\limits_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 \Pr(X = x) \\
&= \sum_{x=0}^{n} x^2 \binom{n}{x} \pi^x (1-\pi)^{n-x} \\
{}^{[1]} &= \sum_{x=0}^{n} x^2 \binom{n}{x} \pi^x q^{n-x} \\
&= 0^2 \frac{n!}{0!(n-0)!} \pi^0 q^n + 1^2 \frac{n!}{1!(n-1)!} \pi^1 q^{n-1} + \cdots + n^2 \frac{n!}{n!(n-n)!} \pi^n q^{n-n} \\
&= 0 + 1 \frac{n!}{(n-1)!} \pi q^{n-1} + 2 \frac{n!}{1 \cdot (n-2)!} \pi^2 q^{n-2} + \cdots + n \frac{n!}{(n-1)!(n-n)!} \pi^n \\
&= n\pi \Big[1 \frac{(n-1)!}{(n-1)!} \pi^0 q^{n-1} + 2 \frac{(n-1)!}{1(n-2)!} \pi^2 q^{n-2} + \cdots + n \frac{(n-1)!}{(n-1)!(n-n)!} \pi^{n-1} \Big] \\
&= n\pi \Big[1 \frac{(n-1)!}{(1-1)!((n-1)-(-1-1))!} \pi^{1-1} q^{n-1} + \cdots + n \frac{(n-1)!}{(n-1)!((n-1)-(n-1))!} \pi^{n-1} q^{(n-1)-(n-1)} \Big] \\
&= n\pi \sum_{x=1}^{n} x \binom{n-1}{x-1} \pi^{x-1} 1^{(n-1)-(x-1)} \\
{}^{[2]} &= \sum_{y=0}^{m} (y+1) \binom{m}{y} \pi^y q^{m-y} \\
&= n\pi \Big[\sum_{y=0}^{m} y \binom{m}{y} \pi^y q^{m-y} + \binom{m}{y} \pi^y q^{m-y} \Big] \\
&= n\pi \Big[\sum_{y=0}^{m} y \binom{m}{y} \pi^y q^{m-y} + \sum_{y=0}^{m} \binom{m}{y} \pi^y q^{m-y} \Big] \\
{}^{[3]} &= n\pi(m\pi + 1) \\
&= n\pi[(n-1)\pi + 1] \\
&= n\pi(n\pi - \pi + 1) \\
&= n^2\pi^2 - n\pi^2 + n\pi
\end{aligned}
$$

1. $q = (1-\pi)$
2. Let $y = x - 1$ and $n = m + 1$
   $\Rightarrow x = y + 1$ and $m = n - 1$
3. $\sum_{y=0}^{m} y\binom{m}{y}\pi^y q^{m-y}$ is of the form of the expected value of Y, and $E(Y) = m\pi = (n-1)\pi$.
   $\sum_{y=0}^{m} \binom{m}{y}\pi^y q^{m-y}$ is the sum of all probabilities over the domain of Y which is 1.

$$
\begin{aligned}
\mu &= E(X) \\
&= n\pi
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= E(X^2) - E(X)^2 \\
&= n^2\pi^2 - n\pi^2 + n\pi - n^2\pi^2 \\
&= -n\pi^2 + n\pi \\
&= n\pi(-\pi - 1) \\
&= n\pi(1 - \pi)
\end{aligned}
$$

4.4 Moment Generating Function

$$
\begin{aligned}
M_X(t) = E(e^{tX}) &= \sum_{x=0}^{n} e^{tx} p(x) \\
&= \sum_{x=0}^{n} e^{tx} \binom{n}{x} \pi^x (1-\pi)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} e^{tx} \pi^x (1-\pi)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (\pi e^{tx})^x (1-\pi)^{n-x} \\
{}^{[1]} &= [(1-\pi) + \pi e^t]^n
\end{aligned}
$$

1. By the Binomial Theorem (6.1.2), $\sum\limits_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$

$$M_X^{(1)}(t) = n[(1-\pi) + \pi e^t]^{n-1}\pi e^t$$

$$\begin{aligned} M_X^{(2)}(t) &= n[(1-\pi) + \pi e^t]^{n-1}\pi e^t + n(n-1)[(1-\pi) + \pi e^t]^{n-2}(\pi e^t)^2 \\ &= n\pi e^t[(1-\pi) + \pi e^t]^{n-1} + n(n-1)\pi e^{2t}[(1-\pi) + \pi e^t]^{n-2} \end{aligned}$$

$$\begin{aligned} E(X) &= M_X^{(1)}(0) \\ &= n[(1-\pi) + \pi e^0]^{n-1}\pi e^0 \\ &= n[1-\pi+\pi^{n-1}\pi \\ &= n(1)^{n-1}\pi & = n\pi \end{aligned}$$

$$\begin{aligned} E(X^2) &= M_X^{(2)}(0) \\ &= n\pi e^0[(1-\pi) + \pi e^0]^{n-1} + n(n-2)\pi e^{2\cdot 0}[(1-\pi) + \pi e^0]^{n-2} \\ &= n\pi(1-\pi+\pi)^{n-2} + n(n-1)\pi^2(1-\pi+\pi^{n-2} \\ &= n\pi(1)^{n-1} + n(n-1)\pi^2(1)^{n-2} \\ &= n\pi + n(n-1)\pi^2 \\ &= n\pi + (n^2-n)\pi^2 \\ &= n\pi + n^2 + n^2\pi^2 - n\pi^2 \end{aligned}$$

$$\begin{aligned} \mu &= E(X) \\ &= n\pi \end{aligned}$$

$$\begin{aligned} \sigma^2 &= E(X^2) - E(X)^2 \\ &= n\pi + n^2\pi^2 - n\pi^2 - n^2\pi^2 \\ &= n\pi - n\pi^2 \\ &= n\pi(1-\pi) \end{aligned}$$

4.5 Maximum Likelihood Estimator

Since n is fixed in each Binomial experiment, and must therefore be given, it is unnecessary to develop an estimator for n. The mean and variance can both be estimated from the single parameter π.

Let X be a Binomial random variable with parameter π and n outcomes $(x_1, x_2, \ldots, x_n)$. Let $x_i = 0$ for a failure and $x_i = 1$ for a success. In other words, X is the sum of n Bernoulli trials with equal probability of success and $X = \sum\limits_{i=1}^{n} x_i$.

4.5.1 Likelihood Function

$$
\begin{aligned}
L(\theta) &= L(x_1, x_2, \dots, x_n|\theta)\\
&= P(x_1|\theta)P(x_2|\theta)\cdots P(x_n|\theta)\\
&= [\theta^{x_1}(1-\theta)^{1-x_1}][\theta^{x_2}(1-\theta)^{1-x_2}]\cdots[\theta^{x_n}(1-\theta)^{1-x_n}]\\
&= \exp_\theta\left\{\sum_{i=1}^n x_i\right\}\exp_{(1-\theta)}\left\{n-\sum_{i=1}^n x_i\right\}\\
&= \theta^X(1-\theta)^{n-X}
\end{aligned}
$$

4.5.2 Log-likelihood Function

$$
\begin{aligned}
\ell(\theta) &= \ln L(\theta)\\
&= \ln\left(\theta^X(1-\theta)^{n-X}\right)\\
&= X\ln(\theta) + (n-X)\ln(1-\theta)
\end{aligned}
$$

4.5.3 MLE for π

$$
\frac{d\ell(p)}{d\pi} = \frac{X}{\pi} - \frac{n-X}{1-\pi}
$$

$$
\begin{aligned}
0 &= \frac{X}{\pi} - \frac{n-X}{1-\pi}\\
\Rightarrow \frac{X}{\pi} &= \frac{n-X}{1-\pi}\\
\Rightarrow (1-\pi)X &= \pi(n-X)\\
\Rightarrow X - \pi X &= n\pi - \pi X\\
\Rightarrow X &= n\pi\\
\Rightarrow \frac{X}{n} &= \pi
\end{aligned}
$$

So $\hat{p} = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^n x_i$ is the maximum likelihood estimator for π.

4.6 Theorems for the Binomial Distribution

4.6.1 Validity of the Distribution

$$
\sum_{x=0}^n \binom{n}{x} p^x(1-p)^{n-x} = 1
$$

Proof:

$$\begin{aligned}
\sum_{x=0}^{n} \binom{n}{x} \pi^x (1-\pi)^{n-x} \\
&\overset{[1]}{=} \big(\pi + (1-\pi)\big)^n \\
&= (1)^n \\
&= 1
\end{aligned}$$

1. By the Binomial Theorem (6.1.2), $\sum\limits_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$

4.6.2 Sum of Binomial Random Variables

Let $X_1, X_2, \ldots, X_k$ be independent random variables where X_i comes from a Binomial distribution with parameters n_i and π. That is $X_i \sim (n_i, \pi)$.

Let $Y = \sum\limits_{i=1} kX_i$. Then Y ~Binomial($\sum\limits_{i=1}^{k} n_i, \pi$).

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{t(x_1 + X_2 + \cdots + X_k)} \\
&= E(e^{tX_1} e^{tX_2} \cdots e^{tX_k}) \\
&= E(e^{tX_1}) E(e^{tX_2}) \cdots E(e^{tX_k}) \\
&= \prod_{i=1}^{k} [(1-\pi) + \pi e^t]^{n_i} \\
&= [(1-\pi) + \pi e^t]^{\sum\limits_{i=1}^{k} n_i}
\end{aligned}$$

Which is the mgf of a Binomial random variable with parameters $\sum\limits_{i=1}^{k} n_i$ and π.

Thus Y ~Binomial($\sum\limits_{i=1}^{k} n_i, \pi$).

4.6.3 Sum of Bernoulli Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from a Bernoulli distribution with parameter π. Let $Y = \sum\limits_{i=1}^{n} X_i$.
Then Y ~Binomial(n, π)

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{tX_1} e^{tX_2} \cdots e^{tX_n}) \\
&= E(e^{tX_1}) E(e^{tX_2}) \cdots E(e^{tX_n}) \\
&= (\pi e^t + (1-\pi))(\pi e^t + (1-\pi)) \cdots (\pi e^t + (1-\pi)) \\
&= (\pi e^t + (1-\pi))^n
\end{aligned}$$

Which is the mgf of a Binomial random variable with parameters n and π. Thus, $Y \sim$ Binomial(n, π).

Chapter 5

Binomial Test

5.1 Binomial Test

The binomial test is used to look for evidence that the proportion of a Binomial distributed random variable may differ from a hypothesized (or previously observed) value.

5.1.1 Test Statistic

The test statistic for a binomial test is the observed frequency of experimental subjects that exhibit the trait of interest.

5.1.2 Definitions

Let X be a random variable following a binomial distribution with parameters n and π. Let x be the observed frequency of experimental subjects exhibiting the trait of interest.

5.1.3 Hypotheses

The hypotheses for the Binomial test may take the following forms:

For a two-sided test:

$$H_0 : \pi = \pi_0$$
$$H_a : \pi \neq \pi_0$$

For a one-sided test:

$$H_0 : \pi < \pi_0$$
$$H_a : \pi \geq \pi_0$$

or

$$H_0 : \pi > \pi_0$$
$$H_a : \pi \leq \pi_0$$

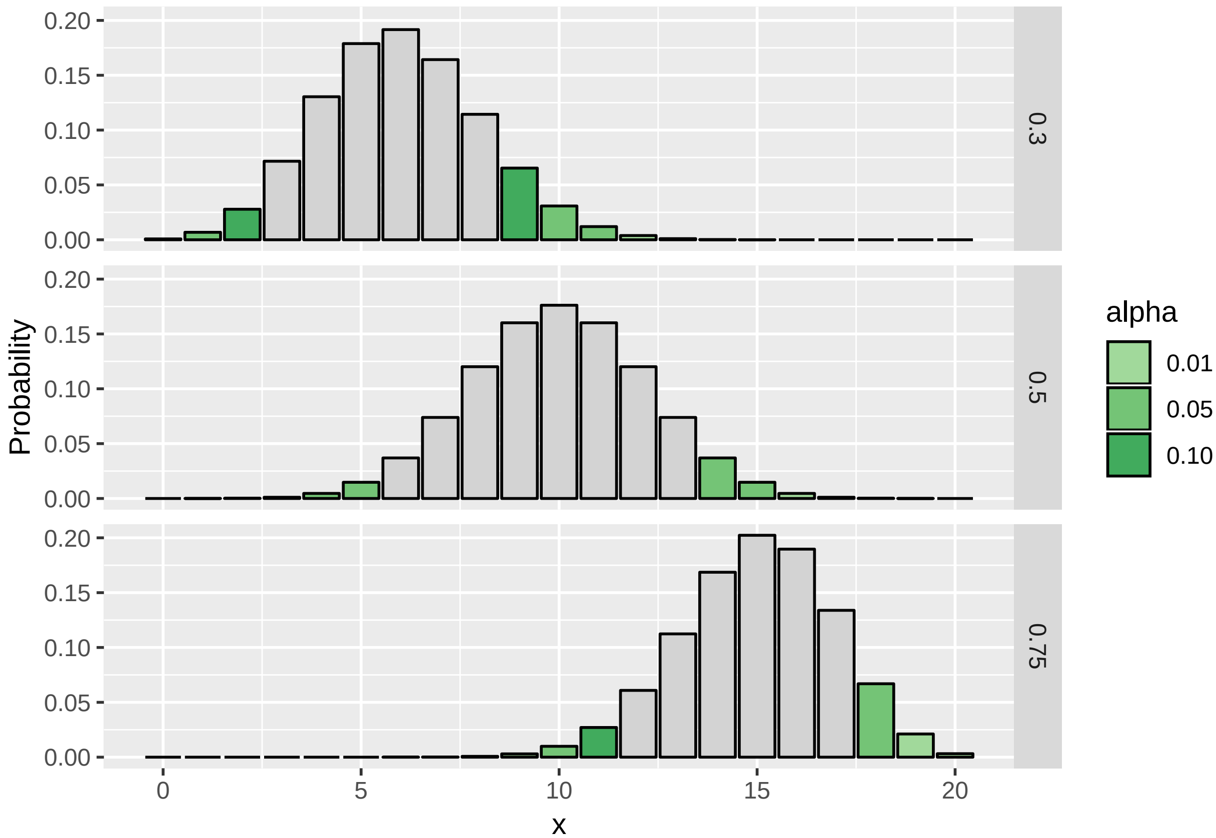

Figure 5.1: The examples displayed use $n = 20$. For the top, middle, and bottom examples, π is set at 0.3, 0.5, and 0.75, respectively. Notice that in some cases, the rejection regions for $\alpha = 0.10$ and $\alpha = 0.05$ are identical.

5.1.4 Decision Rule

The decision to reject the null hypothesis is made when the observed value of x lies in the critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

In a two-sided test, the upper bound is shared equally in both tails. Due to the discrete nature of the distribution, the total probability in the tails may not equal α. The figures below depict examples of rejection regions for selected values of the Binomial distribution parameters. The decision rule is:

Reject H_0 if $x < Binomial(\alpha/2, n, \pi_0)$ or $x > Binomial(1 - \alpha/2, n, \pi_0)$

In the one-sided test, α is placed in only one tail. The figures below depict examples of rejection regions for selected values of the Binomial distribution parameters. In each case, α is the area in the tail of the figure. It follows, then, that the decision rule for a lower tailed test is:

Reject H_0 when $x \leq Binomial(\alpha, n, \pi_0)$

For an upper tailed test, the decision rule is:

Reject H_0 when $x \geq Binomial(1 - \alpha, n, \pi_0)$

5.1.5 Power

The derivations below make use of the following symbols:

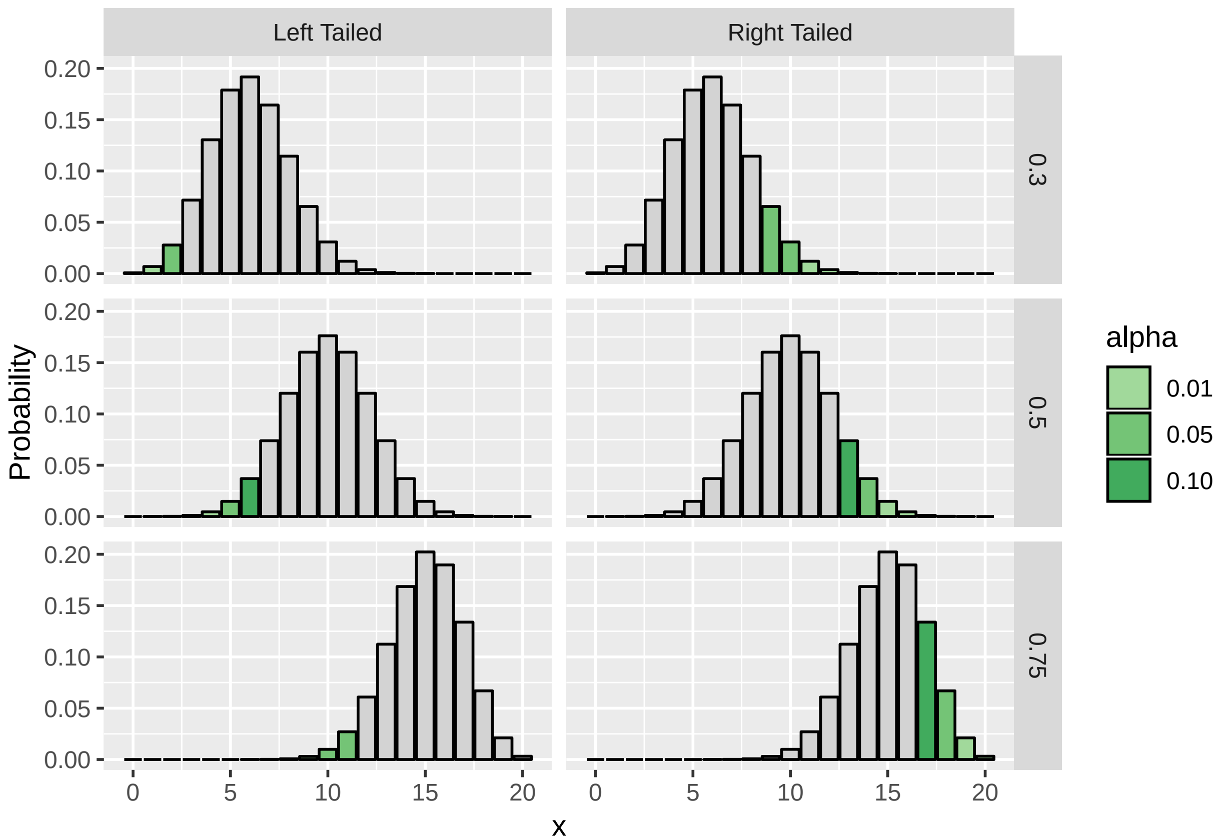

Figure 5.2: The examples displayed use $n = 20$. For the top, middle, and bottom examples, π is set at 0.3, 0.5, and 0.75, respectively.

- x: The observed frequency of experimental units exhibiting the trait of interest.
- n: The total number of experimental units.
- π_0: The proportion of the population that exhibits the trait of interest under the null hypothesis.
- π_a: The proportion of the population that exhibits the trait of interest under the alternative hypothesis.
- α: The significance level.
- $\gamma(\pi)$: The power of the test for the parameter π.
- $Binomial(\alpha, n, \pi)$: A quantile of the Binomial distribution with a probability α, and parameters n and π.
- C: The critical region.

Two Sided Test

$$\begin{aligned}\gamma(\pi_a) &= P_{\pi_a}(x \in C)\\ &= P_{\pi_a}(Binomial(\alpha/2, n, \pi_0) \leq Binomial(\alpha/2, n, \pi_a)) +\\ &\qquad P_{\pi_a}(Binomial(1-\alpha/2, n, \pi_0) \geq Binomial(1-\alpha/2, n, \pi_a))\end{aligned}$$

Left Sided Test

$$\begin{aligned}\gamma(\pi_a) &= P_{\pi_a}(x \in C)\\ &= P_{\pi_a}(Binomial(\alpha, n, \pi_0) \leq Binomial(\alpha, n, \pi_a))\end{aligned}$$

Right Sided Test

$$\begin{aligned}\gamma(\pi_a) &= P_{\pi_a}(x \in C)\\ &= P_{\pi_a}(Binomial(1-\alpha, n, \pi_0) \geq Binomial(1-\alpha, n, \pi_a))\end{aligned}$$

Since the Binomial distribution is discrete, the power curve has the interesting characteristic of not being monotonic. It is sometimes described as having a “sawtooth” appearance. This behavior means that a larger sample size is not always preferred. For example, in the following figure, a sample size of 10 has better power than a sample size of 12.

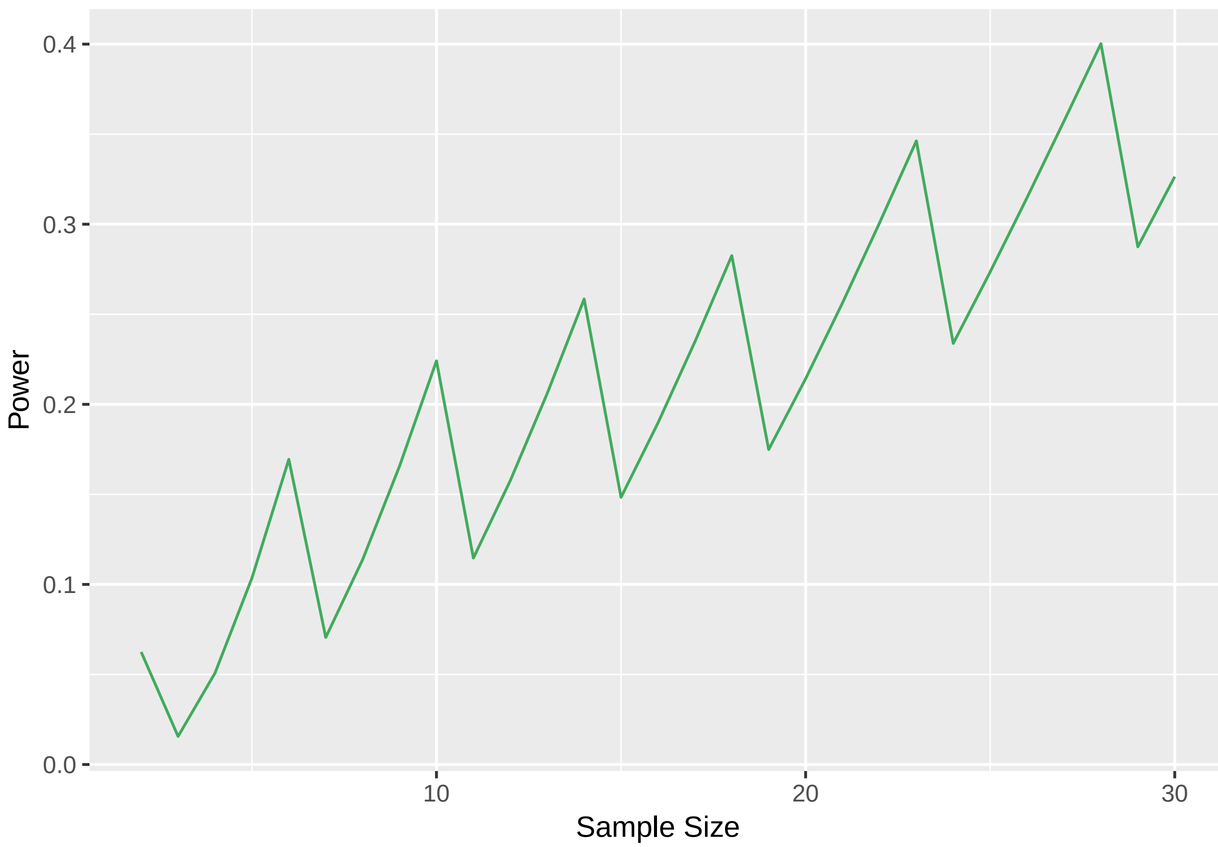

Figure 5.3: Power for a Binomial test with $\pi_0 = .15$ and $\pi_a = 0.25$

Chapter 6

Binomial Theorem

The Binomial Theorem is useful in developing theory around the Binomial and Hypergeometric Distributions. Two proofs of the Theorem are provided here; one using the traditional approach, and one using a more general approach. Other useful theorems are provided at the end of this chapter.

6.1 Traditional Proof

6.1.1 Lemma: Pascal's rule

Let n and x be non-negative integers such that $x \le n$.

Then $\binom{n-1}{x} + \binom{n-1}{x-1} = \binom{n}{x}$.

Proof:

$$\begin{aligned}
\binom{n-1}{x} + \binom{n-1}{x-1} &= \frac{(n-1)!}{x!(n-1-x)!} + \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} \\
&= \frac{(n-1)!}{x!(n-x-1)!} + \frac{(n-1)!}{(x-1)!(n-1-x+1)!} \\
&= \frac{(n-1)!}{x!(n-x-1)!} + \frac{(n-1)!}{(x-1)!(n-x)!} \\
&= \frac{(n-1)!}{x(x-1)!(n-x-1)!} + \frac{(n-1)!}{(x-1)!(n-x)(n-x-1)!} \\
&= \frac{x(n-1)!}{x(x-1)!(n-x)(n-x-1)!} + \frac{(n-x)(n-1)!}{x(x-1)!(n-x)(n-x-1)!} \\
&= \frac{x(n-1)! + (n-x)(n-1)!}{x(x-1)!(n-x)(n-x-1)!} \\
&= \frac{(x+n-x)(x-1)!}{x(x-1)!(n-x)(n-x-1)!} \\
&= \frac{n(n-1)!}{x(x-1)!(n-x)(n-x-1)!} \\
&= \frac{n!}{x!(n-x)!} \\
&= \binom{n}{x}
\end{aligned}$$

6.1.2 The Binomial Theorem

Let a and b be constants and let n be any positive integer. Then

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x$$

Proof:

This proof is completed by mathematical induction.

Base Step: $n = 1$

$$\begin{aligned}
(a+b)^1 &= \sum_{x=0}^{1} \binom{1}{x} a^{1-x} b^x \\
&= \binom{1}{0} a^{1-0} b^0 + \binom{1}{1} a^{1-1} b^1 \\
&= 1 \cdot a \cdot 1 + 1 \cdot 1 \cdot b \\
&= a + b
\end{aligned}$$

Inductive Step: Assume that the Theorem holds for n, and show it is true for $n+1$.

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)(a+b)^n \\
&= a(a+b)^n + b(a+b)^n \\
&= a(a^n + \sum_{x=1}^{n-1} \binom{n}{x} a^{n-x} b^x + b^n) + b(a^n + \sum_{x=1}^{n-1} \binom{n}{x} a^{n-x} b^x + b^n) \\
&= (a^{n+1} + a \sum_{x=1}^{n-1} \binom{n}{x} a^{n-x} a b^x) + (a^n b + \sum_{x=1}^{n-1} \binom{n}{x} a^{n-x} b^x + b^{n+1}) \\
&= (a^{n+1} + \sum_{x=1}^{n-1} \binom{n}{x} a^{n-x+1} a b^x) + (a^n b + \sum_{x=1}^{n-1} \binom{n}{x} a^{n-x} b^{x+1} + b^{n+1}) \\
{}^{[1]} &= (a^{n+1} + \sum_{x=1}^{n} a^{n-x+1} b^x) + (\sum_{x=0}^{n-1} \binom{n}{x} a^{n-x} b^{x+1} + b^{n+1}) \\
{}^{[2]} &= (a^{n+1} + \sum_{x=1}^{n} \binom{n}{x} a^{n-x+1} b^x) + \sum_{x-1}^{n-1} \binom{n}{x-1} a^{n-x+1} b^{x+1-1} + b^{n+1}) \\
{}^{[3]} &= a^{n+1} + \sum_{x+1}^{n} \binom{n+1}{x} a^{n-x+1} b^x + b^{n+1} \\
&= a^{n+1} + \sum_{x=1}^{n} \binom{n+1}{x} a^{(n+1)-x} b^x + b^{n+1} \\
{}^{[4]} &= \sum_{x=0}^{n+1} \binom{n+1}{x} a^{(n+1)-x} b^x
\end{aligned}$$

This completes both the inductive step and the proof.

1. $ab^n = \binom{n}{n} a^{n-n+1} b^n$ which is the term for $x = n$ in the first summation.
   $a^n b = \binom{n}{0} a^{n-0} b^1$ which is the term for $x = 0$ in the second summation.

2. $\sum_{x=0}^{n-1} \binom{n}{x} a^{n-x} b^{x+1}$
   $= \sum_{x=1}^{n} \binom{n}{x-1} a^{n-(x-1)} b^{(x-1)+1}$
   $= \sum_{x=1}^{n} \binom{n}{x-1} a^{n-x+1} b^{x}$
3. This step is made using Pascal's Rule with $n = n - 1$.
4. $a^{n+1} = \binom{n+1}{0} a^{(n+1)-0} b^0$ which is the term for $x = 0$ in the summation.
   $b^{n+1} = \binom{n+1}{n+1} a^{(n+1)-(n+1)} b^{n+1}$ which is the term for $x = n + 1$ in the summation

6.2 General Approach

6.2.1 A Binomial Expansion Theorem

This theorem and its corrolary are provided by Brunette (Brunette 2003–2007a).

For any positive integer n, let $B_n = (x_1 + y_1)(x_2 + y_2) \cdots (x_n + y_n)$. In the expansion B_n, before combining possible like terms, the following are true:

i. There will be 2^n terms.
ii. Each of these terms will be a product of n factors.
iii. In each such product there will be one factor from each binomial (in B_n).
iv. Every such product of n factors, one from each binomial, is represented in the expansion.

Proof:

Proof is done by induction.

For the case $n = 1$, the result is clear.

Now assume that the theorem is true for a particular n and consider B_{n+1}.

$$B_{n+1} = B_n(x_{n+1} + y_{n+1}) = B_n x_{n+1} + B_n y_{n+1}$$

By the inductive assumption, $B_n = T_1 + T_2 + \cdots + T_{2^n}$ where each T_i is a product of n factors, one factor from each binomial. It follows that every term in the expansion of $B_n + 1$ is either of the type $T_i x_{n+1}$ or $T_i y_{n+1}$, for some $1 \leq i \leq 2^n$. But each term of either of the above types is clearly a product of $n + 1$ factors with one factor coming from each binomial. thus, if *(ii)* and *(iii)* are true for B_n, then they are true for $B_n + 1$.

Next, by the inductive assumption, the expansion of B_n is a sum of $2^n + 2^n$ terms, i.e., 2^{n+1} terms. This completes the inductive step for *(i)*.

Lastly, it remains for us to consider a product of the type $p_1 p_2 \cdots p_n p_{n+1}$ where, for each $1 \leq i \leq n + 1$, $p_i = x_i$ or $p_i = y_i$. By the inductive hypothesis, $p_1 p_2 \cdots p_n$ is a term in the expansion of B_n. If $p_{n+1} = x_{n+1}$, then $p_1 p_2 \cdots p_n p_{n+1}$ is a term in the expansion of $B_n x_{n+1}$, and so of B_{n+1}. Likewise, if $p_{n+1} = y_{n+1}$, then $p_1 p_2 \cdots p_n p_{n+1}$ is a term in the expansion of $B_n y_{n+1}$, and so of B_{n+1}. This completes the inductive step and the proof.

6.2.2 Corollary: Binomial Theorem

Let x and y be constants and let n be any positive integer.

Then $(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$

Proof:

Since each term in the expansion will have n terms, each term must follow the form $x^{n-i}y^i$ for $0 \leq i \leq n$, and in all, there are 2^n such terms. For any given value of i, the number of terms of the form $x^{n-i}y^i$ is clearly the number of ways one can choose the i factors of y from the n available binomials, i.e., $\binom{n}{i}$, which gives

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$$

6.3 Other Theorems

6.3.1 Theorem

$$\binom{N_1}{0}\binom{N_2}{n} + \binom{N_1}{2}\binom{N_2}{n-1} + \cdots + \binom{N_1}{n-1}\binom{N_2}{1} + \binom{N_1}{n}\binom{N_2}{0} = \binom{N_1+N_2}{n}$$

where $0 \leq n \leq N_1 + N_2$.

Proof:

Using the Binomial Theorem we establish

$$(1+a)^{N-1}(1+a)^{N_2} = (1+a)^{N_1+N_2} \Rightarrow [\binom{N_1}{0}a^0 + \cdots + \binom{N_1}{N_1}a^{N_1}]\cdot[\binom{N_2}{0}a^0 + \cdots + \binom{N_2}{N_2}a^{N_2}] \quad = \binom{N_1+N_2}{0} + \binom{N_1+N_2}{1}a + \cdot$$

Expanding the left side of the equation gives

$$\binom{N_1}{0}\binom{N_2}{0} + \binom{N_1}{0}\binom{N_2}{1}a + \cdots + \binom{N_1}{0}\binom{N_2}{N_2}a^{N_2} + \binom{N_1}{1}\binom{N_2}{0}a \quad + \cdots + \binom{N_1}{1}\binom{N_2}{N_2}a^{N_2+1} + \cdots + \binom{N_1}{N_1}\binom{N_2}{0}a^{N_1} + \binom{N_1}{N_1}($$

Notice that for any n where $0 \leq n \leq N_1 + N_2$, the coefficient for a^n, found by combining like terms, is $\binom{N_1}{0}\binom{N_2}{n} + \binom{N_1}{1}\binom{N_2}{n-1} + \cdots + \binom{N_1}{n-1}\binom{N_2}{1} + \binom{N_1}{0}\binom{N_2}{n}$ and, by the equivalence of the first equation in the proof, is equal to the coefficient $\binom{N_1+N_2}{n}$.

6.3.2 Theorem

$$\frac{\sum_{i=1}^{n}\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}} = 1$$

for $0 \leq n \leq N_1 + N_2$.\

Proof:

Theorem 6.3.1 establishes the equality

$$\binom{N_1}{0}\binom{N_2}{n} + \binom{N_1}{2}\binom{N_2}{n-1} + \cdots + \binom{N_1}{n-1}\binom{N_2}{1} + \binom{N_1}{n}\binom{N_2}{0} = \binom{N_1+N_2}{n} \Rightarrow \sum_{i=1}^{n}\binom{N_1}{i}\binom{N_2}{n-i} = \binom{N_1+N_2}{n} \Rightarrow \frac{\sum_{i=1}^{n}(}{(}$$

Chapter 7

Central Limit Theorem

7.1 Theorem: Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$.

Define

$$U = \sqrt{n}\Big(\frac{\bar{X} - \mu}{\sigma}\Big)$$

where

$$\bar{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n}$$

Then the distribution function of U converges to a standard normal distribution function as $n \to \infty$.

Proof:

Recall the definition of the moment generating function 27.2.1 for a random variable X can be written

$$1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$$

Let $Z_i = \frac{X_i - \mu}{\sigma}$. We observe the following preliminary results:

Preliminary 1

$$
\begin{aligned}
E(Z_i) &= E\Big(\frac{X_i - \mu}{\sigma}\Big) \\
&= \frac{1}{\sigma} \cdot E(X_i - \mu) \\
&= \frac{1}{\sigma} \cdot \big(E(X_i) - \mu\big) \\
&= \frac{1}{\sigma} \cdot (\mu - \mu) \\
&= \frac{1}{\sigma} \cdot 0 \\
&= 0
\end{aligned}
$$

Preliminary 2

$$
\begin{aligned}
V(Z_i) &= V\Big(\frac{X_i - \mu}{\sigma}\Big) \\
&= \frac{1}{\sigma^2} V(X_i - \mu) \\
&= \frac{1}{\sigma^2} \cdot \big(V(X_i) - V(\mu)\big) \\
&= \frac{1}{\sigma^2} \cdot \big(V(X_i) - 0\big) \\
&= \frac{1}{\sigma^2} \cdot V(X_i) \\
&= \frac{1}{\sigma^2} \cdot \sigma^2 \\
&= \frac{\sigma^2}{\sigma^2} \\
&= 1 \\
\Rightarrow E(Z^2) - E(Z)^2 &= 1 \\
\Rightarrow E(Z^2) &= 1 - E(Z)^2 \\
{}^{[1]} &= 1 - 0 \\
&= 1
\end{aligned}
$$

From Preliminary 1, $E(Z) = 0$.

Preliminary 3

$$
\begin{aligned}
U &= \sqrt{n}\Big(\frac{\bar{X} - \mu}{\sigma}\Big) \\
&= \sqrt{n}\Big(\frac{\frac{\sum_{i=1}^{n} X_i}{n} - \mu}{\sigma}\Big) \\
&= \sqrt{n}\Big(\frac{\frac{\sum_{i=1}^{n} X_i}{n} - \frac{n \cdot \mu}{n}}{\sigma}\Big) \\
&= \sqrt{n}\Big(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma}\Big) \\
&= \frac{\sqrt{n}}{n}\Big(\frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma}\Big) \\
&= \frac{1}{\sqrt{n}}\Big(\frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma}\Big) \\
&= \frac{1}{\sqrt{n}}\Big(\frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma}\Big) \\
&= \frac{1}{\sqrt{n}}\Big(\frac{\sum_{i=1}^{n} (X_i - \mu)}{\sigma}\Big) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i
\end{aligned}
$$

Note that, due to the independence of the X_i random variables, it follows that the Z_i random variables are also independent.

The moment generating function of U is derived as

$$
\begin{aligned}
M_U(t) &= M_Z\Big(t\sum_{i=1}^{n}\frac{Z_i}{\sqrt{n}}\Big)\\
&= M_Z\Big(\frac{t}{\sqrt{n}}\sum_{i=1}^{n}Z_i\Big)\\
&= E\Big(\exp\Big\{\sum_{i=1}^{n}\frac{t}{\sqrt{n}}Z_i\Big\}\Big)\\
&= E\Big(\prod_{i=1}^{n}\exp\Big\{\frac{t}{\sqrt{n}}Z_i\Big\}\Big)\\
&= \prod_{i=1}^{n}E\Big(\exp\Big\{\frac{t}{\sqrt{n}}Z_i\Big\}\Big)\\
&= \prod_{i=1}^{n}M_{Z_i}(t)\\
{}^{[1]} &= \prod_{i=1}^{n}M_Z(t)\\
&= \big(M_{Z_i}(t)\big)^n\\
{}^{[2]} &= \Big(1+\frac{t}{\sqrt{n}}E(Z)+\frac{1}{2!}\cdot(\frac{t}{\sqrt{n}})^2E(Z^2)+\frac{1}{3!}\cdot(\frac{t}{\sqrt{n}})^3E(Z^3)+\cdots\Big)^n\\
{}^{[3]} &= \Big(1+\frac{t}{\sqrt{n}}\cdot 0+\frac{t}{2!\cdot n}E(Z^2)+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)^n\\
&= \Big(1+0+\frac{t}{2!\cdot n}E(Z^2)+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)^n\\
&= \Big(1+\frac{t}{2!\cdot n}E(Z^2)+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)^n\\
{}^{[4]} &= \Big(1+\frac{t}{2!\cdot n}\cdot 1+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)\\
&= \Big(1+\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)^n\\
\Rightarrow \ln M_U(t) &= \ln\Big[\Big(1+\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)^n\Big]\\
&= n\cdot\ln\Big(1+\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots\Big)\\
{}^{[5]} &= n\cdot\ln(1+k)\\
{}^{[6]} &= n\cdot(k-\frac{k^2}{2}+\frac{k^3}{3}-\frac{k^4}{4})\\
{}^{[7]} &= n\cdot\Big((\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)-\frac{1}{2}(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)^2+\cdots\Big)\\
&= n\cdot(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)-\frac{n}{2}(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)^2+\cdots\\
&= (\frac{n\cdot t}{2!\cdot n}+\frac{n\cdot t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)-\frac{n}{2}(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)^2+\cdots\\
&= \frac{n\cdot t}{2!\cdot n}+(\frac{n\cdot t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)-\frac{n}{2}(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)^2+\cdots\\
&= \frac{t}{2!}+(\frac{n\cdot t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)-\frac{n}{2}(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)^2+\cdots\\
&= \frac{t}{2}+(\frac{n\cdot t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)-\frac{n}{2}(\frac{t}{2!\cdot n}+\frac{t^3}{3!\cdot n^{\frac{3}{2}}}E(Z^3)+\cdots)^2+\cdots
\end{aligned}
$$

1. Each of the Z_i are identically distributed, so they each have the same moment generating function.
2. Definition of the moment generating function @ref{moment-definition-mgf}
3. From Preliminary 1, $E(Z) = 0$.
4. From Preliminary 2, $E(Z^2) = 1$.
5. Let $k = \frac{t}{2! \cdot n} + \frac{t^3}{3! \cdot n^{\frac{3}{2}}} E(Z^3) + \cdots$
6. $\ln(1+x)$ may be rewritten with a series expansion: $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$

To complete the proof, we take the limit as $n \to \infty$.

$$\lim_{n\to\infty} \ln[M_Z(t)] = \lim_{n\to\infty} \left[\frac{t}{2} + \big(\frac{n \cdot t^3}{3! \cdot n^{\frac{3}{2}}} E(Z^3) + \cdots \big) - \frac{n}{2} \big(\frac{t}{2! \cdot n} + \frac{t^3}{3! \cdot n^{\frac{3}{2}}} E(Z^3) + \cdots \big)^2 + \cdots \right]$$

By noting that, with the exception of the $\frac{t}{2}$ term, every term in the series involves n in the denominator. Additionally, in each of those terms, the n in the denominator is larger than the n in the numerator. Thus, as $n \to \infty$, each of those terms approaches 0, yielding.

$$\begin{aligned} \lim_{n\to\infty} \ln[M_Z(t)] &= \frac{t}{2} \\ \Rightarrow \lim_{n\to\infty} \exp\{\ln[M_Z(t)]\} &= \exp\left\{ \frac{t}{2} \right\} \\ \Rightarrow \lim_{n\to\infty} M_Z(t) &= e^{\frac{t}{2}} \end{aligned}$$

This is the form of the moment generating function for a normal random variable with $\mu = 0$ and $\sigma^2 = 1$.

7.2 References

D Wackerly, W Mendenhall, R Scheaffer, *Mathematical Statistics with Applications* 6th ed., Duxbury Thomson Learning, 2002 pp 352 - 354.

Chapter 8

Chebychev's Theorem

8.1 Chebychev's Theorem

In any finite set of numbers and for any real number $h > 1$, at least $(1 - \frac{1}{h^2}) \cdot 100\%$ of the numbers lie within h standard deviations of the mean. In other words, they lie within the interval $(\mu - h \cdot \sigma, \mu + h \cdot \sigma)$.

Proof:

For a set $\{x_1, x_2, \ldots, x_r, x_{r+1}, \ldots, x_n\}$ where, by choice of labeling, $\{x_1, x_2, \ldots, x_r\}$ lie outside of $(\mu - h \cdot \sigma, \mu + h \cdot \sigma)$. Also, $\{x_{r+1}, \ldots, x_n\}$ are within the interval. Under these conditions we know

$$|x_1 - \mu| > h\sigma,\ |x_2 - \mu| > h\sigma, \ldots,\ |x_r - \mu| > h\sigma$$

Squaring gives

$$(x_1 - \mu)^2 > h^2\sigma^2,\ (x_2 - \mu)^2 > h^2\sigma^2, \ldots,\ (x_r - \mu)^2 > h^2\sigma^2 \qquad \Rightarrow \sum_{i=1}^{r}(x_1 - \mu)^2 > \sum_{i=1}^{r} h^2\sigma^2 = rh^2\sigma^2$$

Since all $(x_i - \mu)^2$ must necessarily be positive,

$$\begin{array}{rrcl} & \sum_{i=1}^{r}(x_i - \mu)^2 & < & \sum_{i=1}^{n}(x_i - \mu)^2 \\ \Rightarrow & rh^2\sigma^2 & < & \sum_{i=1}^{n}(x_i - \mu)^2 \\ {}^{[1]} \Rightarrow & rh^2\sigma^2 & < & n\sigma^2 \\ \Rightarrow & rh^2 & < & n \\ \Rightarrow & \frac{r}{n} & < & \frac{1}{h^2} \end{array}$$

1. $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

   $\Rightarrow n\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2$

and $\frac{r}{n}$ is the fraction of numbers outside $(\mu - h \cdot \sigma, \mu + h \cdot \sigma)$. By the law of complements, the fraction of numbers inside the interval is $1 - \frac{r}{n}$, which implies $1 - \frac{r}{n} > 1 - \frac{1}{h^2}$. Thus, more than $(1 - \frac{1}{h^2}) \cdot 100\%$ of the points lie within h standard deviations of the mean, or within the interval $(\mu - h \cdot \sigma, \mu + h \cdot \sigma)$.

8.2 Alternate Proof of Chebychev's Theorem

In any finite set of numbers and for any real number $h > 1$, at least $(1 - \frac{1}{h^2}) \cdot 100\%$ of the numbers lie within h standard deviations of the mean. In other words, they lie within the interval $(\mu - h \cdot \sigma, \mu + h \cdot \sigma)$.\

Proof:

The proof here is done for the discrete case, but is applicable also in the continuous case by replacing the summations with integrals (with integrals, the limits will be from $-\infty$ to ∞).

$$\begin{array}{rrcl} & \sigma^2 & = & E(x-\mu)^2 \\ & & = & \sum_{y=0}^{\infty}(y-\mu)^2 p(y) \\ & & = & \sum_{y=0}^{\mu-h\sigma}(y-\mu)^2 p(y) + \sum_{y=\mu-h\sigma+1}^{\mu+h\sigma-1}(y-\mu)^2 p(y) + \sum_{y=\mu+h\sigma}^{\infty}(y-\mu)^2 p(y) \\ {}^{[1]} \Rightarrow & \sigma^2 & \geq & \sum_{y=0}^{\mu-h\sigma}(y-\mu)^2 p(y) + \sum_{y=\mu+h\sigma}^{\infty}(y-\mu)^2 p(y) \end{array}$$

1. Since all the $(y-\mu)^2$ must be positive, removing the middle term will surely result in this inequality.

In both of these summations y is outside the interval $(\mu - h \cdot \sigma, \mu + h \cdot \sigma)$, so

$$\begin{array}{rrcl} & |y-\mu| & \geq & h\sigma \\ \Rightarrow & (y-\mu^2) & \geq & h^2\sigma^2 \\ \Rightarrow & \sigma^2 & \geq & \sum_{y=0}^{\mu-h\sigma} h^2\sigma^2 \Pr(Y=y) + \sum_{\mu+h\sigma}^{\infty} h^2\sigma^2 \Pr(Y=y) \\ \Rightarrow & \sigma^2 & \geq & h^2\sigma^2 \Big[\sum_{y=0}^{\mu-h\sigma} \Pr(Y=y) + \sum_{\mu+h\sigma}^{\infty} \Pr(Y=y)\Big] \end{array}$$

The first summation is the sum of all probabilities that $y - \mu < h\sigma$, i.e. $P(y - \mu < h\sigma)$. Likewise, the second summation is $P(y - \mu > h\sigma)$.

$$\begin{array}{rrcl} \Rightarrow & \sigma^2 & \geq & h^2\sigma^2[P(y-\mu < h\sigma) + P(y-\mu > h\sigma)] \\ \Rightarrow & \sigma^2 & \geq & h^2\sigma^2[P(|y-\mu| > h\sigma)] \\ \Rightarrow & \frac{1}{h^2} & \geq & \Pr(|y-\mu| > h\sigma) \\ \Rightarrow & 1-\frac{1}{h^2} & \leq & \Pr(|y-\mu| > h\sigma) \end{array}$$

8.3 Chebychev's Theorem for Absolute Deviation

This theorem is provided by Brunette (Brunette 2003–2007b)

In any finite set of numbers, and for any real number $h > 1$, at least $1 - \frac{1}{h}$ of the numbers lie within h absolute deviations of the mean, where the absolute deviation is defined $Ab = \frac{1}{n}\sum_{i=1} n|x_i - \bar{x}|$. In other words, $1 - \frac{1}{h}$ of the numbers are in the interval $(\bar{x} - h \cdot Ab, \bar{x} + h \cdot Ab)$.

Proof:

For a set $\{x_1, x_2, \ldots, x_r, x_{r+1}, \ldots, x_n\}$ where, by choice of labeling, $\{x_1, x_2, \ldots, x_r\}$ lie outside of $(\mu - h \cdot Ab, \mu + h \cdot Ab)$. Also, $\{x_{r+1}, \ldots, x_n\}$ are within the interval. Accordingly,

$$h \cdot Ab \leq |x_1 - \bar{x}|,\ h \cdot Ab \leq |x_1 - \bar{x}|, \ldots,\ h \cdot Ab \leq |x_1 - \bar{x}|$$

$$\begin{array}{rrcl} & \Rightarrow \quad r \cdot h \cdot Ab & \leq & \sum_{i=1}^{r} |x_i - \bar{x}| \\ & \Rightarrow \quad r \cdot h \cdot Ab & \leq & \sum_{i=1}^{n} |x_i - \bar{x}| \\ [1] & \Rightarrow \quad r \cdot h \cdot Ab & \leq & n \cdot Ab \\ & \Rightarrow \quad \frac{r}{n} & \leq & \frac{1}{h} \\ & \Rightarrow \quad -\frac{r}{n} & \geq & -\frac{1}{h} \\ & \Rightarrow \quad 1 - \frac{r}{n} & \geq & 1 - \frac{1}{h} \end{array}$$

1. $Ab = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$

   $\Rightarrow n \cdot Ab = \sum_{i=1}^{n} |x_i - \bar{x}|$

Now $\frac{r}{n}$ is the fraction of numbers outside the interval. So $1 - \frac{r}{n}$ is the fraction of numbers within h absolute deviations of the mean, or within the interval $(\mu - h \cdot Ab, \mu + h \cdot Ab)$.

Chapter 9

Chi-Square Distribution

9.1 Probability Distribution Function

A random variable X is said to have a Chi-Square Distribution with parameter ν if its probability distribution function is

$$f(x) = \begin{cases} \frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} & 0 < x,\ 0 < \nu \\ 0 & otherwise \end{cases}$$

ν is commonly referred to as the *degrees of freedom*.

9.2 Cumulative Distribution Function

The cumulative distribution function for the Chi-Square Distribution cannot be written in closed form. It's integral form is expressed as

$$F(x) = \begin{cases} \int\limits_0^x \frac{t^{\frac{\nu}{2}-1}e^{-\frac{t}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} dt & 0 < x,\ 0 < \nu \\ \\ 0 & otherwise \end{cases}$$

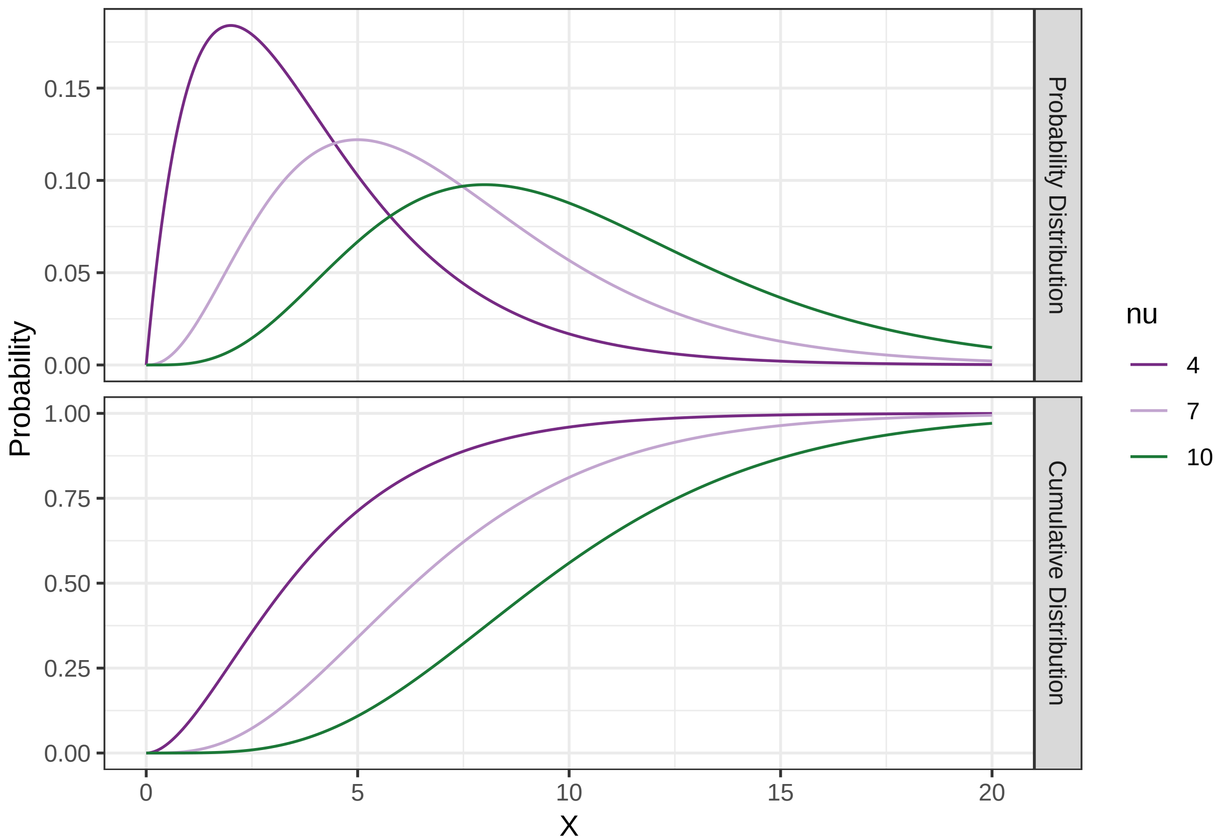

Figure 9.1: (#fig:ChiSquare_Distribution)The graphs on the top and bottom depict the Chi-Square probability distribution and cumulative distribution functions, respectively, for $\nu = 4, 7, 10$. As ν gets larger, the distribution becomes flatter with thicker tails.

9.3 Expected Values

$$
\begin{aligned}
E(X) &= \int_0^\infty x \frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}dx \\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty x \cdot x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}dx \\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty x^{\frac{\nu}{2}}e^{-\frac{x}{2}}dx \\
\overset{[1]}{=}& \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\Big[\Gamma\Big(\frac{\nu}{2}+1\Big)2^{\frac{\nu}{2}+1}\Big] \\
&= \frac{\Gamma(\frac{\nu}{2}+1)2^{\frac{\nu}{2}+1}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \\
&= \frac{\frac{\nu}{2}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}+1}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \\
&= \frac{2\nu}{2} \\
&= \nu
\end{aligned}
$$

1. $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$

$$
\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}dx \\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty x^2 \cdot x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}dx \\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty x^{\frac{\nu}{2}+1}e^{-\frac{x}{2}}dx \\
\overset{[1]}{=}& \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\Big[\Gamma(\frac{\nu}{2}+2)2^{\frac{\nu}{2}+2}\Big] \\
&= \frac{\Gamma\Big(\frac{\nu}{2}+2\Big)2^{\frac{\nu}{2}+2}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \\
&= \frac{(\frac{\nu}{2}+1)\Gamma(\frac{\nu}{2}+1)2^{\frac{\nu}{2}+2}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \\
&= \frac{\Big(\frac{\nu}{2}+1\Big)\frac{\nu}{2}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}+2}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \\
&= \Big(\frac{\nu}{2}+1\Big)\frac{\nu}{2}\cdot 2^2 = 2\Big(\frac{\nu}{2}+1\Big)\nu \\
&= (\nu+2)\nu = \nu^2+2\nu
\end{aligned}
$$

1. $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$

$$\begin{aligned}\mu &= E(X)\\ &= \nu\end{aligned}$$

$$\begin{aligned}\sigma^2 &= E(X^2) - E(X)^2\\ &= \nu^2 + 2\nu - \nu^2\\ &= 2\nu\end{aligned}$$

9.4 Moment Generating Function

$$\begin{aligned}
M_X(t) &= E(e^{tX})\\
&= \int\limits_0^\infty e^{tx}\frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty e^{tx}\cdot x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{tx}e^{-\frac{x}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{tx-\frac{x}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{\frac{2tx}{2}-\frac{x}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{-\frac{2tx-x}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{-x\frac{-2t+1}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{-x\frac{1-2t}{2}}dx\\
&= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int\limits_0^\infty x^{\frac{\nu}{2}-1}e^{\frac{\frac{-x}{2}}{1-2t}}dx\\
{}^{[1]} &= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\Big[\Big(\frac{2}{1-2t}\Big)^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\Big]\\
&= \frac{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})(1-2t)^{\frac{\nu}{2}}}\\
&= \frac{1}{(1-2t)^{\frac{\nu}{2}}}\\
&= (1-2t)^{-\frac{\nu}{2}}
\end{aligned}$$

1. $\int\limits_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$

$$\begin{aligned}
M_X^{(1)}(t) &= -\frac{\nu}{2}(1-2t)^{-\frac{\nu}{2}-1}(-2) \\
&= \frac{2\nu}{2}(1-2t)^{-\frac{\nu}{2}-1} \\
&= \nu(1-2t)^{-\frac{\nu}{2}-1}
\end{aligned}$$

$$\begin{aligned}
M_X^{(2)}(t) &= (-\frac{\nu}{2}-1)\nu(1-2t)^{-\frac{\nu}{2}-2}(-2) \\
&= (\frac{2\nu}{2}+2)\nu(1-2t)^{-\frac{\nu}{2}-2} \\
&= (\nu+2)\nu)(1-2t)^{-\frac{\nu}{2}-2} \\
&= (\nu^2+2\nu)(1-2t)^{-\frac{\nu}{2}-2}
\end{aligned}$$

$$\begin{aligned}
M_X^{(1)}(0) &= \nu(1-2\cdot 0)^{-\frac{\nu}{2}-1} \\
&= \nu(1-0)^{-\frac{\nu}{2}-1} \\
&= \nu(1)^{-\frac{\nu}{2}-1} \\
&= \nu \\
M_X^{(2)}(0) &= (\nu^2+2\nu)(1-2\cdot 0)^{-\frac{\nu}{2}-2} \\
&= (\nu^2+2\nu)(1-0)^{-\frac{\nu}{2}-2} \\
&= (\nu^2+2\nu)(1)^{-\frac{\nu}{2}-2} \\
&= (\nu^2+2\nu)
\end{aligned}$$

$$\begin{aligned}
E(X) &= M_X^{(1)}(0) \\
&= \nu
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= M_X^{(2)}(0) \\
&= (\nu^2+2\nu)
\end{aligned}$$

$$\begin{aligned}
\mu &= E(X) \\
&= \nu \\
\sigma^2 &= E(X^2) - E(X)^2 \\
&= \nu^2 + 2\nu - \nu^2 \\
&= 2\nu
\end{aligned}$$

9.5 Maximum Likelihood Function

Let $x_1, x_2, \ldots, x_n$ be a random sample from a Chi-square distribution with parameter ν.

9.5.1 Likelihood Function

$$\begin{aligned}
L(\theta) &= f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta) \\
&= \frac{x_1^{\nu/2-1}e^{-x_1/2}}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{x_2^{\nu/2-1}e^{-x_2/2}}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} \cdots \frac{x_n^{\nu/2-1}e^{-x_n/2}}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} \\
&= \prod_{i=1}^{n} \frac{x_i^{\nu/2-1}e^{-x_i/2}}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} \\
&= \left(2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right) \prod_{i=1}^{n} x_i^{\nu/2-1}e^{-x_i/2} \\
&= \left(2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right) \cdot \exp\left\{\sum_{i=1}^{n} \frac{x_i}{2}\right\} \cdot \prod_{i=1}^{n} x_i^{\nu/2-1} \\
&= \left(2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right) \cdot \exp\left\{\frac{1}{2}\sum_{i=1}^{n} x_i\right\} \cdot \prod_{i=1}^{n} x_i^{\nu/2-1}
\end{aligned}$$

9.5.2 Log-likelihood Function

$$\begin{aligned}
\ell(\theta) &= \ln\big(L(\theta)\big) \\
&= \ln\left[\left(2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right) \cdot \exp\left\{\frac{1}{2}\sum_{i=1}^{n} x_i\right\} \cdot \prod_{i=1}^{n} x_i^{\nu/2-1}\right] \\
&= \ln\left[\left(2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right)\right] + \ln\left(\exp\left\{\frac{1}{2}\sum_{i=1}^{n} x_i\right\}\right) + \ln\left(\prod_{i=1}^{n} x_i^{\nu/2-1}\right) \\
&= -n\ln\left(2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{1}{2}\sum_{i=1}^{n} x_i + \left(\frac{\nu}{2}-1\right)\ln\left(\prod_{i=1}^{n} x_i\right) \\
&= -n\left(\ln(2^{\nu/2}) + \Gamma\left(\frac{\nu}{2}\right)\right) + \frac{1}{2}\sum_{i=1}^{n} x_i + \left(\frac{\nu}{2}-1\right)\sum_{i=1}^{n} \ln x_i \\
&= -n\left(\frac{\nu}{2}\ln 2 + \ln\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{1}{2}\sum_{i=1}^{n} x_i + \left(\frac{\nu}{2}-1\right)\sum_{i=1}^{n} \ln x_i \\
&= -\frac{n\nu}{2}\ln 2 - n\ln\Gamma\left(\frac{\nu}{2}\right) + \frac{1}{2}\sum_{i=1}^{n} x_i + \left(\frac{\nu}{2}-1\right)\sum_{i=1}^{n} \ln x_i
\end{aligned}$$

9.5.3 MLE for ν

$$\begin{aligned}
\frac{d\ell}{d\nu} &= -\frac{n}{2}\ln 2 - \frac{n}{\Gamma\left(\frac{\nu}{2}\right)}\Gamma'\left(\frac{\nu}{2}\right)\cdot\frac{1}{2} + 0 + \frac{1}{2}\sum_{i=1}^{n}\ln x_i \\
&= -\frac{n}{2}\ln 2 - \frac{n}{2\Gamma\left(\frac{\nu}{2}\right)}\Gamma'\left(\frac{\nu}{2}\right) + \frac{1}{2}\sum_{i=1}^{n}\ln x_i
\end{aligned}$$

$$\begin{aligned}
0 &= -\frac{n}{2}\ln 2 - \frac{n}{2\Gamma\left(\frac{\nu}{2}\right)}\Gamma'\left(\frac{\nu}{2}\right) + \frac{1}{2}\sum_{i=1}^{n}\ln x_i \\
\Rightarrow \frac{n}{2}\ln 2 - \frac{1}{2}\sum_{i=1}^{n}\ln x_i &= -\frac{n}{2\Gamma\left(\frac{\nu}{2}\right)}\Gamma'\left(\frac{\nu}{2}\right) \\
\Rightarrow n\ln 2 - \sum_{i=1}^{n}\ln x_i &= -\frac{n}{\Gamma\left(\frac{\nu}{2}\right)}\Gamma'\left(\frac{\nu}{2}\right) \\
\Rightarrow \frac{\sum_{i=1}^{n}\ln x_i - n\ln 2}{n} &= \frac{\Gamma'\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}
\end{aligned}$$

Due to the complexity of the Gamma function in this equation, no solution can be developed for ν in closed form. Thus, we have to rely on numerical methods to obtain a solution to the equation and find the maximum likelihood estimator.

9.6 Theorems for the Chi-Square Distribution

9.6.1 Validity of the Distribution

$$\int_0^\infty \frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}dx = 1$$

Proof:

$$\begin{aligned}
\int_0^\infty \frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}dx &= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}dx \\
{}^{[1]} &= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\left[2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\right] \\
&= \frac{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \\
&= 1
\end{aligned}$$

1. $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$

9.6.2 Sum of Chi-Square Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent Chi-Square random variables with parameter ν_i, that is $X_i \sim \chi^2(\nu_i),\ i = 1, 2, \ldots, n$.

Suppose $Y = \sum_{i=1}^{n} X_i$. Then $Y \sim \chi^2(\sum_{i=1}^{n} \nu_i)$.

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY} = E(e^{t(X_1+X_2+\cdots+X_n})\\
&= E(e^{tX_1}e^{tX_2}\cdots e^{tX_n})\\
&= E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n})\\
&= (1-2t)^{-\frac{\nu_1}{2}}(1-2t)^{-\frac{\nu_2}{2}}\cdots(1-2t)^{-\frac{\nu_n}{2}}\\
&= (1-2t)^{\sum_{i=1}^{n}\nu_i}
\end{aligned}$$

Which is the mgf of a Chi-Square random variable with parameter $\sum_{i=1}^{n} \nu_i$.
Thus $Y \sim \chi^2\left(\sum_{i=1}^{n} \nu_i\right)$.

9.6.3 Multiple of a Chi-Square Random Variable

Let X be a Chi-Square random variable with parameter ν, that is $X \sim \chi^2(\nu),\ i = 1, 2, \ldots, n$.

Suppose $Y = c \cdot X$. Then $Y \sim Gamma(\frac{\nu}{2}, 2 \cdot c)$.

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) = E(e^{tcX})\\
&= (1-2tc)^{-\frac{\nu}{2}}\\
&= (1-2c\cdot t)^{-\frac{\nu}{2}}
\end{aligned}$$

Which is the mgf of a Gamma distributed variable with parameters $\alpha = \frac{\nu}{2}$ and $\beta = 2c$. Thus, $Y \sim Gamma(\frac{\nu}{2}, 2 \cdot c)$.

9.6.4 Square of a Standard Normal Random Variable

If $Z \sim N(0, 1)$, then $Z^2 \sim \chi^2(1)$.

Proof:

$$
\begin{aligned}
M_{Z^2}(t) &= E(e^{tZ^2}) \\
&= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz^2} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(-2t+1)} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(1-2t)} dz \\
[1] &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{z^2}{2}(1-2t)} dz \\
[2] &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-u} \frac{\sqrt{2}u^{-\frac{1}{2}}}{2(1-2t)^{\frac{1}{2}}} du \\
&= \frac{2\sqrt{2}}{2\sqrt{2\pi}(1-2t)^{\frac{1}{2}}} \int_{0}^{\infty} e^{-u} u^{-\frac{1}{2}} du \\
&= \frac{2\sqrt{2}}{2\sqrt{2\pi}(1-2t)^{\frac{1}{2}}} \int_{0}^{\infty} u^{\frac{1}{2}-1} e^{-u} du \\
[3] &= \frac{1}{\sqrt{\pi}(1-2t)^{\frac{1}{2}}} \Gamma(\frac{1}{2}) \\
&= \frac{\sqrt{\pi}}{\sqrt{\pi}(1-2t)^{\frac{1}{2}}} \\
&= \frac{1}{(1-2t)^{\frac{1}{2}}} = (1-2t)^{-\frac{1}{2}}
\end{aligned}
$$

1. $\int_{-\infty}^{\infty} f(x)dx = 2\int_{0}^{\infty} f(x)dx$ when f(x) is an even function (**??**)
2. Let $u = \frac{z^2}{2}(1-2t) \quad \Rightarrow z = \frac{\sqrt{2}u^{\frac{1}{2}}}{(1-2t)^{\frac{1}{2}}}$

   So $dz = \frac{\sqrt{2}u^{-\frac{1}{2}}}{2(1-2t)^{\frac{1}{2}}}$
3. $\int_{0}^{\infty} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^{\alpha}\Gamma(\alpha)$

Which is the mgf of a Chi-Square random variable with 1 degree of freedom. Thus $Z^2 \sim \chi^2(1)$.

Chapter 10

Combinations

10.0.1 Lemma

A set of n elements may be partitioned into m distinct groups containing $k_1, k_2, \ldots, k_m$ objects, respectively, where each object appears in exactly one group and $\sum_{i=1}^{m} k_i = n$, in $N = \binom{n}{k_1 k_2 \ldots k_m} = \frac{n!}{k_1! k_2! \ldots k_m!}$ ways.\

Proof:

N is the number of ways all n of the elements of the set can be arranged in m groups where the order within each group is not important (i.e. rearrangements of elements in a group do not qualify as distinct groups).

The number of distinct arrangements of the n elements in which the order of selection is important, P_k^n, is equal to N multiplied by the number of ways each individual group of k_i can be selected in which the order is important, i.e.

$$\begin{aligned} P_n^n &= N \cdot P_{k_1}^{k_1} P_{k_2}^{k_2} \cdots P_{k_m}^{k_m} \\ \Rightarrow n! &= N \cdot k_1! k_2! \cdots k_m! \\ \Rightarrow N &= \frac{n!}{k_1! k_2! \cdots k_m!} \end{aligned}$$

10.0.2 Combinations Theorem

Given a set of n elements, the number of possible ways to select a subset of size k, without regard to the order of their selection, is $\frac{n!}{k!(n-k)!}$.\

Proof:

This theorem is a special case of the Lemma with $n = n$, $m = 2$, $k_1 = k$ and $k_2 = n - k$. thus,

$$N = \frac{n!}{k!(n-k)!}$$

The formula $\dfrac{n!}{k!(n-k)!}$ is denoted in a number of ways, depending on the author. Denotations may be C_k^n, ${}_nC_k$, $C_{n,k}$, $C(n,k)$, and $\binom{n}{k}$.
Throughout this book, the form $\binom{n}{k}$ is used and may be read "n choose k objects."

10.0.3 Theorem

For any integer a such that $0 \leq a \leq k$,

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-a+1)}{k(k-1)(k-2)\cdots(k-a+1)}\binom{n-a}{k-a}$$

Proof:

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n(n-1)!}{k(k-1)!(n-k)!} \\
&= \frac{n(n-1)(n-2)!}{k(k-1)(k-2)!(n-k)!} \\
&= \frac{n(n-1)(n-2)\cdots(n-a+1)(n-a)!}{k(k-1)(k-2)\cdots(k-a+1)(k-a)!(n-k)!} \\
&= \frac{n(n-1)(n-2)\cdots(n-a+1)}{k(k-1)(k-2)\cdots(k-a+1)} \cdot \frac{(n-a)!}{(k-a)!(n-k)!} \\
&= \frac{n(n-1)(n-2)\cdots(n-a+1)}{k(k-1)(k-2)\cdots(k-a+1)} \cdot \frac{(n-a)!}{(k-a)!(n-a+a-k)!} \\
&= \frac{n(n-1)(n-2)\cdots(n-a+1)}{k(k-1)(k-2)\cdots(k-a+1)} \cdot \frac{(n-a)!}{(k-a)![(n-a)+(a-k)]!} \\
&= \frac{n(n-1)(n-2)\cdots(n-a+1)}{k(k-1)(k-2)\cdots(k-a+1)} \cdot \frac{(n-a)!}{(k-a)![(n-a)-(k-a)]!} \\
&= \frac{n(n-1)(n-2)\cdots(n-a+1)}{k(k-1)(k-2)\cdots(k-a+1)} \cdot \binom{n-a}{k-a}
\end{aligned}$$

Chapter 11

Correlation (Pearson's)

Pearson's correlation coefficient of the variables X and Y is a measure of the linear relationship between X and Y. It is defined

$$\rho = \frac{Cov(X,Y)}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}}$$

Notice that if X and Y are independent then $Cov(X,Y,) = 0$ and $\rho = 0$ and there is no linear relationship between the variables.

11.1 Theorems on Pearson's Correlation

11.2 Computational Formula for ρ

$$\rho = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}(x_i - \mu_X)(y_j - \mu_Y)}{\sum\limits_{i=1}^{n}(x_i - \mu_X)\sum\limits_{j=1}^{m}(y_i - \mu_Y)}$$

Proof:

$$
\begin{aligned}
\rho &= \frac{Cov(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} \\
&= \frac{Cov(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} \\
&= \frac{\sum\limits_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)\frac{1}{N}}{\sqrt{\frac{\sum\limits_{i=1}^{N}(x_i-\mu_X)^2}{N} \frac{\sum\limits_{i=1}^{N}(y_i-\mu_Y)^2}{N}}} \\
&= \frac{\frac{1}{N}\sum\limits_{i=1}^{N}(x_i - \mu_X)(y_j - \mu_Y)}{\frac{1}{N}\sqrt{\sum\limits_{i=1}^{N}(x_i - \mu_X)^2 \sum\limits_{i=1}^{N}(y_i - \mu_Y)^2}} \\
&= \frac{\sum\limits_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum\limits_{i=1}^{N}(x_i - \mu_X) \sum\limits_{i=1}^{N}(y_i - \mu_Y)}}
\end{aligned}
$$

Chapter 12

Covariance

12.1 Definition of Covariance

For any two random variables X and Y, the covariance of X and Y is defined as

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

12.2 Theorems on Covariance

12.2.1 Theorem

Let X be a random variable. Then

$$\text{Cov}(X, X) = V(X)$$

Proof:

$$\begin{aligned}
\text{Cov}(X, X) &= E[(X - \mu)(X - \mu)] \\
&= E[(X - \mu)^2] \\
&= V(X)
\end{aligned}$$

12.2.2 Theorem

Let X and Y be random variables. Then

$$\text{Cov(X, Y)} = E(XY) - E(X)E(Y)$$

Proof:

$$\begin{aligned}
\text{Cov(X, Y)} &= E[(X - \mu_x)(Y - \mu_Y)] \\
&= E[XY - X\mu_y - Y\mu_X + \mu_X\mu_Y] \\
&= E(XY) - E(X)\mu_Y - \mu_X E(Y) + \mu_X\mu_Y \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - 2E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

12.2.3 Covariance

Let X and Y be random variables and let a and b be constants. Then

$$\text{Cov(aX, bY)} = ab\text{Cov}(X, Y)$$

Proof:

$$\begin{aligned}
\text{Cov}(aX, bY) &= E(aXbY) - E(aX)E(bY) \\
&= abE(XY) - abE(X)E(Y) \\
&= ab[E(XY) - E(X)E(Y)] \\
&= ab \cdot \text{Cov}(X, Y)
\end{aligned}$$

12.2.4 Theorem

Let $X_1, X_2, \ldots, X_n$ be random variables with $E(X_i) = \mu_i$ for $i = 1, 2, \ldots, n$ and let $Y_1, Y_2, \ldots, Y_m$ be random variables with $E(Y_j) = \phi_j$ for $j = 1, 2, \ldots, m$. Also, let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$ be constants.\

If $U_1 = \sum_{i=1}^{n} a_i X_i$ and $U_2 = \sum_{i=1}^{m} b_i Y_i$, then

$$\text{Cov}(U_1, U_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \cdot \text{Cov}(X_i, Y_j)$$

Proof:

$$\begin{aligned}
\text{Cov}(U_1, U_2) &= E[(U_1 - E(U_1))(U_2 - E(U_2))] \\
&= E\Big[\big(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i\mu_i\big)\big(\sum_{j=1}^{m} b_j Y_j - \sum_{j=1}^{m} b_j\phi_j\big)\Big] \\
&= E\Big[\big(\sum_{i=1}^{n} a_i(X_i - \mu_i)\big)\big(\sum_{j=1}^{m} b_j(Y_j - \phi_j)\big)\Big] \\
&= E\Big[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j (X_i - \mu_i)(Y_j - \phi_j)\Big] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E[(X_i - \mu_i)(Y_j - \phi_j)] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \ \text{Cov}(X_i, Y_j)
\end{aligned}$$

12.2.5 Theorem

Let $X_1, X_2, \ldots, X_n$ be random variables with $E(X_i) = \mu_i$ for $i = 1, 2, \ldots, n$ and let $a_1, a_2, \ldots, a_n$ be constants. If $Y = \sum\limits_{i=1}^{n} a_i X_i$ then

$$V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i<} \sum_{j} a_i a_j \ \mathrm{Cov}(X_i, X_j)$$

Proof:

$$\begin{aligned}
V(Y) &= E[(Y - \mu_Y)^2] \\
&= E[(Y - \mu_Y)(Y - \mu_Y)] \\
&= E\Big[\big(\sum_{i=1}^{n} a_i X_i - a_i \mu_i\big)\big(\sum_{n=1}^{n} a_j X_j - a_j \mu_j\big)\Big] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= \sum_{i=1}^{n} a_i^2 \ \mathrm{Cov}(X_i, X_i) + \sum_{i\neq} \sum_{j} a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i<} \sum_{j} a_i a_j \ \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

Chapter 13

Experimental Designs

13.1 Designs in Categorical Data Analysis

Studies in Categorical Data Analysis can be classified into several designs. These designs fall into the following two categories:

1. *Retrospective Design*: looks at and analyzes measurements that have already been taken.
2. *Prospective Design*: specifies the measurements to be collected at a future time.

13.1.1 Case Control Study

In case control studies, the marginal distribution of the *response variable* is fixed by the sampling design. In other words, researchers select particular numbers of each category of the response variable in order to ensure that enough of each case are included in the sample. The result is that the marginal distribution of the response is non-random.

Unfortunately, in order to calculate conditional probabilities, the marginal distribution of interest must be random. The difference of proportions for the response and the relative risk are both based on the marginal distribution of the response, and are both invalid procedures in case-control studies.

In taking the measurements, researchers idenitfy people who are already classified into the response variable, making the design retrospective.

13.1.2 Cross Sectional Study

13.1.3 Cohort Study

In Cohort Studies, subjects make their own choice about which group in the explanatory variable to join and researchers monitor the subjects with respect to a response variable over a period of time. Both the explanatory and response variables are random and only the total sample size is fixed by the researcher. Thus, conditional probabilities may be computed for both the predictor and response variables; differences in proportions may be estimated; and the relative risk is defined for the response variable.

Since subjects select the group in which they will be and a measurement of their response is taken later, cohort studies are prospective.

	conditional probability explanatory	conditional probability response	difference of proportions	Relative Risk	Odds Ratio
Case Control	xxx				xxx
Cross-Sectional	xxx	xxx	xxx	xxx	xxx
Cohort	xxx	xxx	xxx	xxx	xxx
Randomized		xxx	xxx	xxx	xxx

13.1.4 Randomized Study

In randomized Studies, the researcher randomly assigns subjects to the explanatory variable and then observes their response (making this a prospective study). The marginal distribution of the explanatory variable is therefore fixed, and conditional probabilities may not be computed.

The response variable, on the other hand, is random and conditional probabilites may be computed, as well as the difference of proportions and relative risk.

13.1.5 Summary of Designs

Chapter 14

Exponential Distribution

14.1 Probability Distribution Function

A random variable is said to have an Exponential Distribution with parameter β if its probability distribution function is

$$f(x) = \begin{cases} \frac{1}{\beta} e^{\frac{-x}{\beta}}, & 0 < x, \quad 0 < \beta \\ 0 & otherwise \end{cases}$$

14.2 Cumulative Distribution Function

$$\begin{aligned} F(x) &= \int_0^x \frac{1}{\beta} \exp\left\{\frac{-t}{\beta}\right\} dt \\ &= -\exp\left\{\frac{-t}{\beta}\right\}\Big|_0^x \\ &= -\exp\left\{\frac{-x}{\beta}\right\} - \left(-\exp\left\{\frac{-0}{\beta}\right\}\right) \\ &= \exp\left\{\frac{0}{\beta}\right\} - \exp\left\{\frac{-x}{\beta}\right\} \\ &= 1 - \exp\left\{\frac{-x}{\beta}\right\} \end{aligned}$$

And so the cumulative distribution function is given by

$$F(x) = \begin{cases} 1 - e^{\frac{-x}{\beta}}, & 0 < x, \ 0 < \beta \\ 0 & otherwise \end{cases}$$

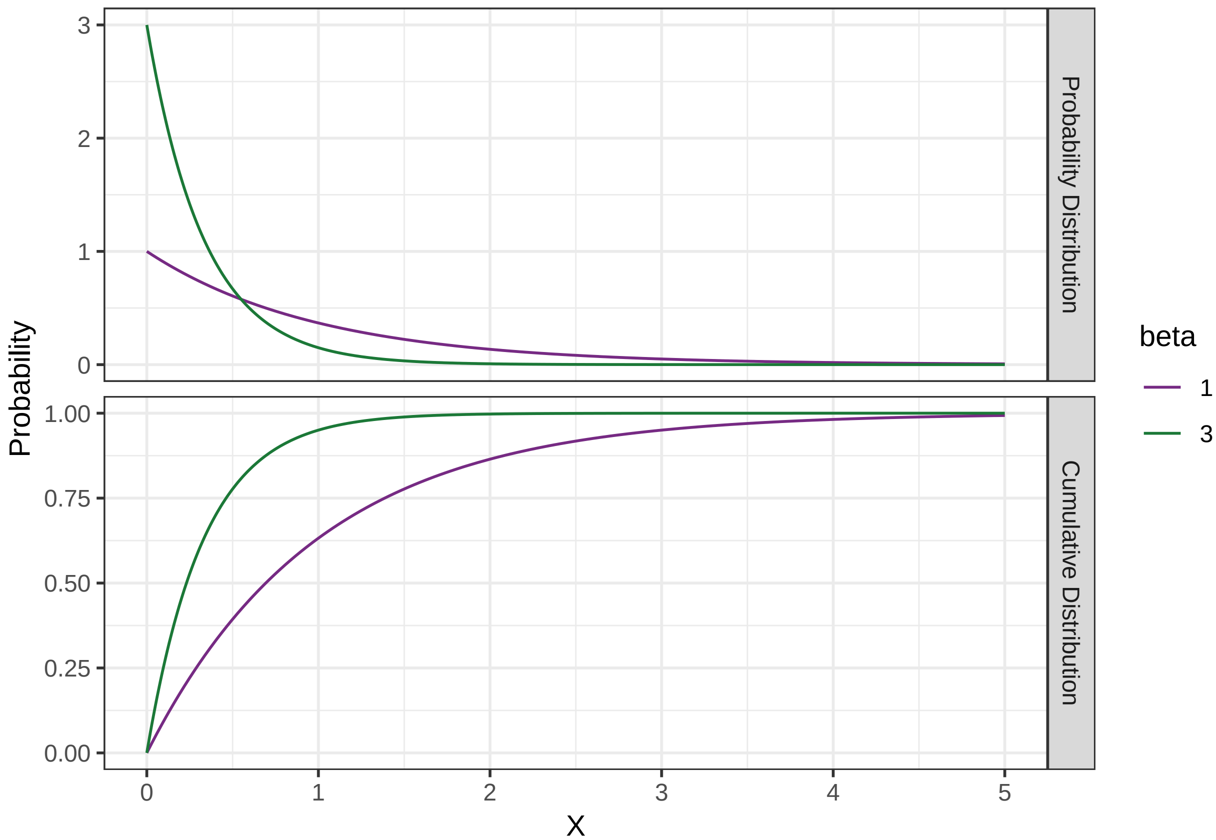

Figure 14.1: (#fig:Exponential_Distribution)The figures on the top and bottom display the Exponential probability and cumulative distirubtion functions, respectively, for $\beta = 1, 3$.

14.3 Expected Values

$$
\begin{aligned}
E(X) &= \int_0^\infty x f(x) dx \\
&= \int_0^\infty x \frac{1}{\beta} e^{\frac{-x}{\beta}} dx \\
&= \frac{1}{\beta} \int_0^\infty x e^{\frac{-x}{\beta}} dx \\
&= \frac{1}{\beta} \int_0^\infty x^{2-1} e^{\frac{-x}{\beta}} dx \\
{}^{[1]} &= \frac{1}{\beta}(\beta^2 \Gamma(2)) \\
&= \frac{\beta^2 \cdot 1!}{\beta} \\
&= \beta
\end{aligned}
$$

1. $\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^\alpha \Gamma(\alpha)$

$$
\begin{aligned}
E(X^2) &= \int_0^\infty x^2 f(x) dx \\
&= \int_0^\infty x^2 \frac{1}{\beta} e^{\frac{-x}{\beta}} dx \\
&= \frac{1}{\beta} \int_0^\infty x^2 e^{\frac{-x}{\beta}} dx \\
&= \frac{1}{\beta} \int_0^\infty x^{3-1} e^{\frac{-x}{\beta}} dx \\
{}^{[1]} &= \frac{1}{\beta}(\beta^3 \Gamma(3)) \\
&= \frac{\beta^3 \cdot 2!}{\beta} \\
&= 2\beta^2
\end{aligned}
$$

1. $\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^\alpha \Gamma(\alpha)$

$$\begin{aligned}\mu &= E(X)\\ &= \beta\end{aligned}$$

$$\begin{aligned}\sigma^2 &= E(X^2) - E(X)^2\\ &= 2\beta^2 - \beta^2\\ &= \beta^2\end{aligned}$$

14.4 Moment Generating Function

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_0^\infty e^{tx} \frac{1}{\beta} e^{\frac{-x}{\beta}} \, dx \\
&= \frac{1}{\beta} \int_0^\infty e^{tx} e^{\frac{-x}{\beta}} \, dx \\
&= \frac{1}{\beta} \int_0^\infty e^{tx - \frac{x}{\beta}} \, dx \\
&= \frac{1}{\beta} \int_0^\infty e^{\frac{\beta t x}{\beta} - \frac{x}{\beta}} \, dx \\
&= \frac{1}{\beta} \int_0^\infty e^{\frac{\beta t x - x}{\beta}} \, dx \\
&= \frac{1}{\beta} \int_0^\infty e^{\frac{-x(\beta 1 - \beta t}{\beta}} \, dx \\
&= \frac{1}{\beta} \left(\frac{-\beta}{1-\beta t}\right) e^{\frac{-x(1-\beta t}{\beta}} |_0^\infty \\
&= \frac{-1}{1-\beta t} e^{\frac{-x(1-\beta t}{\beta}} |_0^\infty \\
&= \frac{-1}{1-\beta t} \cdot 0 - \frac{-1}{1-\beta t} e^0 \\
&= \frac{1}{1-\beta t} = (1-\beta t)^{-1}
\end{aligned}$$

$$\begin{aligned}
M_X^{(1)}(t) &= -1(1-\beta t)^{-2}(-\beta) \\
&= \beta(1-\beta t)^{-2}
\end{aligned}$$

$$\begin{aligned}
M_X^{(2)}(t) &= -2\beta(1-\beta t)^{-3}(-\beta) \\
&= 2\beta^2(1-\beta t)^{-3}
\end{aligned}$$

$$\begin{aligned}
E(X) &= M_X^{(1)}(0) \\
&= \beta(1-\beta \cdot 0)^{-2} \\
&= \beta(1-0)^{-2} \\
&= \beta(1)^{-2} \\
&= \beta
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= M_X^{(2)}(0) \\
&= 2\beta^2(1-\beta \cdot 0)^{-3} \\
&= 2\beta^2(1-0)^{-3} \\
&= 2\beta^2(1)^{-3}
\end{aligned}$$

14.5 Maximum Likelihood Estimator

Let $x_1, x_2, \ldots, x_n$ be a random sample from an Exponential distribution with parameter β.

14.5.1 Likelihood Function

$$\begin{aligned} L(\theta) &= L(x_1, x_2, \ldots, x_n|\theta) \\ &= f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta) \\ &= \frac{1}{\theta}\exp\left\{-\frac{x_1}{\theta}\right\}\cdot\frac{1}{\theta}\exp\left\{-\frac{x_n}{\theta}\right\}\cdots\frac{1}{\theta}\exp\left\{-\frac{x_n}{\theta}\right\} \\ &= \frac{1}{\theta^n}\exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right\} \end{aligned}$$

14.5.2 Log-likelihood Function

$$\begin{aligned} \ell(\theta) &= \ln(L(\theta)) \\ &= \ln(1) - n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} x_i \\ &= 0 - n\ln(\theta) - \theta^{-1}\sum_{i=1}^{n} x_i \\ &= -n\ln(\theta) - \theta^{-1}\sum_{i=1}^{n} x_i \end{aligned}$$

14.5.3 MLE for β

$$\frac{d\ell(\beta)}{d\beta} = -\frac{n}{\beta} + \beta^2\sum_{i=1}^{n} x_i$$

$$\begin{aligned} 0 &= -\frac{n}{\beta} + \beta^2\sum_{i=1}^{n} x_i \\ \Rightarrow \frac{n}{\beta} &= \beta^2\sum_{i=1}^{n} x_i \\ \Rightarrow \frac{n\beta^2}{\beta} &= \sum_{i=1}^{n} x_i \\ \Rightarrow n\beta &= \sum_{i=1}^{n} x_i \\ \Rightarrow \beta &= \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned}$$

So $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the maximum likelihood estimator for β.

14.6 Theorems for the Exponential Distribution

14.6.1 Validity of the Distribution

$$\int_0^\infty \frac{1}{\beta} e^{\frac{-x}{B}} dx = 1$$

Proof:

$$\begin{aligned}
\int_0^\infty \frac{1}{\beta} e^{\frac{-x}{\beta}} dx &= -e^{\frac{-x}{\beta}} \Big|_0^\infty \\
&= -e^{\frac{-\infty}{\beta}} - (-e^{\frac{-0}{\beta}}) \\
&= e^{\frac{0}{\beta}} - e^{\frac{-\infty}{\beta}} \\
&= 1 - 0 \\
&= 1
\end{aligned}$$

14.6.2 Sum of Exponential Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent random variables from an Exponential distribution with parameter β, i.e. $X_i \sim$Exponential(β). Let $Y = \sum\limits_{i=1}^{n} X_i$. Then $Y \sim$Gamma(n, β).

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{t(X_1+X_2+\cdots+X_n)}) \\
&= E(e^{tX_1} e^{tX_2} \cdots e^{tX_n}) \\
&= (1-\beta t)^{-1} (1-\beta t)^{-1} \cdots (1-\beta t)^{-1} \\
&= (1-\beta t)^{-n}
\end{aligned}$$

Which is the mgf for a Gamma random variable with parameters n and β. Thus $Y \sim$Gamma(n, β).

Chapter 15

Functions

15.1 Fundamental Concepts and Definitions

Much of this chapter is taken from the lectures of Dr. John Brunette, University of Southern Maine

A **function** is a collection of ordered pairs in which no two pairs have the same first element.

The set of all *first* members of the pairs is called the **domain**.

The set of all *second* members of the pairs is called the **range**.

Suppose now that for any function f we have two items x and y such that $x \in dom(f)$ and $y \in ran(x)$ where $dom(f)$ and $ran(f)$ denote the domain and range of f, respectively. It is said that f maps x onto y, written

$$f:\ x \mapsto y$$

It is common to write the $ran(f)$ as some expression of x. For example, $f : x \mapsto x^2$ takes each element in the domain, and pairs it with it's square. The common shorthand for this is $f(x) = x^2$, meaning that whatever appears between the parentheses following the f is to be squared.

15.1.1 Function Operations

The three basic operations that can be performed on functions are addition, multipilication, and composition. For any two functions f and g these operations are defined as:

Addition	$[f+g](x) =: \big\{\big(x, f(x)+g(x)\big) \mid x \in dom(f) \cap dom(g)\big\}$
Multiplication	$[f \cdot g](x) := \big\{\big(x, f(x) \cdot g(x)\big) \mid x \in dom(f) \cap dom(g)\big\}$
Composition	$[f \circ g](x) = \big\{\big(x, f\big(g(x)\big) \mid x \in dom(g) \text{ and } g(x) \in dom(f)\big\}$

Notice that the composition $[f \circ g](x) = f \circ g : g(x) \mapsto f(x)$. In other words, the result of g is then applied to f to produce the result of the composition.

15.2 Identities and Inverses

Recall that addition and multiplication have identity properties. Specifically, for any real number x, applying one of these identities returns the value x, i.e. $x + 0 = x$ and $x \cdot 1$=x. Functions also have an identity, denoted $id(x)$, that is defined as

$$id:\ x \mapsto x$$

Furthermore, the composition of id with f behaves in this way:

$$id \circ f = f \circ id = f$$

Functions may also exhibit the property of inverses that are exhibited by addition and multiplication. In the latter two, combining any real number x and its inverse returns the identity of that operation, i.e. $x + -x = 0$ and $x \cdot x^{-1} = 1,\ x \neq 0$. Likewise, some functions have an inverse function. If a function f has an inverse f^{-1}, then

$$f \circ f^{-1} = f^{-1} \circ f = id$$

On closer observation, we see

$$f^{-1} \circ f\big(dom(x)\big) = f^{-1}\Big(f\big(dom(x)\big)\Big) = f^{-1}\big(ran(x)\big) = dom(x)$$

So f^{-1} must be the set of all ordered pairs (y, x) where $x \in dom(x)$ and $y \in ran(x)$, i.e. $f^{-1}(x) = \{(y, x) \mid x \in dom(x) \text{ and } y \in ran(x)\}$. By the definition of functions, no two first elements in f^{-1} can be the same. But the first elements in f^{-1} are the second elements in f. So f^{-1} only exists if no two second elements in f are the same. We thus make the following definition:

A function f is called a **one-to-one** function if it has no two ordered pairs with the same second element.

For any one-to-one function f, no two of the first elements are the same, and no two of the second elements are the same. Thus, f^{-1} is a function, because no two of its first elements are the same, and because the range of f^{-1} is the domain of f, no two second elements in f^{-1} are the same, and f^{-1} is a one-to-one function. Thus, every one-to-one function has an inverse.

If a function f is not one-to-one, however, then there exist two pairs in f that have the same second element. The inverse f^{-1} therefore has two pairs where the first element is the same. When such is the case, the definition of a function is violated, and f^{-1} cannot be a function. Thus, if a function is invertible, it must be one-to-one.

15.3 Odd and Even Functions

A function is said to be *even* if for any real number $x,\ f(-x) = f(x)$.

A function is said to be *odd* if for any real number $x,\ f(-x) = -f(x)$.

If neither of these criteria are met, the function is simply said to be neither odd nor even.

15.4 Theorems

15.4.1 Operations on Even Functions

Let f and g both be even functions. Then:

i. $[f + g](x)$ is an even function
ii. $[f \cdot g](x)$ is an even function
iii. $[f \circ g](x)$ is an even function.

Proof:

i.
$$\begin{aligned}(-x) &= f(-x) + g(-x)\\ &= f(x) + g(x)\\ &= [f+g](x)\end{aligned}$$

so $[f+g](x)$ is an even function.

ii.
$$\begin{aligned}(-x) &= f(-x) \cdot g(-x)\\ &= f(x) \cdot g(x)\\ &= [f \cdot g](x)\end{aligned}$$

so $[f \cdot g](x)$ is an even function.

iii.
$$\begin{aligned}(-x) &= f\big(g(-x)\big)\\ &= f\big(g(x)\big)\\ &= [f \circ g](x)\end{aligned}$$

so $[f \circ g](x)$ is an even function.

15.4.2 Operations on Odd Functions

Let f and g both be odd functions. Then:

i. $[f+g](x)$ is an odd function
ii. $[f \cdot g](x)$ is an even function
iii. $[f \circ g](x)$ is an odd function.

Proof:

i.
$$\begin{aligned}(-x) &= f(-x) + g(-x)\\ &= -f(x) - g(x)\\ &= -[f+g](x)\end{aligned}$$

so $[f+g](x)$ is an odd function.

ii.
$$\begin{aligned}(-x) &= f(-x) \cdot g(-x)\\ &= -f(x) \cdot -g(x)\\ &= f(x) \cdot g(x)\\ &= [f \cdot g](x)\end{aligned}$$

so $[f \cdot g](x)$ is an even function.

iii.
$$\begin{aligned}(-x) &= f\big(g(-x)\big)\\ &= f\big(-g(x)\big)\\ &= -f\big(g(x)\big)\\ &= -[f \circ g](x)\end{aligned}$$

so $[f \circ g](x)$ is an odd function.

15.4.3 Operations on an Odd and Even Function

Let f be an even function and let g both be an odd function. Then:

i. $[f+g](x)$ is neither an odd nor an even function
ii. $[f \cdot g](x)$ is an odd function
iii. $[f \circ g](x)$ is an even function
iv. $[g \circ f](x)$ is an even function.

Proof:

i.

$$\begin{aligned}(-x) &= f(-x) + g(-x) \\ &= -f(x) - g(x)\end{aligned}$$

so $[f+g](x)$ is neither an odd nor an even function.

ii.

$$\begin{aligned}(-x) &= f(-x) \cdot g(-x) \\ &= f(x) \cdot -g(x) \\ &= -\big(f(x) \cdot g(x)\big) \\ &= -[f \cdot g](x)\end{aligned}$$

so $[f \cdot g](x)$ is an odd function.

iii.

$$\begin{aligned}(-x) &= f\big(g(-x)\big) \\ &= f\big(-g(x)\big) \\ &= f\big(g(x)\big) \\ &= [f \circ g](x)\end{aligned}$$

so $[f \circ g](x)$ is an even function.

iv.

$$\begin{aligned}(-x) &= g\big(f(-x)\big) \\ &= g\big(f(x)\big) \\ &= [g \circ f](x)\end{aligned}$$

so $[f \circ g](x)$ is an even function. \end{itemize}

15.4.4 Derivatives and Anti-derivatives of Odd Functions

Let f be an odd function and let f' and F denote the derivative and anti-derivative of f, respectively. Then f' and F are both even functions.

Proof:

$$\begin{aligned}f(-x) &= -f(x) \\ \Rightarrow \frac{d}{dx}\big[f(-x)\big] &= \frac{d}{dx}\big[-f(x)\big] \\ \Rightarrow f'(-x) \cdot -1 &= -f'(x) \\ \Rightarrow -f'(-x) &= -f'(x) \\ \Rightarrow f'(-x) &= f'(x)\end{aligned}$$

So f' is an even function.

$$\begin{aligned}
f(-x) &= -f(x) \\
\Rightarrow \int f(-x) &= \int -f(x) \\
\Rightarrow F(-x) \cdot -1 &= -F(x) \\
\Rightarrow -F(-x) &= -F(x) \\
\Rightarrow F(-x) &= F(x)
\end{aligned}$$

So F is also an even function. ∎

15.4.5 Derivatives and Anti-derivatives of Even Functions

Let g be an even function, and let g' and G denote the derivative and anti-derivative of g, respectively. Then g' and G are both odd functions.

Proof:

$$\begin{aligned}
g(-x) &= g(x) \\
\Rightarrow \frac{d}{dx}\big[g(-x)\big] &= \frac{d}{dx}\big[g(x)\big] \\
\Rightarrow g'(-x) \cdot -1 &= g'(x) \\
\Rightarrow -g'(-x) &= g'(x) \\
\Rightarrow g'(-x) &= -g'(x)
\end{aligned}$$

So g' is an odd function.

$$\begin{aligned}
g(-x) &= g(x) \\
\Rightarrow \int g(-x) &= \int g(x) \\
\Rightarrow G(-x) \cdot -1 &= G(x) \\
\Rightarrow -G(-x) &= G(x) \\
\Rightarrow G(-x) &= -G(x)
\end{aligned}$$

So G is also an odd function.

Chapter 16

Gamma Distribution

16.1 Probability Distribution Function

A random variable X is said to have a Gamma Distribution with parameters α and β if its probability distribution function is

$$f(x) = \begin{cases} \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, & 0 < x,\ 0 < \alpha,\ 0 < \beta \\ 0 & otherwise \end{cases}$$

Where α is a scale parameter and\ β is a shape parameter.

16.2 Cumulative Distribution Function

The cumulative distribution function for the Gamma Distribution cannot be expressed in closed form. It's interval form is expressed here.

$$F(x) = \begin{cases} \int\limits_0^x \frac{t^{\alpha-1}e^{-\frac{t}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, & 0 < t,\ 0 < \alpha,\ 0 < \beta \\ \\ 0 & otherwise \end{cases}$$

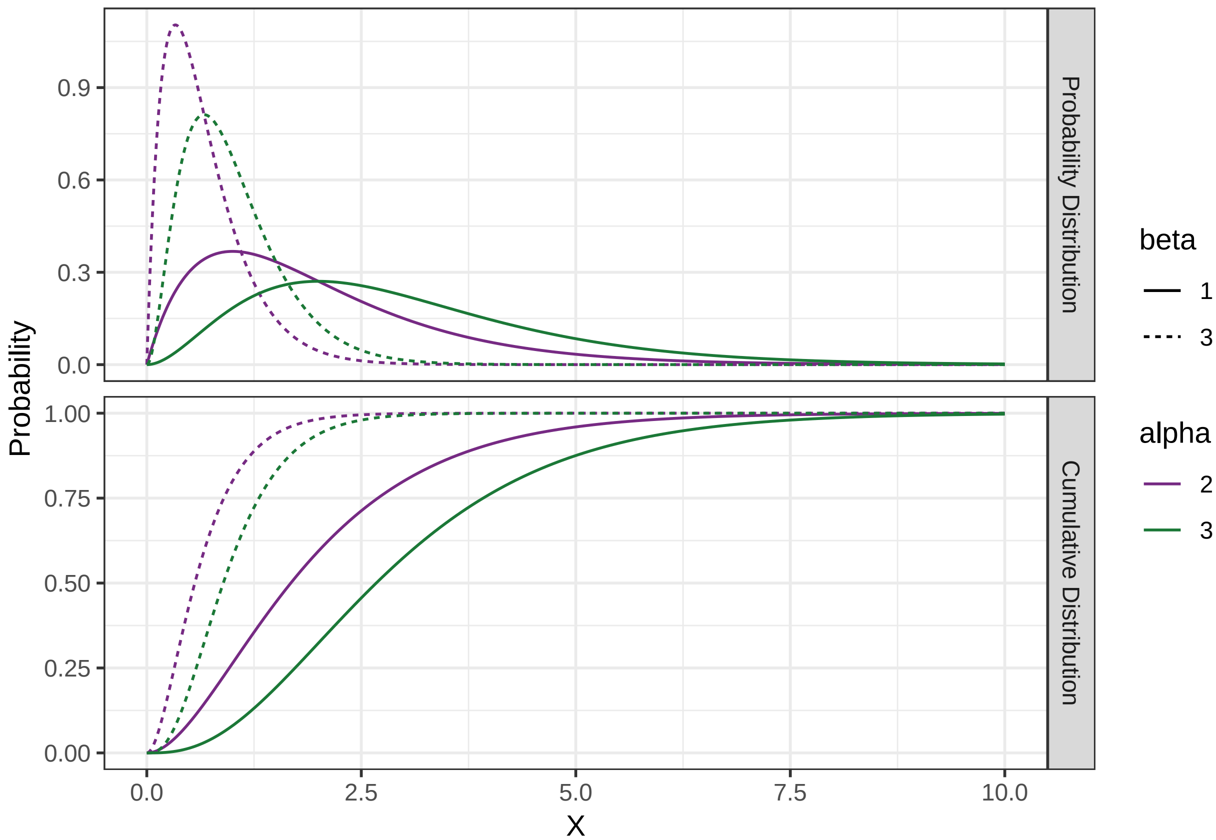

Figure 16.1: (#fig:Gamma_Distribution)The figures on the left and right display the Gamma probability and cumulative distirubtion functions, respectively, for the combinations of $\alpha = 2, 3$ and $\beta = 1, 3$.

16.3 Expected Values

$$\begin{aligned}
E(X) &= \int_0^\infty x \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}\,dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x\cdot x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha}e^{-\frac{x}{\beta}}\,dx \\
\overset{[1]}{}&= \frac{1}{\Gamma(\alpha)\beta^\alpha}[\Gamma(\alpha+1)\beta^{\alpha+1}] \\
&= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha} \\
&= \frac{\alpha\Gamma(\alpha)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha} \\
&= \alpha\beta
\end{aligned}$$

1. $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx = \beta^\alpha\Gamma(\alpha)$

$$\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}\,dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^2\cdot x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+1}e^{-\frac{x}{\beta}}\,dx \\
\overset{[1]}{}&= \frac{1}{\Gamma(\alpha)\beta^\alpha}[\Gamma(\alpha+2)\beta^{\alpha+2}] \\
&= \frac{\Gamma(\alpha+2)\beta^{\alpha+2}}{\Gamma(\alpha)\beta^\alpha} \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)\beta^{\alpha+2}}{\Gamma(\alpha)\beta^\alpha} \\
&= \frac{(\alpha+1)\alpha\Gamma(\alpha)\beta^{\alpha+2}}{\Gamma(\alpha)\beta^\alpha} \\
&= \alpha(\alpha+1)\beta^2 \\
&= (\alpha^2+\alpha)\beta^2 \\
&= \alpha^2\beta^2+\alpha\beta^2
\end{aligned}$$

1. $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx = \beta^\alpha\Gamma(\alpha)$

$$\begin{aligned}\mu &= E(X) \\ &= \alpha\beta\end{aligned}$$

$$\begin{aligned}\sigma^2 &= E(X^2) - E(X)^2 \\ &= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 \\ &= \alpha\beta^2\end{aligned}$$

16.4 Moment Generating Function

$$\begin{aligned}M_X(t) &= E(e^{tX}) \\ &= \int_0^\infty e^{tx}\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{tx}x^{\alpha-1}e^{-\frac{x}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{tx}e^{-\frac{x}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{tx-\frac{x}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{\frac{\beta tx}{\beta}-\frac{x}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{\frac{\beta tx-x}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-x\frac{-\beta t+1}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-x\frac{1-\beta t}{\beta}}dx \\ {}^{[1]} &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\left[\Gamma(\alpha)\Big(\frac{\beta}{1-\beta t}\Big)\alpha)\right] \\ &= \frac{\Gamma(\alpha)\beta^\alpha}{\Gamma(\alpha)\beta^\alpha(1-\beta t)^\alpha} \\ &= \frac{1}{(1-\beta t)^\alpha} = (1-\beta t)^{-\alpha}\end{aligned}$$

1. $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$

$$\begin{aligned}
M_X^{(1)}(t) &= -\alpha(1-\beta t)^{-\alpha-1}(-\beta) \\
&= \alpha\beta(1-\beta t)^{-\alpha-1} \\
M_X^{(2)}(t) &= (-\alpha-1)\alpha\beta(1-\beta t)^{-\alpha-2}(-\beta) \\
&= (\alpha+1)\alpha\beta^2(1-\beta t)^{-\alpha-2} \\
&= (\alpha^2\beta^2+\alpha\beta^2)(1-\beta t)^{-\alpha-2}
\end{aligned}$$

$$\begin{aligned}
E(X) &= M_X^{(1)}(0) = \alpha\beta(1-\beta\cdot 0)^{-\alpha-1} \\
&= \alpha\beta(1-0)^{\alpha-1} = \alpha\beta(1)^{-\alpha-1} \\
&= \alpha\beta
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= M_X^{(2)}(0) = (\alpha^2\beta^2+\alpha\beta^2)(1-\beta 0)^{-\alpha-2} \\
&= (\alpha^2\beta^2+\alpha\beta^2)(1-0)^{-\alpha-2} \\
&= (\alpha^2\beta^2+\alpha\beta^2)(1)^{-\alpha-2} \\
&= \alpha^2\beta^2+\alpha\beta^2
\end{aligned}$$

$$\begin{aligned}
\mu &= E(X) \\
&= \alpha\beta
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= E(X^2)-E(X)^2 \\
&= \alpha^2\beta^2+\alpha\beta^2-\alpha^2\beta^2 \\
&= \alpha\beta^2
\end{aligned}$$

16.5 Maximum Likelihood Estimators

Let $x_1, x_2, \ldots, x_n$ denote a random sample from a Gamma Distribution with parameters α and β.

16.5.1 Likelihood Function

$$
\begin{aligned}
L(\theta) &= L(x_1, x_2, \ldots, x_n|\theta) \\
&= f(x_1|\theta) f(x_2|\theta) \cdots f(x_n|\theta) \\
&= \frac{x_1^{\alpha-1} e^{-x_1/\beta}}{\Gamma(\alpha)\beta^\alpha} \frac{x_2^{\alpha-1} e^{-x_2/\beta}}{\Gamma(\alpha)\beta^\alpha} \cdots \frac{x_n^{\alpha-1} e^{-x_n/\beta}}{\Gamma(\alpha)\beta^\alpha} \\
&= \prod_{i=1}^{n} \frac{x_i^{\alpha-1} e^{-x_i/\beta}}{\Gamma(\alpha)\beta^\alpha} \\
&= \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^n \prod_{i=1}^{n} x_i^{\alpha-1} e^{-x_i/\beta} \\
&= \big(\Gamma(\alpha)\beta^\alpha\big)^{-n} \prod_{i=1}^{n} x_i^{\alpha-1} e^{-x_i/\beta} \\
&= \big(\Gamma(\alpha)\beta^\alpha\big)^{-n} \exp\left\{ \sum_{i=1}^{n} -\frac{x_i}{\beta} \right\} \prod_{i=1}^{n} x_i^{\alpha-1} \\
&= \big(\Gamma(\alpha)\beta^\alpha\big)^{-n} \exp\left\{ -\frac{1}{\beta} \sum_{i=1}^{n} x_i \right\} \prod_{i=1}^{n} x_i^{\alpha-1}
\end{aligned}
$$

16.5.2 Log-likelihood Function

$$
\begin{aligned}
\ell(\theta) &= \ln\left[\big(\Gamma(\alpha)\beta^\alpha\big)^{-n} \exp\left\{ -\frac{1}{\beta} \sum_{i=1}^{n} x_i \right\} \prod_{i=1}^{n} x_i^{\alpha-1} \right] \\
&= \ln\big(\Gamma(\alpha)\beta^\alpha\big)^{-n} + \ln\left(\exp\left\{ -\frac{1}{\beta} \sum_{i=1}^{n} x_i \right\} \right) + \ln\left(\prod_{i=1}^{n} x_i^{\alpha-1} \right) \\
&= -n \ln\big(\Gamma(\alpha)\beta^\alpha\big) - \frac{1}{\beta} \sum_{i=1}^{n} x_i + \ln\left(\prod_{i=1}^{n} x_i^{\alpha-1} \right) \\
&= -n\big[\ln\big(\Gamma(\alpha)\beta^\alpha\big) \big] - \frac{1}{\beta} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} (\alpha-1) \ln x_i \\
&= -n \ln\Gamma(\alpha) - n\alpha \ln\beta - \frac{1}{\beta} \sum_{i=1}^{n} x_i + (\alpha-1) \sum_{i=1}^{n} \ln x_i
\end{aligned}
$$

16.5.3 MLE for α

$$\begin{aligned}\frac{d\ell}{d\alpha} &= -n\frac{1}{\Gamma(\alpha)}\Gamma'(\alpha) - n\ln\beta - 0 + \sum_{i=1}^{n}\ln x_i \\ &= -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\ln\beta + \sum_{i=1}^{n}\ln x_i\end{aligned}$$

$$\begin{aligned}0 &= -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\ln\beta + \sum_{i=1}^{n}\ln x_i \\ \Rightarrow n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} &= \sum_{i=1}^{n}\ln x_i - n\ln\beta \\ \Rightarrow \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} &= \frac{1}{n}\left(\sum_{i=1}^{n}\ln x_i - n\ln\beta\right)\end{aligned}$$

However, this must be solved numerically. Notice also that the MLE for α depends on β.

16.5.4 MLE for β

$$\begin{aligned}\frac{d\ell}{d\beta} &= 0 - n\alpha\frac{1}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n}x_i + 0 \\ &= -\frac{n\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n}x_i\end{aligned}$$

$$\begin{aligned}0 &= -\frac{n\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n}x_i \\ \Rightarrow \frac{n\alpha}{\beta} &= \frac{1}{\beta^2}\sum_{i=1}^{n}x_i \\ \Rightarrow n\alpha\beta &= \sum_{i=1}^{n}x_i \\ \Rightarrow \beta &= \frac{1}{n\alpha}\sum_{i=1}^{n}x_i\end{aligned}$$

This estimate, however, depends on α. Since each estimator depends on the value of the other parameter, we must maximize the likelihood functions simulatneously. That is, we must simultaneously solve the system

$$\begin{cases} -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\ln\beta + \sum\limits_{i=1}^{n}\ln x_i &= 0 \\ -\frac{n\alpha}{\beta} + \frac{1}{\beta^2}\sum\limits_{i=1}^{n}x_i &= 0\end{cases}$$

Solving this system will require numerical methods.

16.5.5 Approximation of $\hat{\alpha}$ and $\hat{\beta}$

Approximations of $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by noticing that\

$$\begin{aligned}\frac{d\ell}{d\beta} &= 0\\ \Rightarrow \beta &= \frac{1}{n\alpha}\sum_{i=1}^{n} x_i\\ \Rightarrow \alpha\beta &= \frac{1}{n}\sum_{i=1}^{n} x_i\end{aligned}$$

So $\widehat{\alpha\beta} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$. Recall that $\alpha\beta$ and $\alpha\beta^2$ are the mean and variance of the Gamma Distribution, respectively. We utilize

$$\frac{\alpha\beta^2}{\alpha\beta} = \beta$$

If we assume that $\widehat{\alpha\beta^2} = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$, then

$$\frac{\widehat{\alpha\beta^2}}{\widehat{\alpha\beta}} = \beta^* \approx \hat{\beta}$$

Where β^* denotes an approximation to $\hat{\beta}$

We now substitute β^* into

$$\begin{aligned}\widehat{\alpha\beta} &= \frac{1}{n}\sum_{i=1}^{n} x_i\\ \Rightarrow \alpha^*\beta^* &= \frac{1}{n}\sum_{i=1}^{n} x_i\\ \Rightarrow \alpha^* &= \frac{1}{n\beta^*}\sum_{i=1}^{n} x_i \approx \hat{\alpha}\end{aligned}$$

Where α^* denotes an approximation to $\hat{\alpha}$.

This method of estimation is prone to error because β^* is found through two levels of estimation and α^* is found through three levels of estimation. Surely, this process inflates the error of estimation. At this point, however, I have no information to indicate how badly the error of estimation is inflated, nor have I performed any investigation into this problem.

16.6 Theorems for the Gamma Distribution

16.6.1 Validity of the Distribution

$$\int_0^{\infty} x\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}dx = 1$$

Proof:

$$
\begin{aligned}
\int\limits_0^\infty \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}dx &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int\limits_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx \\
&\overset{[1]}{=} \frac{1}{\Gamma(\alpha)\beta^\alpha}[\Gamma(\alpha)\beta^\alpha] \\
&= \frac{\Gamma(\alpha)\beta^\alpha}{\Gamma(\alpha)\beta^\alpha} \\
&= 1
\end{aligned}
$$

1. $\int\limits_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$

16.6.2 Sum of Gamma Random Variables

Let $X_1, X_2, \ldots, X_n$ be Gamma distributed random variables with parameters α_i and β, that is $X_i \sim$Gamma(α_i, β). Let $Y = \sum\limits_{i=1}^n X_i$.\ Then $Y \sim$Gamma$(\sum\limits_{i=1}^n \alpha_i, \beta)$.

Proof:

$$
\begin{aligned}
M_Y(t) = E(e^{tY}) &= E(e^{t(X_1+X_2+\cdots+X_n)} \\
&= E(e^{tX_1}e^{tX_2}\cdots e^{tX_n}) \\
&= E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n}) \\
&= (1-\beta t)^{-\alpha_1}(1-\beta t)^{-\alpha_2}\cdots(1-\beta t)^{-\alpha_n} = (1-\beta t)^{-\sum\limits_{i=1}^n \alpha_i}
\end{aligned}
$$

Which is the moment generating function of a Gamma random variable with parameters $\sum\limits_{i=1}^n \alpha_i$ and β. Thus $Y \sim$Gamma$(\sum\limits_{i=1}^n \alpha_i, \beta)$.

16.6.3 Sum of Exponential Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent random variables from an Exponential distribution with parameter β, i.e. $X_i \sim$Exponential(β). Let $Y = \sum\limits_{i=1}^n X_i$. Then $Y \sim$Gamma(n, β).

Proof:

$$
\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{t(X_1+X_2+\cdots+X_n}) \\
&= E(e^{tX_1}e^{tX_2}\cdots e^{tX_n}) \\
&= (1-\beta t)^{-1}(1-\beta t)^{-1}\cdots(1-\beta t)^{-1} \\
&= (1-\beta t)^{-n}
\end{aligned}
$$

Which is the moment generating function for a Gamma random variable with parameters n and β. Thus $Y \sim$Gamma(n, β).

Chapter 17

Gamma Function

The Gamma Function is a function used frequently in statistical theory. It has properties that permit simplified calculations and is used in defining many probability distributions, particularly within the Exponential family of distributions.

17.1 Definition

$$\Gamma(x) = \int\limits_0^\infty t^{x-1} e^{-t} dt$$

Note that the definition defines a function of x that is integrated over t. Thus, for each value of x, a curve is defined, and the Gamma function calculates the area under the curve defined by x.

17.2 Theorems for the Gamma Function

17.2.1 Lemma

$$\left[-t^{x-1} e^{-t}\right]_{t=0}^{t=\infty} = 0$$

Proof:

$$
\begin{aligned}
\left[-t^{x-1}e^{-t}\right]_{t=0}^{t=\infty} &= \lim_{t\to\infty}\left(-t^{x-1}e^{-t}\right) - 0^{x-1}e^{-0} \\
&= \lim_{t\to\infty}\left(-t^{x-1}e^{-t}\right) - 0 \\
&= \lim_{t\to\infty}\left(-t^{x-1}e^{-t}\right) \\
&= -\lim_{t\to\infty}\frac{t^{x-1}}{e^{t}} \\
&= -\lim_{t\to\infty}\{exp\left[(x-1)\ln t - t\right]\} \\
&= -\lim_{t\to\infty}\left\{exp\left[(x-1)\cdot t\cdot\left(\frac{\ln t}{t}-1\right)\right]\right\} \\
{}^{[1]} &= \lim_{t\to\infty}\left\{exp\left[(x-1)\cdot t\cdot\Big(0-1\Big)\right]\right\} \\
&= \lim_{t\to\infty}\{exp\left[-(x-1)\cdot t\right]\} \\
&= \lim_{t\to\infty}\frac{1}{e^{(x-1)\cdot t}} \\
&= 0
\end{aligned}
$$

1. L'H^{o}pital's Rule: $\lim_{x\to u}\frac{f(x)}{g(x)} = \lim_{x\to u}\frac{f'(x)}{g'(x)}$. This implies $\lim_{x\to\infty}\frac{\ln x}{x} = \lim_{x\to\infty}\frac{\frac{1}{x}}{1} = \lim_{x\to\infty}\frac{1}{x} = 0$

17.2.2 Theorem: The Reduction Relation

$\Gamma(x) = (x-1)\cdot\Gamma(x)$

Proof:

The proof relies on integration by parts. Let:

$$
\begin{aligned}
u &= t^{x-1} \\
du &= (x-1)\cdot t^{(x-2)} \\
\\
v &= -e^{-t} \\
dv &= e^{-t}dt
\end{aligned}
$$

Integration by parts yields:

$$\begin{aligned}
\Gamma(x) &= \int_0^\infty t^{x-1}e^{-t}dt \\
&= u \cdot v - \int_0^\infty v \cdot du \\
&= \Big[t^{x-1} \cdot -e^{-t}\Big]_{t=0}^{t=\infty} - \int_0^\infty -e^{-t} \cdot (x-1) \cdot t^{(x-2)}dt \\
&= -\Big[t^{x-1} \cdot e^{-t}\Big]_{t=0}^{t=\infty} - (-(x-1)) \int_0^\infty e^{-t} \cdot t^{(x-2)}dt \\
{}^{[1]} &= -0 + (x-1) \int_0^\infty e^{-t} \cdot t^{(x-2)}dt \\
&= (x-1) \int_0^\infty e^{-t} \cdot t^{((x-1)-1)}dt \\
&= (x-1) \int_0^\infty t^{((x-1)-1)} \cdot e^{-t}dt \\
&= (x-1) \cdot \Gamma(x-1)
\end{aligned}$$

1. 17.2.1

17.2.3 Corollary

$$\Gamma(x) = \frac{1}{x} \cdot \Gamma(x+1)$$

Proof:

Theorem 17.2.2 establishes

$$\Gamma(x) = (x-1) \cdot \Gamma(x-1)$$

Let $y = x + 1$. Then

$$\begin{aligned}
\Gamma(y) &= (y-1) \cdot \Gamma(y-1) \\
\Rightarrow \Gamma(x+1) &= (x+1-1) \cdot \Gamma(x+1-1) \\
&= x \cdot \Gamma(x) \\
\Rightarrow \frac{1}{x} \cdot \Gamma(x+1) &= \Gamma(x) \\
\Rightarrow \Gamma(x) &= \frac{1}{x} \cdot \Gamma(x+1)
\end{aligned}$$

This allows the recurrence relation to move toward $\Gamma(0)$ for any value of x. Note, however, that $\Gamma(0)$ is undefined. Thus, solutions for the Gamma Function may be determined for positive integers, since $\Gamma(1)$ can be solved. On the other hand, $\Gamma(-1)$ can not be solved, and the recurrence relation results in a zero denominator. Hence, the Gamma Function is defined for all $x \in \mathbb{R}$ so long as $x \notin \mathbb{Z}^-$

17.2.4 Theorem:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Proof:

$$\begin{aligned}
\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty t^{-\frac{1}{2}} e^{-t} dt \\
&= \frac{2}{2} \int_0^\infty t^{-\frac{1}{2}} e^{-t} dt \\
&= 2 \int_0^\infty \frac{1}{2} \cdot t^{-\frac{1}{2}} e^{-(\sqrt{t})^2} dt \\
&= 2 \int_0^\infty \frac{1}{2} \cdot t^{-\frac{1}{2}} e^{-(\sqrt{t})^2} dt \\
{}^{[1]} &= 2 \int_0^\infty \frac{1}{2} \cdot x^{-1} e^{-x^2} \cdot 2 \cdot x \; dx \\
&= 2 \int_0^\infty \frac{2x}{2x} e^{-x^2} dx \\
&= 2 \int_0^\infty e^{-x^2} dx \\
{}^{[2]} &= 2 \cdot \frac{\sqrt{\pi}}{2} \\
&= \sqrt{\pi}
\end{aligned}$$

1. $x = \sqrt{t}$; $t = x^2$; and $dt = 2xdx$
2. Theorem 18.1.2

17.2.5 Theorem:

Let c be a constant such that $c > 0$. Then

$$\int_0^\infty t^x e^{-ct} \cdot dt = \frac{\Gamma(x+1)}{c^{x+1}}$$

Proof:

$$\begin{aligned}
\int_0^\infty t^x e^{-ct} dt \\
[1] &= \int_0^\infty \left(\frac{y}{c}\right)^x e^{-c\frac{y}{c}} \frac{dy}{c} \\
&= \frac{1}{c^x} \cdot \frac{1}{c} \int_0^\infty y^x e^{-y} dy \\
&= \frac{1}{c^{x+1}} \int_0^\infty y^{(x+1)-1} e^{-y} dy \\
[2] &= \frac{1}{c^{x+1}} \cdot \Gamma(x+1) \\
&= \frac{\Gamma(x+1)}{c^{x+1}}
\end{aligned}$$

1. Let $y = ct$. Then $y = \frac{y}{c}$ and $dt = \frac{dy}{c}$.
2. $\int_0^\infty y^{(x+1)-1} e^{-y} dy$ satisfies the form of a Gamma function. 17.1

17.3 References

- Pennsylvania State University, Elberly College of Science, STAT 414/415, https://onlinecourses.science.psu.edu/stat414/node/142
- Theodore Hatch Whitfield, Lecture Notes, E156 *Mathematical Foundations of Statistical Software.*

Chapter 18

Gaussian Integral

The Gaussian Integral is defined by

$$\int_{-\infty}^{\infty} e^{-x^2} \cdot dx$$

The Gaussian Integral may be generalized to the form

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2} \cdot dx$$

18.1 Theorems for the Gaussian Integral

18.1.1 Theorem

The Gaussian Integral is an even function.

Proof:

Recall that an even function is a function $f(x)$ such that $f(-x) = f(x)$.

Let $f(x)$ be the Gaussian Integral

$$f(x) = \int_{-\infty}^{\infty} e^{-x^2} \cdot dx$$

$$\begin{aligned} f(-x) &= \int_{-\infty}^{\infty} e^{-(-x)^2} \cdot dx \\ &= \int_{-\infty}^{\infty} e^{-x^2} \cdot dx \\ &= f(x) \end{aligned}$$

18.1.2 Theorem

$$\int_{-\infty}^{\infty} e^{-x^2} \cdot dx = \sqrt{\pi}$$

Proof:

Let $y = x$, and let $I = \int_{-\infty}^{\infty} e^{-x^2} \cdot dx$. This permits the equation

$$I = \int_{-\infty}^{\infty} e^{-x^2} \cdot dx = \int_{-\infty}^{\infty} e^{-y^2} \cdot dy$$

We use this equality to define the double integral for I^2.

$$\begin{aligned}
I^2 &= I \cdot I \\
&= \Big(\int_{-\infty}^{\infty} e^{-x^2} \cdot dx\Big) \cdot \Big(\int_{-\infty}^{\infty} e^{-y^2} \cdot dx\Big) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2} e^{-y^2} \cdot dx\ dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} \cdot dx\ dy \\
{}^{[1]} &= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2} \cdot r \cdot dr\ d\theta \\
&= \int_{0}^{2\pi} -\frac{1}{2} e^{-r^2}\Big|_{r=0}^{r=\infty} \cdot d\theta \\
&= \int_{0}^{2\pi} -0 - \Big(-\frac{1}{2}\Big) \cdot d\theta \\
&= \int_{0}^{2\pi} \frac{1}{2} \cdot d\theta \\
&= \frac{\theta}{2}\Big|_{\theta=0}^{\theta=2\pi} \\
&= \frac{2\pi}{2} - \frac{0}{2} \\
&= \frac{2\pi}{2} \\
&= \pi
\end{aligned}$$

1. Conversion to polar coordinates. Let the radius be $r = x^2 + y^2$ on the domain of $[0, \infty]$ and let the angle be θ on the domain of $[0, 2\pi]$. $dx\ dy = rdr\ d\theta$.

This establishes that $I^2 = \pi$. It follows:

$$I^2 = \pi$$
$$I = \sqrt{\pi}$$
$$\int_{-\infty}^{\infty} e^{-x^2} = \sqrt{\pi}$$

18.1.3 Theorem

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2} \cdot dx = \sqrt{\frac{\pi}{a}}$$

Proof:

Let $y = x$, and let $I = \int_{-\infty}^{\infty} e^{-a(x+b)^2} \cdot dx$. This permits the equation

$$I = \int_{-\infty}^{\infty} e^{-a(x+b)^2} \cdot dx = \int_{-\infty}^{\infty} e^{-a(y+b)^2} \cdot dy$$

We use this equality to define the double integral for I^2.

$$
\begin{aligned}
I^2 &= I \cdot I \\
&= \Big(\int\limits_{-\infty}^{\infty} e^{-a(x+b)^2} \cdot dx\Big) \cdot \Big(\int\limits_{-\infty}^{\infty} e^{-a(y+b)^2} \cdot dx\Big) \\
&= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{-a(x+b)^2} e^{-a(y+b)^2} \cdot dx\ dy \\
&= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{-(a(x+b)^2 + a(y+b)^2)} \cdot dx\ dy \\
{}^{[1]} &= \int\limits_{0}^{2\pi} \int\limits_{0}^{\infty} e^{-2a(r+b)^2} \cdot r \cdot dr\ d\theta \\
&= \int\limits_{0}^{2\pi} -\frac{1}{2a} e^{-r^2} \Big|_{r=0}^{r=\infty} \cdot d\theta \\
&= \int\limits_{0}^{2\pi} -0 - \big(-\frac{1}{2a}\big) \cdot d\theta \\
&= \int\limits_{0}^{2\pi} \frac{1}{2a} \cdot d\theta \\
&= \frac{\theta}{2a}\Big|_{\theta=0}^{\theta=2\pi} \\
&= \frac{2\pi}{2a} - \frac{0}{2} \\
&= \frac{2\pi}{2a} \\
&= \frac{\pi}{a}
\end{aligned}
$$

1. Conversion to polar coordinates. Let the radius be $r = a(x+b)^2 + a(y+b)^2$ on the domain of $[0, \infty]$ and let the angle be θ on the domain of $[0, 2\pi]$. $dx\ dy = rdr\ d\theta$.

This establishes that $I^2 = \frac{\pi}{a}$. It follows:

$$
\begin{aligned}
I^2 &= \frac{\pi}{a} \\
I &= \sqrt{\frac{\pi}{a}} \\
\int\limits_{-\infty}^{\infty} e^{-a(x+b)^2} &= \sqrt{\frac{\pi}{a}}
\end{aligned}
$$

18.2 References

- Theodore Hatch Whitfield, Lecture Notes, E156 *Mathematical Foundations of Statistical Software*.

Chapter 19

Geometric Distribution

19.1 First Success as a Random Variable

The Geometric Distribution random variable may be considered as the number of trials required to achieve the first success. In this consideration, if x is the total number of trials and k is the number of failures, $x = k + 1$.

19.1.1 Probability Mass Function

A random variable X is said to have a Geometric Distribution with parameter p if its probability mass function is

$$p(x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, 3, \dots \\ 0 & otherwise \end{cases}$$

where p is the probability of a success on any given trial and x is the number of trials until the first success.

19.1.2 Cumulative Distribution Function

The cumulative probability of x is computed as the x^{th} partial sum of the Geometric Series See 20.1.1.

$$\begin{aligned} P(x) &= \sum_{i=1}^{x} p(1-p)^{x-1} \\ &= \frac{p - p(1-p)^{x-1}}{1-(1-p)} \\ &= \frac{p[1-(1-p)^{x-1}]}{p} \\ &= 1-(1-p)^{x-1} \end{aligned}$$

So

$$P(x) = \begin{cases} 1-(1-p)^{x-1}, & x = 1, 2, 3, \dots \\ 0 & otherwise \end{cases}$$

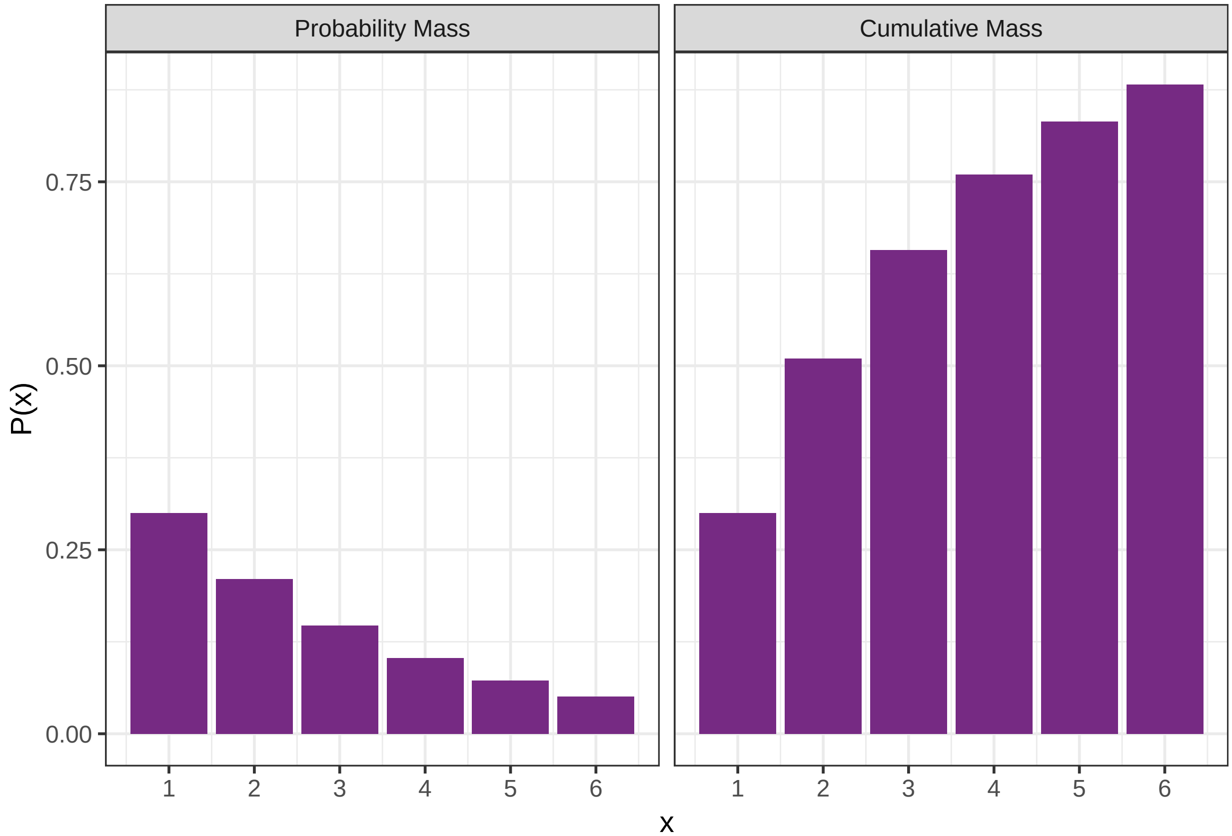

Figure 19.1: The figures on the left and right display the Geometric probability and cumulative distirubtion functions, respectively, for the combinations of $p = .3$.

A recursive form of the cdf can be derived and has some usefulness in computer applications. With it, one need only initiate the first value and additional cumulative probabilities can be calculated. It is derived as follows:

$$\begin{aligned} P(X = x + 1) &= p(1-p)^x \\ &= (1-p)p(1-p)^{x-1} \\ &= (1-p)P(X = x) \end{aligned}$$

19.1.3 Expected Values

$$
\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x \cdot p(1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} x \cdot (1-p)^{x-1} \\
{}^{[1]} &= p \sum_{x=1}^{\infty} x \cdot q^{x-1} \\
{}^{[2]} &= p \frac{d}{dq} \Big(\sum_{x=1}^{\infty} q^{x} \Big) \\
&= p \frac{d}{dq} \Big(\sum_{x=1}^{\infty} q \cdot q^{x-1} \Big) \\
{}^{[3]} &= p \frac{d}{dq} \Big(\frac{q}{1-q} \Big) \\
{}^{[4]} &= p \Big(\frac{(1-q) \cdot 1 - q(-1)}{(1-q)^2} \Big) \\
&= p \Big(\frac{1-q+q}{(1-q)^2} \Big) \\
&= p \Big(\frac{1}{(1-q)^2} \Big) \\
{}^{[5]} &= p \cdot \frac{1}{p^2} && = \frac{p}{p^2} = \frac{1}{p}
\end{aligned}
$$

1. Let $1 - p = 1$
2. $a \cdot x^{a-1} = \frac{d}{dx}(x^a)$
3. $\sum\limits_{k=1}^{\infty} ar^{k-1} = \frac{a}{1-r}$ (Geometric Series 20.1.1).
4. Product Rule for Differentiation:
   $\frac{d}{dx}\big(\frac{f(x)}{g(x)}\big) = \frac{g\prime(x) \cdot f(x) - f\prime(x) \cdot g(x)}{g^2(x)}$
5. $1 - p = q \quad \Rightarrow p = 1 - q$

$$
\begin{aligned}
E[X(X-1)]^{[1]} &= \sum_{x=2}^{\infty} x(x-1)p(1-p)^{x-1} \\
&= p(1-p)\sum_{x=2}^{\infty} x(x-1)(1-p)^{x-2} \\
{}^{[2]} &= pq\sum_{x=2}^{\infty} x(x-1)q^{x-2} \\
{}^{[3]} &= pq\frac{d^2}{dq^2}\Big(\sum_{x=2}^{\infty} q^x\Big) \\
&= pq\frac{d^2}{dq^2}\Big(\sum_{x=2}^{\infty} q\cdot q^{x-1}\Big) \\
{}^{[4]} &= pq\frac{d^2}{dq^2}\Big(\frac{2q-1}{1-q}\Big) \\
{}^{[5]} &= pq\frac{d}{dq}(-(1-q)^{-2}) \\
{}^{[6]} &= pq\frac{2}{(1-q)^3} \\
&= \frac{2pq}{(1-q)^3} \\
{}^{[7]} &= \frac{2p(1-p)}{p^3} \\
&= \frac{2(1-p)}{p^2}
\end{aligned}
$$

1. We start the summand at $x = 2$ because the term for $x = 1$ is 0.
2. Let $1 - p = 1$
3. $a(a-1)x^{a-2} = \frac{d^2}{dx^2}x^a$
4. $\sum\limits_{k=1}^{\infty} ar^{k-1} = \frac{a}{1-r} = 1 + a + ar + ar^2 + \cdots$. The current series leaves out the first term,\
$\sum\limits_{k=2}^{\infty} ar^{k-1} = (\sum\limits_{k=1}^{\infty} ar^{k-1}) - 1 = \frac{a}{1-r} - 1 = \frac{1}{1-r} - \frac{1-r}{1-r} = \frac{a-(1-r)}{1-r} = \frac{a+r-1}{r-1}$
5. $\frac{d}{dx}(\frac{2x-1}{1-x}) = \frac{-(2x-1)-2(1-x)}{(1-x)^2} = \frac{2x+1-2+2x}{(1-x)^2} = \frac{-1}{(1-x)^2} = -(1-x)^{-2}$
6. $\frac{d}{dx} - (1-x)^{-2} = 2(1-x)^{-3} = \frac{2}{(1-x)^3}$
7. See note number 5.

$$
\begin{aligned}
\mu &= E(X) \\
&= \frac{1}{p}
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= E(X^2) - E(X)^2 \\
&= E(X^2) - E(X) + E(X) - E(X)^2 \\
&= [E(X^2) - E(X)] + E(X) - E(X)^2 \\
&= E(X^2 - X) + E(X) - E(X)^2 \\
&= E[(X(X-1)] + E(X) - E(X)^2 \\
&= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{2(1-p)}{p^2} + \frac{p}{p^2} - \frac{1}{p^2} \\
&= \frac{2(1-p)+p-1}{p^2} \\
&= \frac{2-2p+p-1}{p^2} \\
&= \frac{2-1+p-2p}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

19.1.4 Moment Generating Function

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=1}^{\infty} e^{tx} p(1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} e^{t^x}(1-p)^{x-1} \\
&= pe^t \sum_{x=1}^{\infty} e^{t^{(x-1)}} \\
&= pe^t \sum_{x=1}^{\infty} [(1-p)e^t]^{x-1} \\
{}^{[1]} &= pe^t \frac{1}{1-(1-p)e^t} \\
&= \frac{pe^t}{1-(1-p)e^t}
\end{aligned}
$$

1. $\sum\limits_{k=1}^{\infty} ar^{k-1} = \frac{a}{1-r}$, (Geometric Series 20.1.1)

$$\begin{aligned}
M_X^{(1)}(t) &= \frac{[1-(1-p)e^t]pe^t - pe^t[-(1-p)e^t]}{\left(1-(1-p)e^t\right)^2} \\
&= \frac{pe^t[1-(1-p)e^t + (1-pe^t)]}{\left(1-(1-p)e^t\right)^2} \\
&= \frac{pe^t(1)}{\left(1-(1-p)e^t\right)^2} \\
&= \frac{pe^t}{\left(1-(1-p)e^t\right)^2}
\end{aligned}$$

$$\begin{aligned}
M_X^{(2)}(t) &= \frac{(1-(1-p)e^t)^2pe^t - pe^t[-2(1-(1-p)e^t)(1-p)e^t}{\left(1-(1-p)e^t\right)^4} \\
&= \frac{pe^t[(1-(1-p)e^t)^2 + 2(1-(1-p)e^t(1-p)e^t]}{\left(1-(1-p)e^t\right)^4}
\end{aligned}$$

$$\begin{aligned}
E(X) &= M_X^{(1)}(0) \\
&= \frac{pe^0}{\left(1-(1-p)e^0\right)^2} \\
&= \frac{p}{\left(1-(1-p)\right)^2} \\
&= \frac{p}{(1-1+p)^2} \\
&= \frac{p}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= M_X^{(2)}(0) \\
&= \frac{pe^0[(1-(1-p)e^0)^2 + 2(1-(1-p)e^0)(1-p)e^0]}{(1-(1-p)e^0)^4} \\
&= \frac{p[(1-(1-p))^2 + 2(1-(1-p))(1-p)]}{(1-(1-p))^4} \\
&= \frac{p[(1-1+p)^2 + 2(1-1+p)(1-p)]}{(1-1+p)^4} \\
&= \frac{p[p^2 = 2p(1-p)]}{p^4} \\
&= \frac{p(p^2 + 2p - 2p^2)}{p^4} \\
&= \frac{p(2p - p^2)}{p^4} \\
&= \frac{p^2(2-p)}{p^4} \\
&= \frac{2-p}{p^2}
\end{aligned}$$

$$\mu = E(X)$$

19.1.5 Maximum Likelihood Estimator

Let $x_1, x_2, \ldots, x_n$ be a random sample from a Geometric distribution with parameter p.

19.1.5.1 Likelihood Function

$$\begin{aligned}L(\theta) &= P(x_1, x_2, \ldots, x_n|\theta) \\ &= p(1-p)^{x-1}\end{aligned}$$

19.1.5.2 Log-likelihood Function

$$\begin{aligned}\ell(\theta) &= \ln p(1-p)^{x-1} \\ &= \ln p + \ln(1-p)^{x-1} \\ &= \ln p + (x-1)\ln(1-p)\end{aligned}$$

19.1.5.3 MLE for p

$$\begin{aligned}\frac{d\ell}{dp} &= \frac{1}{p} + \frac{-(x-1)}{1-p} \\ &= \frac{1}{p} + \frac{1-x}{1-p}\end{aligned}$$

$$\begin{aligned}0 &= \frac{1}{p} + \frac{1-x}{1-p} \\ \Rightarrow \frac{1-x}{1-p} &= -\frac{1}{p} \\ \Rightarrow 1-x &= \frac{-1+p}{p} \\ \Rightarrow -x &= \frac{-1+p}{p} - 1 \\ \Rightarrow x = 1 - \frac{-1+p}{p} &= \frac{p}{p} - \frac{-1+p}{p} \\ &= \frac{p+1-p}{p} \\ &= \frac{1}{p} \\ \Rightarrow p &= \frac{1}{x}\end{aligned}$$

So

$$\hat{p} = \frac{1}{x}$$

is the Maximum Likelihood Estimator for p.

19.1.6 Theorems for the Geometric Distribution

19.1.6.1 Validity of the Distribution

$$\sum_{i=1}^{\infty} p(1-p)^{x-1} = 1$$

Proof:

$$\begin{aligned}\sum_{i=1}^{\infty} p(1-p)^{x-1} &= \frac{p}{1-(1-p)} \\ &= \frac{p}{p} \\ &= 1\end{aligned}$$

1. $S = \lim\limits_{k\to\infty} S_k = \lim\limits_{k\to\infty} \frac{a-ar^k}{1-r}$ (Geometric Series 20.1.1)

19.1.6.2 Sum of Geometric Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent random variables from a Geometric Distribution with parameter p, that is, $X_i \sim$Geometric(p), $i = 1, 2, 3, \ldots$. Let $Y = \sum\limits_{i=1}^{n} X_i$. Then $Y \sim$Negative Binomial(n, p).

Proof:

$$\begin{aligned}M_Y(t) &= E(e^{tY}) \\ &= E(e^{t(X_1+X_2+\cdots+X_n)}) \\ &= E(e^{tX_1}e^{tX_2}\cdots e^{tX_n}) \\ &= E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n}) \\ &= \frac{pe^t}{1-(1-p)e^t}\cdot\frac{pe^t}{1-(1-p)e^t}\cdot\ \cdots\ \cdot\frac{pe^t}{1-(1-p)e^t} \\ &= \Big(\frac{pe^t}{1-(1-p)e^t}\Big)^n\end{aligned}$$

Which is the mgf of a Negative Binomial random variable with parameters n and p. Thus $Y \sim$Negative Binomial(n, p).

19.1.6.3 Lemma

Let X be a Geometric random variable with parameter p. Then $P(X > a) = (1-p)^a$.

Proof:

$$
\begin{aligned}
P(X > a) = 1 - P(X \leq a) = 1 - \sum_{i=1}^{a} p(1-p)^i - 1 \\
{}^{[1]} &= 1 - \frac{p - p(1-p)^a}{1-(1-p)} \\
&= 1 - \frac{p\big(1-(1-p)^a\big)}{1-1+p} \\
&= 1 - \frac{p\big(1-(1-p)^a\big)}{p} \\
&= 1 - \big(1-(1-p)^a\big) \\
&= 1 - 1 + (1-p)^a \\
&= (1-p)^a
\end{aligned}
$$

1. $S_k = \lim_{k\to\infty} \frac{a-ar^k}{1-r}$ (Geometric Series 20.1.1)

19.1.6.4 Memoryless Property

$$P(X \geq k+j | X \geq k) = P(X \geq k)$$

Proof:

$$P(X > a+b)^{[1]} = (1-p)^{a+b} = (1-p)^a(1-p)^b$$

$$
\begin{aligned}
P(X > k+j | X > k) &= \frac{P(X > k+j \cap X > k)}{P(X > k)} \\
{}^{[2]} &= \frac{P(X > k+j)}{P(X > k)} \\
&= \frac{(1-p)^k(1-p)^j}{(1-p)^k} \\
&= (1-p)^j \\
&= P(X > j)
\end{aligned}
$$

1. Geometric Distribution 19.1.6.3
2. Since j must be a positive integer in the Geometric Distribution, it is certain that \$(k+j) \cap k = k+j$.

19.2 Number of Failures as Random Variable

The Geometric Distribution random variable may be considered the number of failures observed before a success is observed. In this consideration, if x is the number of failures, then the total number of trials is $x+1$.

19.2.1 Probability Mass Function

A random variable X is said to have a Geometric Distribution with parameter p if its probability mass function is

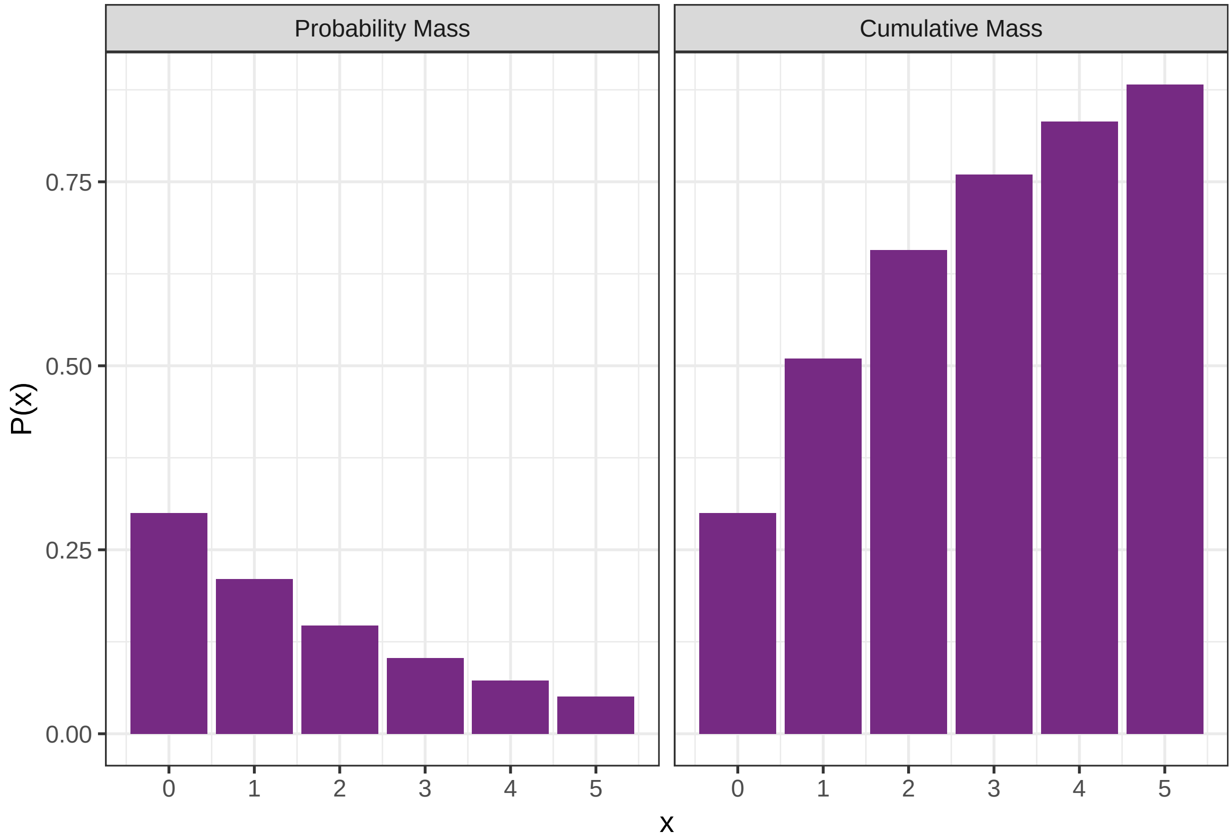

Figure 19.2: The figures on the left and right display the Geometric probability and cumulative distirubtion functions, respectively, for the combinations of $p = .3$.

$$p(x) = \begin{cases} p(1-p)^x, & x = 0, 1, 2, 3, \ldots \\ 0 & otherwise \end{cases}$$

19.2.2 Cumulative Distribution Function

The cumulative probability of x is computed as the x^{th} partial sum of the Geometric Series See 20.1.1.

$$\begin{aligned} P(x) &= \sum_{i=0}^{x} p(1-p)^x \\ &= \frac{p - p(1-p)^{x+1}}{1-(1-p)} \\ &= \frac{p[1-(1-p)^{x+1}]}{p} \\ &= 1-(1-p)^{x+1} \end{aligned}$$

So

$$P(x) = \begin{cases} 1-(1-p)^{x+1}, & x = 0, 1, 2, 3, \ldots \\ 0 & otherwise \end{cases}$$

Chapter 20

Geometric Series

20.1 Series Starts at $k = 1$

A Geometric Series is a series of the form $\sum\limits_{k=1}^{\infty} ar^{k-1}$ where $a \neq 0,\ r \neq 0, 1$. Expanding the series gives $\sum\limits_{k=1}^{\infty} ar^{k-1} = a + ar + ar^2 + ar^3 + \cdots$.

20.1.1 Partial and Infinite Summations

Let S_k denote the sum of a series over k terms (or the k^{th} partial sum). For the Geometric Series\

$$\begin{aligned} S_k &= \sum_{k=1}^{k} ar^{k-1} \\ &= a + ar + ar^2 + ar^3 + \cdots + ar^{k-1} \\ &= ar^0 + ar^2 + ar^2 + \cdots + ar^{k-1} \\ &= a + ar + ar^2 + \cdots + ar^{k-1} \end{aligned}$$

Notice that $rS_k = ar + ar^2 + ar^3 + \cdots + ar^k$. So

$$\begin{aligned} S_k - rS_k &= (a + ar + \cdots + ar^{k-1}) - (ar + ar^2 + \cdots + ar^k) \\ &= a + ar - ar + ar^2 - ar^2 + \cdots + ar^{k-2} - ar^{k-2} + ar^{k-1} - ar^{k-1} - ar^k \\ &= a - ar^k \end{aligned}$$

Observing that $S_k - rS_k = S_k(1 - r)$, we may conclude

$$\begin{aligned} S_k(1 - r) &= S_k - rS_k \\ &= a - ar^k \\ \Rightarrow S_k &= \frac{a - ar^k}{1 - r} \end{aligned}$$

20.1.2 Proofs of Convergence

$\sum\limits_{k=1}^{\infty} ar^{k-1}$ converges when $|r| < 1$ and diverges when $|r| > 1$.

Proof:

Recall that the k^{th} partial sum of the Geometric Series is

$$S_k = \frac{a - ar^k}{1 - r}$$

And let S denote the sum of the infinite series, i.e. the sum as$k \to \infty$.

Case 1: $|r| < 1$

$$\begin{aligned} S &= \lim_{k\to\infty} S_k \\ &= \lim_{k\to\infty} \frac{a - ar^k}{1 - r} \\ &= \frac{a - \lim\limits_{k\to\infty} ar^k}{1 - r} \\ &= \frac{a}{1 - r} \end{aligned}$$

So $\sum\limits_{k=1}^{\infty} ar^{k-1}$ converges when $|r| < 1$ and $S = \frac{a}{1-r}$

Case 2: $|r| > 1$

$$\begin{aligned} S &= \lim_{k\to\infty} S_k \\ &= \lim_{k\to\infty} \frac{a - ar^k}{1 - r} \\ &= \frac{a - \lim\limits_{k\to\infty} ar^k}{1 - r} \\ &= \frac{a - \infty}{1 - r} \end{aligned}$$

So $\sum\limits_{k=1}^{\infty} ar^{k-1}$ diverges whern $|r| > 1$.

20.2 Series Starts at $k = 0$

A Geometric Series is a series of the form $\sum\limits_{k=0}^{\infty} ar^k$ where $a \neq 0,\ r \neq 0, 1$. Expanding the series gives $\sum\limits_{k=0}^{\infty} ar^k = a + ar + ar^2 + ar^3 + \cdots$.

20.2.1 Partial and Infinite Summations

Let S_k denote the sum of a series over k terms (or the k^{th} partial sum). For the Geometric Series\

$$\begin{aligned} S_k &= \sum_{k=0}^{n} ar^k \\ &= a + ar + ar^2 + ar^3 + \cdots + ar^k \\ &= ar^0 + ar^2 + ar^2 + \cdots + ar^k \\ &= a + ar + ar^2 + \cdots + ar^k \end{aligned}$$

Notice that $rS_k = ar + ar^2 + ar^3 + \cdots + ar^k$. So

$$\begin{aligned} S_k - rS_k &= (a + ar + \cdots + ar^k) - (ar + ar^2 + \cdots + ar^{k+1}) \\ &= a + ar - ar + ar^2 - ar^2 + \cdots + ar^{k-1} - ar^{k-1} + ar^k + ar^k - ar^{k+1} \\ &= a - ar^{k+1} \end{aligned}$$

Observing that $S_k - rS_k = S_k(1 - r)$, we may conclude

$$\begin{aligned} S_k(1 - r) &= S_k - rS_k \\ &= a - ar^{k+1} \\ \Rightarrow S_k &= \frac{a - ar^{k+1}}{1 - r} \end{aligned}$$

20.2.2 Proofs of Convergence

$\sum\limits_{k=0}^{\infty} ar^k$ converges when $|r| < 1$ and diverges when $|r| > 1$.

Proof:

Recall that the k^{th} partial sum of the Geometric Series is

$$S_k = \frac{a - ar^{k+1}}{1 - r}$$

And let S denote the sum of the infinite series, i.e. the sum as$k \to \infty$.

Case 1: $|r| < 1$

$$\begin{aligned} S &= \lim_{k \to \infty} S_k \\ &= \lim_{k \to \infty} \frac{a - ar^{k+1}}{1 - r} \\ &= \frac{a - \lim\limits_{k \to \infty} ar^{k+1}}{1 - r} \\ &= \frac{a}{1 - r} \end{aligned}$$

So $\sum\limits_{k=1}^{\infty} ar^k$ converges when $|r| < 1$ and $S = \frac{a}{1-r}$

Case 2: $|r| > 1$

$$\begin{aligned} S &= \lim_{k\to\infty} S_k \\ &= \lim_{k\to\infty} \frac{a - ar^{k+1}}{1-r} \\ &= \frac{a - \lim_{k\to\infty} ar^{k+1}}{1-r} \\ &= \frac{a-\infty}{1-r} \end{aligned}$$

So $\sum_{k=1}^{\infty} ar^{k-1}$ diverges whern $|r| > 1$.

Chapter 21

Hypergeometric Distribution

21.1 Probability Mass Function

A random variable X is said to follow a Hypergeometric Distribution if its probability mass function is

$$p(x) = \begin{cases} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, & x = 0, 1, 2, \ldots \\ 0 & otherwise \end{cases}$$

where

- N is the number of objects available to choose from
- n is the number of objects chosen from N
- r is the number of objects in N that posses a desired characteristic (successes)
- x is the number of objects in n that possess the desired characterstic

21.2 Cumulative Mass Function

$$P(x) = \begin{cases} \sum\limits_{i=0}^{x} \frac{\binom{r}{i}\binom{N-r}{n-i}}{\binom{N}{n}}, & x = 0, 1, 2, \ldots \\ 0 & otherwise \end{cases}$$

21.3 Expected Values

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \\
&= \sum_{x=0}^{n} x \binom{r}{x} \frac{\binom{N-r}{n-x}}{\binom{N}{n}} \\
{[1]} &= \sum_{x=0}^{n} x \frac{r}{x} \binom{r-1}{x-1} \frac{\binom{N-r}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} \\
&= \frac{rn}{N} \sum_{x=0}^{n} \frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}} \\
{[2]} &= \frac{rn}{N} \sum_{y=0}^{n-1} \frac{\binom{r-1}{y}\binom{N-r}{n-y-1}}{\binom{N-1}{n-1}} \frac{rn}{N} \sum_{y=0}^{n-1} \frac{\binom{r-1}{y}\binom{N-r}{n-y-1}}{\binom{N-1}{n-1}} \\
&= \frac{\frac{rn}{N} \cdot n}{N} r \sum_{y=0}^{n-1} \frac{\binom{r-1}{y}\binom{N-r}{n-y-1}}{\binom{N-1}{n-1}} \\
{[3]} &= \frac{rn}{N} \cdot 1 \\
&= \frac{rn}{N}
\end{aligned}
$$

1. For any integer a such that $0 \le a \le k$, $\binom{n}{k} = \frac{n(n-1)\cdots(n-a+1)}{k(k-1)\cdots(k-a+1)}\binom{n-a}{k-a}$ (Theorem 10.0.3).
2. Let $y = x - 1 \ \Rightarrow x = y + 1$.
3. $\frac{\sum_{i=1}^{n}\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}} = 1\backslash$ with $N_1 = r,\ N_2 = N - r,\ i = x.$ (Theorem 6.3.2

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{n} x(x-1) \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \\
{[1]} &= \sum_{x=0}^{n} \frac{x(x-1)r(r-1)}{x(x-1)} \frac{\binom{r-2}{x-2}\binom{N-r}{n-x}}{\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}} \\
&= \frac{r(r-1)n(n-1)}{N(N-1)} \sum_{x=0}^{n} \frac{\binom{r-2}{x-2}\binom{N-r}{n-x}}{\binom{N-2}{n-2}} \\
{[2]} &= \frac{r(r-1)n(n-1)}{N(N-1)} \sum_{y=0}^{n-2} \frac{\binom{r-2}{y}\binom{N-r}{n-y-2}}{\binom{N-2}{n-2}} \\
{[3]} &= \frac{r(r-1)n(n-1)}{N(N-1)} \cdot 1 \\
&= \frac{r(r-1)n(n-1)}{N(N-1)}
\end{aligned}
$$

1. For any integer a such that $0 \le a \le k$, $\binom{n}{k} = \frac{n(n-1)\cdots(n-a+1)}{k(k-1)\cdots(k-a+1)}\binom{n-a}{k-a}$ (Theorem 10.0.3).
2. Let $y = x - 1 \ \Rightarrow x = y + 1$.
3. $\frac{\sum_{i=1}^{n}\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}} = 1\backslash$ with $N_1 = r,\ N_2 = N - r,\ i = x.$ (Theorem 6.3.2

$$\begin{aligned}
\mu &= E(X) \\
&= \frac{rn}{N}
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= E(X^2) - E(X)^2 \\
&= E(X^2) - E(X) + E(X) - E(X)^2 \\
&= (E(X^2) - E(X) + E(X) - E(X)^2 \\
&= E(X^2 - X) + E(X) - E(X)^2 \\
&= E[X(X-1)] + E(X) - E(X)^2 \\
&= \frac{r(r-1)n(n-1)}{N(N-1)} + \frac{rn}{N} - \frac{r^2n^2}{N^2} \\
&= \frac{r(r-1)n(n-1)N}{N^2(N-1)} + \frac{rnN(N-1)}{N^2(N-1)} - \frac{r^2n^2(N-1)}{N^2(N-1)} \\
&= \frac{(r^2-r)(n^2-n)Nrn(N^2-N) - r^2n^2(N-1)}{N^2(N-1)} \\
&= \frac{(r^2n^2N - r^2n^2N - rn^2N + rnN + rnN^2 - rnN - r^2n^2N + r^2n^2}{N^2(N-1)} \\
&= \frac{-r^2nN - rn^2N + rnN^2 + r^2n^2}{N^2(N-1)} \\
&= \frac{nr(-rN - nN + N^2 + rn}{N^2(N-1)} \\
&= \frac{nr(N^2 - nN - rN + rn}{N^2(N-1)} \\
&= \frac{nr(N-r)(N-n)}{N^2(N-1)} \\
&= \frac{nr(N-r)(N-n)}{N \cdot N(N-1)} \\
&= \frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}
\end{aligned}$$

21.4 Moment Generating Function

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{n} e^{tx} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \\
&= \frac{1}{\binom{N}{n}} \sum_{x=0}^{n} e^{tx} \binom{r}{x}\binom{N-r}{n-x} \\
&= \frac{1}{\binom{N}{n}} [e^{0t}\binom{r}{0}\binom{N-r}{n-0} + e^{1t}\binom{r}{1}\binom{N-r}{n-1} + e^{2t}\binom{r}{2}\binom{N-r}{n-2} + \cdots + e^{nt}\binom{r}{n}\binom{N-r}{n-n}] \\
&= \frac{1}{\binom{N}{n}} [\binom{N-r}{n-0} + e^{t}\binom{r}{1}\binom{N-r}{n-1} + e^{2t}\binom{r}{2}\binom{N-r}{n-2} + \cdots + e^{nt}\binom{r}{n}\binom{N-r}{n-n}]
\end{aligned}$$

This mgf does not reduce to any form which can be differentiated, and we cannot use it to generate moments for the distribution.

21.5 Theorems for the Hypergeometric Distribution

21.5.1 Validity of the Distribution

$$\sum_{x=0}^{n} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = 1$$

Proof:

Theorem 6.3.1 states

$$\binom{N_1}{0}\binom{N_2}{n} + \binom{N_1}{2}\binom{N_2}{n-1} + \cdots + \binom{N_1}{n-1}\binom{N_2}{1} + \binom{N_1}{n}\binom{N_2}{0} = \sum_{x=0}^{n} \binom{N_1}{x}\binom{N_2}{n-x} = \binom{N_1+N_2}{n}$$

Using $N_1 = r$ and $N_2 = N - r$ we have

$$\begin{aligned}
\sum_{x=0}^{n} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} &= \frac{\binom{r+N-r}{n}}{\binom{N}{n}} \\
&= \frac{\binom{N}{n}}{\binom{N}{n}} \\
&= 1
\end{aligned}$$

Chapter 22

Integration: Techniques and Theorems

22.1 Elementary Theorems

22.1.1 Integration of Even Functions about Zero

Suppose f is an integratable function, and let $F(x) = \int_{0}^{x_0} f(x)dx$.

Then $\int_{-x_0}^{0} f(x)dx = \int_{0}^{x_0} f(x)dx$ if and only if f is an even function.\

Proof:

First, let f be an even function. Then, by Theorem 15.4.5, the anti-derivative F is an odd function.

$$\begin{aligned}\int_{-x_0}^{0} f(x)dx &= F(0) - F(-x_0) \\ &= F(0) + F(x_0) \\ {}^{[1]} &= F(x_0)\end{aligned}$$

$$\begin{aligned}\int_{0}^{x_0} f(x)dx &= F(x_0) - F(0) \\ &= F(x_0)\end{aligned}$$

1. $F(0) = \int_{0}^{0} f(x)dx = 0$.

So

$$\int_{-x_0}^{0} f(x)dx = F(x_0)$$
$$= \int_{0}^{x_0} f(x)dx$$

Now suppose

$$\int_{-x_0}^{0} f(x)dx = \int_{0}^{x_0} f(x)dx$$

Then

$$\int_{-x_0}^{0} f(x)dx = F(0) - F(-x_0)$$
$$= -F(-x_0)$$

and

$$\int_{0}^{x_0} f(x)dx = F(x_0) - F(0)$$
$$= F(x_0)$$

So

$$-F(-x_0) = F(x_0)$$
$$\Rightarrow F(-x_0) = -F(x_0)$$

This satisfies the definition of an odd function. So by Theorem 15.4.4, f must be an even function.

22.1.2 Corollary

If f is a continuous and even function and $t \in \Re$, then

$$\int_{-t}^{t} f(x)dx = 2\int_{0}^{t} f(x)dx$$

Furthermore,

$$\int_{-\infty}^{\infty} f(x)dx = 2\int_{0}^{\infty} f(x)dx$$

.

Proof: Since $f(x)$ is even and by Theorem 22.1.1

$$\int_{-t}^{0} f(-x)dx = \int_{-t}^{0} f(x)dx$$
$$= \int_{0}^{t} f(x)dx$$

It follows that

$$\int_{-t}^{t} f(x)dx = \int_{-t}^{0} f(-x)dx + \int_{0}^{t} f(x)dx$$
$$= \int_{0}^{t} f(x)dx + \int_{0}^{t} f(x)dx$$
$$= 2\int_{0}^{t} f(x)dx$$

The second statement is proven by taking the limits as $t \to \infty$.

22.1.3 Integrals of Horizontal Translations

Let x be any real number and a, b, and c be constants. Also, let $f(x)$ be continuous on the interval (a, b). Then

$$\int_{a}^{b} f(x)dx = \int_{a+c}^{b+c} f(x+c)dx$$

Proof:

The proof of this theorem is completed by applying a change of variable to

$$\int_{a}^{b} f(x)dx$$

We let

$$y = x + c$$
$$\Rightarrow x = y - c$$

So $dx = dy$.

$$\begin{aligned} x = a \quad &\Rightarrow \quad y = a + c \\ x = b \quad &\Rightarrow \quad y = b + c \end{aligned}$$

.

Thus

$$\begin{aligned}
\int_a^b f(x)dx &= \int_{a+c}^{b+c} f(y)dy \\
&= \int_{a+c}^{b+c} f(x+c)dx
\end{aligned}$$

Chapter 23

Logarithms

23.1 The Natural Logarithm

23.1.1 Definition: The Natural Logarithm

For a positive number x the natural logarithm of x is defined as the integral

$$\ln x = \int_0^x \frac{1}{t}\ dt$$

Additionally, the base of the natural logarithm, denoted e, is the value such that $\ln e = 1$. That is

$$1 = \ln e = \int_0^e \frac{1}{t}\ dt$$

23.1.2 Theorem

23.1.2.1 Lemma 1

$$\int_1^{1+\frac{1}{n}} \frac{1}{1+\frac{1}{n}}\ dt = \frac{1}{n+1}$$

Proof:

$$
\begin{aligned}
\int_{1}^{1+\frac{1}{n}} \frac{1}{1+\frac{1}{n}}\ dt &= \frac{1}{1+\frac{1}{n}} \cdot t\ \Big|_{1}^{1+\frac{1}{n}} \\
&= \frac{1}{1+\frac{1}{n}} \cdot \left(1+\frac{1}{n}\right) - \frac{1}{1+\frac{1}{n}} \cdot 1 \\
&= 1 - \frac{1}{1+\frac{1}{n}} \\
&= \frac{1+\frac{1}{n}}{1+\frac{1}{n}} - \frac{1}{1+\frac{1}{n}} \\
&= \frac{\frac{1}{n}}{1+\frac{1}{n}} \\
&= \frac{\frac{1}{n}}{\frac{n}{n}+\frac{1}{n}} \\
&= \frac{\frac{1}{n}}{\frac{n+1}{n}} \\
&= \frac{1}{n} \cdot \frac{n}{n+1} \\
&= \frac{n}{n \cdot (n+1)} \\
&= \frac{1}{n+1}
\end{aligned}
$$

23.1.2.2 Lemma 2

$$
\int_{1}^{1+\frac{1}{n}} 1\ dt = \frac{1}{n}
$$

Proof:

$$
\begin{aligned}
\int_{1}^{1+\frac{1}{n}} 1\ dt &= t\Big|_{1}^{1+\frac{1}{n}} \\
&= 1 + \frac{1}{n} - 1 \\
&= \frac{1}{n}
\end{aligned}
$$

23.1.2.3 Theorem

$$
\lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = e
$$

Proof:

Let t be any real number such that $\frac{1}{1+\frac{1}{n}} \leq \frac{1}{t} \leq 1$. It follows that

$$\int_0^{1+\frac{1}{n}} \frac{1}{1+\frac{1}{n}}\,dt \leq \int_0^{1+\frac{1}{n}} \frac{1}{t}\,dt \quad \leq \int_0^{1+\frac{1}{n}} \frac{1}{1}\,dt$$

Using Lemmas 23.1.2.1 and 23.1.2.2, we may write

$$\frac{1}{n+1} \leq \int_0^{1+\frac{1}{n}} \frac{1}{t}\,dt \leq \frac{1}{n}$$

Next, by applying the definition of the natural logarithm (23.1.1)

$$\begin{aligned} \frac{1}{n+1} &\leq \ln\left(1+\frac{1}{n}\right) && \leq \frac{1}{n} \\ \Rightarrow \exp\left\{\frac{1}{n+1}\right\} &\leq \frac{1}{n+1} && \leq \exp\left\{\frac{1}{n}\right\} \end{aligned}$$

Now we isolate the left inequality to show

$$\begin{aligned} \exp\left\{\frac{1}{n+1}\right\} &\leq 1+\frac{1}{n} \\ \Rightarrow \exp\left\{\frac{1}{n+1}\right\}^{n+1} &\leq \left(1+\frac{1}{n}\right)^{n+1} \\ \Rightarrow \exp\left\{\frac{n+1}{n+1}\right\} &\leq \left(1+\frac{1}{n}\right)^{n+1} \\ \Rightarrow e &\leq \left(1+\frac{1}{n}\right)^{n+1} \end{aligned}$$

Isolating the right inequality allows us to show

$$\begin{aligned} 1+\frac{1}{n} &\leq \exp\left\{\frac{1}{n}\right\} \\ \Rightarrow \left(1+\frac{1}{n}\right)^n &\leq \exp\left\{\frac{1}{n}\right\}^n \\ \Rightarrow \left(1+\frac{1}{n}\right)^n &\leq \exp\left\{\frac{n}{n}\right\} \\ \Rightarrow \left(1+\frac{1}{n}\right)^n &\leq e \end{aligned}$$

Aligning these two inequalities, we have

$$\left(1+\frac{1}{n}\right)^n \leq e \quad \leq \left(1+\frac{1}{n}\right)^{n+1}$$

Notice now that, on the right-most inequality

$$\begin{aligned} e &\leq \left(1+\frac{1}{n}\right)^{n+1} \\ \Rightarrow \frac{e}{1+\frac{1}{n}} &\leq \left(1+\frac{1}{n}\right)^n \end{aligned}$$

The left-most inequality shows that $\left(1+\frac{1}{n}\right)^n < e$, so we may construct the inequality that

$$\frac{e}{1+\frac{1}{n}} \leq \left(1+\frac{1}{n}\right)^n \quad \leq e$$

Finally, taking the limit as n goes to ∞

$$\begin{aligned}
\lim_{n\to\infty} \frac{e}{1+\frac{1}{n}} \leq \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n &\quad \leq \lim_{n\to\infty} e \\
\Rightarrow \frac{e}{1+0} \leq \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n &\quad \leq e \\
\Rightarrow \frac{e}{1} \leq \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n &\quad \leq e \\
\Rightarrow e \leq \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n &\quad \leq e
\end{aligned}$$

Since $\lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n$ is bounded on both sides by e, $\lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = e$

23.2 References

D Joyce, "e as the limit of $(1+1/n)^n$," Clark University, https://mathcs.clarku.edu/~djoyce/ma122/elimit.pdf

Chapter 24

Logistic Regression

24.1 The Logit Transformation

Logistic regression intends to model the probability that a given response will occur based on the characteristics in the predictor variables. The response, therefore, is a probability; a value between zero and one. This is in contrast to typical linear regression where the response lies range of real numbers. This is further complicated by the fact that in the observed data, a subject does not have a probabilistic response, but the response is the dichotomous occurrence of an event. The nature of the response variable in logistic regression, therefore, necessitates that a transformation be applied.

24.1.1 Obtaining the Logit Transformation

Let us call the response for our logistic model the probability p that an event will occur. Since p is a probability, by definition, it's domain is from 0 to 1. Ideally, we would like to have a response whose domain is $\Re$. First, let us consider the transformation $\frac{p}{1-p}$ (also called the odds of p) and it's limits as p approaches 0 and 1.

$$\lim_{p \to 0} \frac{p}{1-p} = \frac{0}{1-0} = 0$$

and

$$\lim_{p \to 1} \frac{p}{1-p} = \infty$$

So the domain of $\frac{p}{1-p}$ is $(0, \infty)$. This is handy, as we do know that the ln function takes a variable on the domain $(0, \infty)$ and maps it onto the range $(-\infty, \infty)$. Thus, the equation for our logistic model the transformation (called the logit, or log-odds):

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

24.2 Retrieving the Modelled Probability

While the logit transformation allows us to perform the logistic regression, the resulting measure tells us about the risk of an event associated with a predictor, but does not tell us directly about the probability of the event occuring. If we need to know the probability of the event ocurring, we must back transform the results of our regression equation. Essentially, we extract $\hat{p}$ from the modelled log-odds. This is done as by:

$$
\begin{aligned}
\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \\
{}^{[1]} &= B \\
\Rightarrow \frac{\hat{p}}{1-\hat{p}} &= \exp(B) \\
\Rightarrow \hat{p} &= (1-\hat{p})\exp(B) \\
\Rightarrow \hat{p} &= \exp(B) - \hat{p}\cdot\exp(B) \\
\Rightarrow \hat{p} + \hat{p}\cdot\exp(B) &= \exp(B) \\
\Rightarrow \hat{p}(1+\exp(B)) &= \exp(B) \\
\Rightarrow \hat{p} &= \frac{\exp(B)}{1+\exp(B)} \\
\Rightarrow \hat{p} &= \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}{1+\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}
\end{aligned}
$$

1. Let $B = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$

Chapter 25

Mantel-Haenszel Test of Linear Trend

The Mantel-Haenszel Test is a method for testing independence of categorical variables on an ordinal scale. See Agresti (1996) for more discussion.

Let X be a categorical variable of ordinal type with R levels.

Let Y be a categorical variable of ordinal type with C levels.

Suppose we take a sample of size n and take a measurement on each item in the sample with respect to X and Y. The presence of a progresive between X and Y can be tested using the correlation coefficient ρ (Mantel 1963). We may begin by taking the estimate of ρ

$$
\begin{aligned}
r &= \frac{\widehat{Cov}(X,Y)}{\sqrt{s_X^2 s_Y^2}} \\
&= \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}(x_i-\bar{x})(y_j-\bar{y})p(x_i,y_j)}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2 p(x_i)\sum\limits_{j=1}^{n}(y_j-\bar{y})^2 p(y_j)}}
\end{aligned}
$$

But since X and Y are categorical, we cannot sensibly perform any of the operations. Instead, we define the variables U and V to be the ordinal scoring of X and Y respecitively. In other words, U_i is the score for the category of X_i and V_i is the score for the category of Y_i. Using this replacement we get

$$
r = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}(u_i-\bar{u})(v_j-\bar{v})p(u_i,v_j)}{\sqrt{\sum\limits_{i=1}^{n}(u_i-\bar{u})^2 p(u_i)\sum\limits_{j=1}^{n}(v_j-\bar{v})^2 p(v_j)}}
$$

To obtain the values of $\bar{u}$ and $\bar{v}$, we consider the following table. Recall that there are R levels of the variable X and C levels of the variable Y.

u_1

In the table, $n_{rc},\ r = 1,2,\dots,R,\ c = 1,2,\dots,C$ is the number of observations in the sample with scores r and c. From the table we can understand the marginal distributions of U and V, and we see that for $r = 1,2,\dots,R,\ c = 1,2,\dots,C$

	Category	of	V			
Category of U	1	2	...	C	Total	U
1	$n_{1,1}$	$n_{1,2}$	...	$n_{1,c}$	$n_{1,+}$	U_1
2	$n_{2,1}$	$n_{2,2}$	...	$n_{2,c}$	$n_{2,+}$	U_2
...	...	...	...	...	...	...
R	$n_{r,1}$	$n_{r,2}$	...	$n_{r,c}$	$n_{r,+}$	U_r
Total	$n_{+,1}$	$n_{+,2}$	...	$n_{+,c}$	$n_{+,+}$	
V	V_1	V_2	...	V_c		

$$p(u_r) = \frac{n_{r+}}{n}$$

$$p(v_c) = \frac{n_{+c}}{n}$$

$$p(u_r, v_c) = \frac{n_{rc}}{n}$$

$$\bar{u} = \sum_{r=1}^{R} u_i \frac{n_{r+}}{n}$$

$$\bar{v} = \sum_{c=1}^{C} v_i \frac{n_{+c}}{n}$$

With these observations, we can derive the value of r as

$$
\begin{aligned}
r &= \frac{\widehat{Cov}(U,V)}{\sqrt{s_U^2 s_V^2}}\\
&= \frac{\dfrac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}(u_r-\bar{u})(v_c-\bar{v})n_{rc}}{n-1}}{\sqrt{\dfrac{\sum\limits_{r=1}^{R}(u_r-\bar{u})^2}{n-1}\,\dfrac{\sum\limits_{c=1}^{C}(v_c-\bar{v})^2}{n-1}}}\\
&= \frac{\frac{1}{n-1}\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}(u_r-\bar{u})(v_c-\bar{v})n_{rc}}{\frac{1}{n-1}\sqrt{\sum\limits_{r=1}^{R}(u_r-\bar{u})^2\sum\limits_{c=1}^{C}(v_c-\bar{v})^2}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}(u_r-\bar{u})(v_c-\bar{v})n_{rc}}{\sqrt{\sum\limits_{r=1}^{R}(u_r-\bar{u})^2\sum\limits_{c=1}^{C}(v_c-\bar{v})^2}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}(u_rv_c-u_r\bar{v}-\bar{u}v_c+\bar{u}\bar{v})n_{rc}}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}(u_rv_cn_{rc}-u_r\bar{v}n_{rc}-\bar{u}v_cn_{rc}+\bar{u}\bar{v}n_{rc})}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rv_cn_{rc}-\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_r\bar{v}n_{rc}-\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}\bar{u}v_cn_{rc}+\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}\bar{u}\bar{v}n_{rc}}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rv_cn_{rc}-\bar{v}\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rn_{rc}-\bar{u}\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}v_cn_{rc}+\bar{u}\bar{v}\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}n_{rc}}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rv_cn_{rc}-\bar{v}\sum\limits_{r=1}^{R}u_rn_{r+}-\bar{u}\sum\limits_{c=1}^{C}v_cn_{+c}+\bar{u}\bar{v}n}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rv_cn_{rc}-\dfrac{\sum\limits_{c=1}^{C}v_cn_{+c}\sum\limits_{r=1}^{R}u_rn_{r+}}{n}-\dfrac{\sum\limits_{r=1}^{R}u_rn_{r+}\sum\limits_{c=1}^{C}v_cn_{+c}}{n}+n\dfrac{\sum\limits_{r=1}^{R}u_rn_{r+}\sum\limits_{c=1}^{C}v_cn_{+c}}{n^2}}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}\\
&= \frac{\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rv_cn_{rc}-\dfrac{2\sum\limits_{r=1}^{R}u_rn_{r+}\sum\limits_{c=1}^{C}v_cn_{+c}}{n}+\dfrac{\sum\limits_{r=1}^{R}u_rn_{r+}\sum\limits_{c=1}^{C}v_cn_{+c}}{n}}{\sqrt{\left(\sum\limits_{r=1}^{R}u_r^2n_{r+}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)^2\right)\left(\sum\limits_{c=1}^{C}v_c^2n_{+c}-\frac{1}{n}\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)^2\right)}}
\end{aligned}
$$

$$
\sum\limits_{r=1}^{R}\sum\limits_{c=1}^{C}u_rv_cn_{rc}-\frac{1}{n}\Big(\sum\limits_{r=1}^{R}u_rn_{r+}\Big)\Big(\sum\limits_{c=1}^{C}v_cn_{+c}\Big)
$$

Chapter 26

McNemar Test

This chapter only represents work that needed to be done for a specific application. Some of the formulas and equations provided are not necessarily coherent with the articles originally published on the topic. The majority of the work was derived from Connor's 1987 paper. This chapter could benefit from a great deal of improvement and additional explanation.

The McNemar Test compares proportions of related samples in which the outcome for each sample is a binary response. Th response is the same in each sample. Related samples may mean subjects from one sample are matched with subjects with similiar qualities (subjects are related, but outcomes are not); or it may mean that subjects are paired with themselves, as in a pre-post design (outcomes are related because they are taken on the same subject).

The table demonstrates the possible outcomes of such a experiment. Suppose Y_i deontes the outcome of Trial 1 and Y_j denotes the outcome of Trial 2. Y_{ij} denotes the outcome of the first and second trials, that is $Y_{ij} = Y_i \cap Y_j$. Then:

$$
\begin{aligned}
p_{11} &= P(Y_i = 1 \cap Y_j = 1) \\
p_{10} &= P(Y_i = 1 \cap Y_j = 0) \\
p_{01} &= P(Y_i = 0 \cap Y_j = 1) \\
p_{00} &= P(Y_i = 0 \cap Y_j = 0)
\end{aligned}
$$

Furthermore

$$
\begin{aligned}
p_1 = p_{11} + p_{10} &\quad = P(Y_i = 1) \\
p_2 = p_{11} + p_{01} &\quad = P(Y_j = 1)
\end{aligned}
$$

In upcoming sections, the values of the difference and sum of p_1 and p_2 will be important, so we define

		Trial 2		
		1	0	
Trial 1	1	$p_{1,1}$	$p_{1,0}$	p_1
	0	$p_{0,1}$	$p_{0,0}$	$1 - p_1$
		p_2	$1 - p_2$	

$$\delta = p_1 - p_2 \quad = (p_{11} + p_{10}) - (p_{11} + p_{01}) = p_{10} - p_{01}$$

$$\psi = p_1 + p_2 \quad = (p_{11} + p_{10}) + (p_{11} + p_{01}) = 2p_{11} + p_{10} + p_{01}$$

26.1 Sample Size Calculations for Paired Design

Three methods of calculating power for McNemar's Test have been presented. Miettenen proposed a method of estimating the power in 1968. Duffy provided the exact power in 1984. Connor provided an additional method of estimating the power in 1987.

26.1.1 Miettenen's Sample Size Calculation

Miettenen was the first to provide a popular power calculation for McNemar's test with a paired-design. Duffy would later show that this calculation tends to under-estimate the power. Subsequently, sample sizes derived from this calculation are generally lower than is needed to obtain the designed power.

Let α be the probability of Type I Error, and let β be the probability of Type II Error. Furthermore, let $Z_\alpha = \Phi(1-\alpha)$ and $Z_\beta = \Phi(1-\beta)$. Now suppose we wish to determine the sample size n_m (for Miettenen method) required to find a change in proportion from p_1 to p_2 with significance α and power $1-\beta$. The required sample size is calculated by:

$$n_m = \frac{\left(Z_\alpha \psi^{1/2} + Z_\beta \left(\psi - \delta^2(3+\psi)/(4\psi)\right)^{1/2}\right)^2}{\delta^2}$$

26.1.2 Connor's Sample Size Calculation

Connor proposed a method for sample size calculation in addition to Miettenen's. Connor's method tends to over-estimate the power. Subsequently, sample sizes derived from this calculation are generally higher than is actually needed to obtain the designed power.

Let α be the probability of Type I Error, and let β be the probability of Type II Error. Furthermore, let $Z_\alpha = \Phi(1-\alpha)$ and $Z_\beta = \Phi(1-\beta)$. Now suppose we wish to determine the sample size n_c (for Miettenen method) required to find a change in proportion from p_1 to p_2 with significance α and power $1-\beta$. The required sample size is calculated by:

$$n_c = \frac{\left(Z_\alpha \psi^{1/2} + Z_\beta (\psi - \delta^2)^{1/2}\right)^2}{\delta^2}$$

26.2 Power Calculation for Paired Design

The following power calculations are derived in a backward fashion. In the application I had at the time, I needed to calculate sample sizes, and also wanted to allow functionality in my R function to obtain power with a supplied sample size. Since I had the sample size equations, I solved for the power. Normally this would be done the other way around, ie, take the power function and solve for n. In the future, this should be revised appropriately.

26.2.1 Power Calculation for Miettenen Method

Let α be the probability of Type I Error, and let β be the probability of Type II Error. Furthermore, let $Z_\alpha = \Phi(1-\alpha)$ and $Z_\beta = \Phi(1-\beta)$. The power function can be found from the sample size equation by:

$$\begin{aligned}
n_m &= \frac{\left(Z_\alpha \psi^{1/2} + Z_\beta\big(\psi - \delta^2(3+\psi)/(4\psi)\big)^{1/2}\right)^2}{\delta^2} \\
\Rightarrow n_m\delta^2 &= \left(Z_\alpha \psi^{1/2} + Z_\beta\big(\psi - \delta^2(3+\psi)/(4\psi)\big)^{1/2}\right)^2 \\
\Rightarrow \sqrt{n_m}\delta &= Z_\alpha \psi^{1/2} + Z_\beta\big(\psi - \delta^2(3+\psi)/(4\psi)\big)^{1/2} \\
\Rightarrow \sqrt{n_m}\delta - Z_\alpha \psi^{1/2} &= Z_\beta\big(\psi - \delta^2(3+\psi)/(4\psi)\big)^{1/2} \\
\Rightarrow \frac{\sqrt{n_m}\delta - Z_\alpha \psi^{1/2}}{\big(\psi - \delta^2(3+\psi)/(4\psi)\big)^{1/2}} &= Z_\beta \\
\Rightarrow \Phi^{-1}\left(\frac{\sqrt{n_m}\delta - Z_\alpha \psi^{1/2}}{\big(\psi - \delta^2(3+\psi)/(4\psi)\big)^{1/2}}\right) &= 1-\beta
\end{aligned}$$

26.2.2 Power Calculation for Connor Method

Let α be the probability of Type I Error, and let β be the probability of Type II Error. Furthermore, let $Z_\alpha = \Phi(1-\alpha)$ and $Z_\beta = \Phi(1-\beta)$. The power function can be found from the sample size equation by:

$$\begin{aligned}
n_c &= \frac{\big(Z_\alpha \psi^{1/2} + Z_\beta(\psi - \delta^2)^{1/2}\big)^2}{\delta^2} \\
\Rightarrow n_c\delta^2 &= \big(Z_\alpha \psi^{1/2} + Z_\beta(\psi - \delta^2)^{1/2}\big)^2 \\
\Rightarrow \sqrt{n_c\delta} &= Z_\alpha \psi^{1/2} + Z_\beta(\psi - \delta^2)^{1/2} \\
\Rightarrow \sqrt{n_c\delta} - Z_\alpha \psi^{1/2} &= Z_\beta(\psi - \delta^2)^{1/2} \\
\Rightarrow \frac{\sqrt{n_c\delta} - Z_\alpha \psi^{1/2}}{(\psi - \delta^2)^{1/2}} &= Z_\beta \\
\Rightarrow \Phi^{-1}\left(\frac{\sqrt{n_c\delta} - Z_\alpha \psi^{1/2}}{(\psi - \delta^2)^{1/2}}\right) &= 1-\beta
\end{aligned}$$

Chapter 27

Moments and Moment Generating Functions

27.1 Definitions of Moments

27.1.1 Definition: General Definition of Moments

The k^{th} moment of a random variable X about some point c is defined to be $E[(X-c)^k]$.

There are two moments that are of particular use in statistics. First, the moment of X about the origin; second, the moment of X about the mean.

27.1.2 Definition: Ordinary Moments

The k^{th} moment of a random variable X about the origin is defined to be $E[(X-0)^k] = E(X^k)$.

27.1.3 Definition: Central Moments

The k^{th} moment of a random variable X about the mean μ is defined to be $E[(X-\mu)^k]$.

Using these definitions we can derive the first three central moments as follows:

$$
\begin{aligned}
E[(X-\mu)^1] &= E(X-\mu) \\
&= E(X)-\mu \\
&= E(X)-E(X)
\end{aligned}
$$

$$
\begin{aligned}
E[(X-\mu)^2] &= E[(X-\mu)(X-\mu)] \\
&= E(X^2-\mu X-\mu X+\mu^2) \\
&= E(X^2-2\mu X+\mu^2) \\
&= E(X^2)-E(2\mu X)+E(\mu^2) \\
&= E(X^2)-2\mu E(X)+\mu^2 \\
&= E(X^2)-2\mu\cdot\mu+\mu^2 \\
&= E(X^2)-2\mu^2+\mu^2 \\
&= E(X^2)-\mu^2 \\
&= E(X^2)-E(X)^2
\end{aligned}
$$

$$
\begin{aligned}
E[(X-\mu)^3] &= E[(X-\mu)(X-\mu)(X-\mu)] \\
&= E[(X^2-2\mu X+\mu^2)(X-\mu)] \\
&= E(X^3-\mu X^2-2\mu X^2+2\mu^2X+\mu^2X+\mu^3) \\
&= E(X^3-3\mu X^2+3\mu^2X-\mu^3) \\
&= E(X^3)-E(3\mu X^2)+E(3\mu^2X)-E(\mu^3) \\
&= E(X^3)-3\mu E(X^2)+3\mu^2E(X)-\mu^3 \\
&= E(X^3)-3\mu E(X^2)+3\mu^3-\mu^3 \\
&= E(X^3)-3\mu E(X^2)+2\mu^3
\end{aligned}
$$

It should be noticed that with all of these results, the moment about the mean can be evaluated by finding the ordinary moments. Thus, if we can find a consistent way to generate ordinary moments , we may use these results to find various parameters of a distribution.

27.2 Moment Generating Functions

27.2.1 Definition: Moment Generating Function

The moment generating function of a random variable, denoted $M_X(t)$, is defined to be:

$$M_X(t)=E(e^{tX})$$

The moment generating function of X is said to exist if for any positive constant c, $M_X(t)$ is finite for $|t|<c$. The definition can be expanded to

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{i=1}^{\infty} e^{tx_i} p(x_i) \\
{}^{[1]} &= \sum_{i=1}^{\infty} [\frac{(tx_i)^0}{0!} + \frac{(tx_i)^1}{1!} + \frac{(tx_i)^2}{2!} + \frac{(tx_i)^3}{3!} + \cdots] p(x_i) \\
&= \sum_{i=1}^{\infty} [1 + tx_i + \frac{(tx_i)^2}{2!} + \frac{(tx_i)^3}{3!} + \cdots] p(x_i) \\
&= \sum_{i=1}^{\infty} [p(x_i) + tx_i p(x_i) + \frac{(tx_i)^2}{2!} p(x_i) + \frac{(tx_i)^3}{3!} p(x_i) + \cdots] \\
&= \sum_{i=1}^{\infty} p(x_i) + \sum_{i=1}^{\infty} tx_i p(x_i) + \sum_{i=1}^{\infty} \frac{(tx_i)^2}{2!} p(x_i) + \sum_{i=1}^{\infty} \frac{(tx_i)^3}{3!} p(x_i) + \cdots \\
&= \sum_{i=1}^{\infty} p(x_i) + t \sum_{i=1}^{\infty} x_i p(x_i) + \frac{t^2}{2!} \sum_{i=1}^{\infty} x_i^2 p(x_i) + \frac{t^3}{3!} \sum_{i=1}^{\infty} x_i^3 p(x_i) + \cdots \\
&= 1 + tE(X) + \frac{t^2}{2!} E(X^2) + \frac{t^3}{3!} E(X^3) + \cdots
\end{aligned}
$$

1. Taylor Series Expansion: $e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} \cdots = 1 + x + \frac{x^2}{2!} + \cdots$

27.2.2 Theorem: Extraction of Moments from Moment Generating Functions

Let $M_X^{(k)}(t)$ denote the k^{th} derivative of $M_X(t)$ with respect to t. Then $M_X^{(k)}(0) = E(X^k)$.

Proof:

$$\begin{aligned} M_X(t) &= 1 + tE(X) \\ &= \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!} + \cdots \end{aligned}$$

$$\begin{aligned} M_X^{(1)}(t) &= 0 + E(X) + \frac{2t}{2!}E(X^2) + \frac{3t^2}{3!}E(X^3) + \cdots \\ &= E(X) + tE(X^2) + \frac{t^2}{2!}E(X^3) + \cdots \end{aligned}$$

$$\begin{aligned} M_X^{(2)}(t) &= 0 + E(X^2) + \frac{2t}{2!}E(X^3) + \frac{3t^2}{3!}E(X^4) + \cdots \\ &= E(X^2) + tE(X^3) + \frac{t^2}{2!}E(X^4) + \cdots \\ &\vdots \end{aligned}$$

$$\begin{aligned} M_X^{(k)}(t) &= 0 + E(X^k) + \frac{2t}{2!}E(X^{k+1}) + \frac{3t^2}{3!}E(X^{k+2}) + \cdots \\ &= E(X^k) + tE(X^{k+1}) + \frac{t^2}{2!}E(X^{k+2}) + \cdots \end{aligned}$$

$$\begin{aligned} M_X^{(1)}(0) &= 0 + E(X) + \frac{2 \cdot 0}{2!}E(X^2) + \frac{3 \cdot 0t^2}{3!}E(X^3) + \cdots \\ &= E(X) \end{aligned}$$

$$\begin{aligned} M_X^{(2)}(0) &= 0 + E(X^2) + \frac{2 \cdot 0}{2!}E(X^3) + \frac{3 \cdot 0^2}{3!}E(X^4) + \cdots \\ &= E(X^2) \end{aligned}$$

$$\vdots$$

$$\begin{aligned} M_X^{(0)}(t) &= 0 + E(X^k) + \frac{2 \cdot 0}{2!}E(X^{k+1}) + \frac{3 \cdot 0^2}{3!}E(X^{k+2}) + \cdots \\ &= E(X^k) \end{aligned}$$

Chapter 28

Multinomial Distribution

Let $E_1, E_2, \dots, E_k$ be mutually exclusive and exhaustive events and define a multinomial experiment to have the following characteristics:

i. The experiment consists of N indepedendent trials.
ii. The outcome of each trial belongs to exactly one $E_j,\ j = 1, 2, \dots, k$.
iii. The probability that an outcome belongs to event E_j is p_j.

Let $X_{ij} = \begin{cases} 1 & \text{if the outcome of the } i^{th} \text{ trial belongs to } E_j. \\ 0 & otherwise \end{cases}$

and let $n_j = \sum\limits_{i=1}^{N} X_{ij}$. Under these conditions, $N = \sum\limits_{j=1}^{k} n_j$.

By Lemma 10.0.1 the number of ways to partition N into the k events, without respect to order, is $\frac{N!}{n_1!n_2!\cdots n_k!}$. So the probabilitiy of any particular outcome of the experiment is

$$p(n_1, n_2, \dots, n_{k-1}) = \frac{N!}{n_1!n_2!\cdots n_{k-1}!n'!} p_1^{n_1} p_2^{n_2} \cdots p_{(k-1)} p'^{n'}$$

where $n' = N - n_1 - n_2 - \cdots - n_{k-1}$ and

$p' = 1 - p_1 - p_2 - \cdots - p_{k-1}$.

In other words, the entire distribution is defined by the first $k - 1$ terms.

28.1 Cumulative Distribution Function

$$P(n_1, n_2, \dots, n_{k-1}) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} \cdots \sum_{n_{k-1}=0}^{N'} \frac{N!}{n_1!n_2!\cdots n_{k-1}!n'!} p_1^{n_1} p_2^{n_2} \cdots p_{k-1}^{n_{k-1}} p'^{n'}$$

where $N' = N - n_1 - n_2 - \cdots n_{k-1}$.

28.2 Expected Values

Since this is a multivariate distribution, we discuss finding the expected values for each variate n_j as opposed to an overall mean.\

n_j is a random variable from a multinomial distribution that specifies how many of the N observations were of type j. Each of the N observations willf all into exactly one type, so we can conclude that an observation is either of type j or it isn't. Also, it is of type j with probability p_j, and each trial is independent. Thus, we may consider n_j a binomial random variable and $E(n_j) = Np_j$ and $V(n_j) = Np_j(1 - p_j)$. Now we must derive the Covariance of n_j.

We begin by defnining the random variables for $j \neq m$:

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ results in type } j. \\ 0 & otherwise \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ results in type } m. \\ 0 & otherwise \end{cases}$$

and let $n_j = \sum_{i=1}^{n} X_i$ and $n_m = \sum_{i=1}^{n} Y_i$. Since X_i and Y_i cannot simultaneously equal 1, $X_i \cdot Y_i = 0$ for all i. We thus have the following results so far:

$$E(X_i \cdot Y_i) = 0$$

$$E(X_i) = p_j$$

$$E(Y_i) = p_m$$

$Cov(X_i, Y_i) = 0$ if $i \neq j$ because the trials are independent

$Cov(X_i, Y_i) = E(X_i \cdot Y_i) - E(X_i)E(Y_i) = 0 - p_j p_m = -p_j p_m$. ($Cov(X, Y) = E(XY) - E(X)E(Y)$ (Theorem 12.2.2)

Using these results we find the Covariance of n_j and n_m.

$$\begin{aligned} Cov(n_j, n_m) &= \sum_{j=1}^{N} \sum_{m=1}^{N} Cov(X_i, Y_i) \\ &= \sum_{i=1}^{N} Cov(X_i Y_i) + \sum \sum_{i \neq j} Cov(X_i, Y_i) \\ &= \sum_{i=1}^{n} -p_j p_m + \sum \sum_{i \neq j} 0 = -np_j p_m \end{aligned}$$

The Expected Values of the p_j's can be found by

$$\begin{aligned}
E(\hat{p}_j) &= E(\frac{n_j}{N}) \\
&= \frac{1}{N}E(n_j) \\
&= \frac{1}{N}Np_j \\
&= p_j
\end{aligned}$$

$$\begin{aligned}
V(\hat{p}_j) &= V(\frac{n_j}{N}) \\
&= \frac{1}{N^2}V(n_j) \\
&= \frac{1}{N^2}Np_j(1-p_j) \\
&= \frac{p_j(1-p_j)}{N}
\end{aligned}$$

$$\begin{aligned}
Cov(\hat{p}_j, \hat{p}_m) &= Cov(\frac{n_j}{N}, \frac{n_m}{N}) \\
&= \frac{1}{N^2}Cov(n_j, n_m) \\
&= \frac{1}{N^2}(-Np_jp_m) \\
&= \frac{-p_jp_m}{N}
\end{aligned}$$

Chapter 29

Normal Distribution

29.1 Probability Distribution Function

A random variable X is said to have a Normal Distribution with parameters μ and σ^2 if its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \Re, \ \mu \in \Re, \ 0 < \sigma^2$$

29.2 Cumulative Distribution Function

The cdf of the Normal Distribution cannot be written in closed form.

It is not uncommon in statistical practice to denote the cdf of the Standard Normal Distribution, that is, the Normal distribution with $\mu = 0$ and $\sigma^2 = 1$ to be denoted by $\Phi(\cdot)$. Although it is rare, the $\Phi(\cdot)$ notation is sometimes used to denote any cumulative density function. Because such citations are rare they are usually accompanied by a statement about the distribution $\Phi(\cdot)$ represents. If no such statement is given, it is reasonably safe to assume that the function refers to the Standard Normal Distribution.

The importance of the Standard Normal Distribution is made evident by the fact that the area under any Normal curve is proportional to the distribution's variance. A proof of this is provided in Theorem **??**

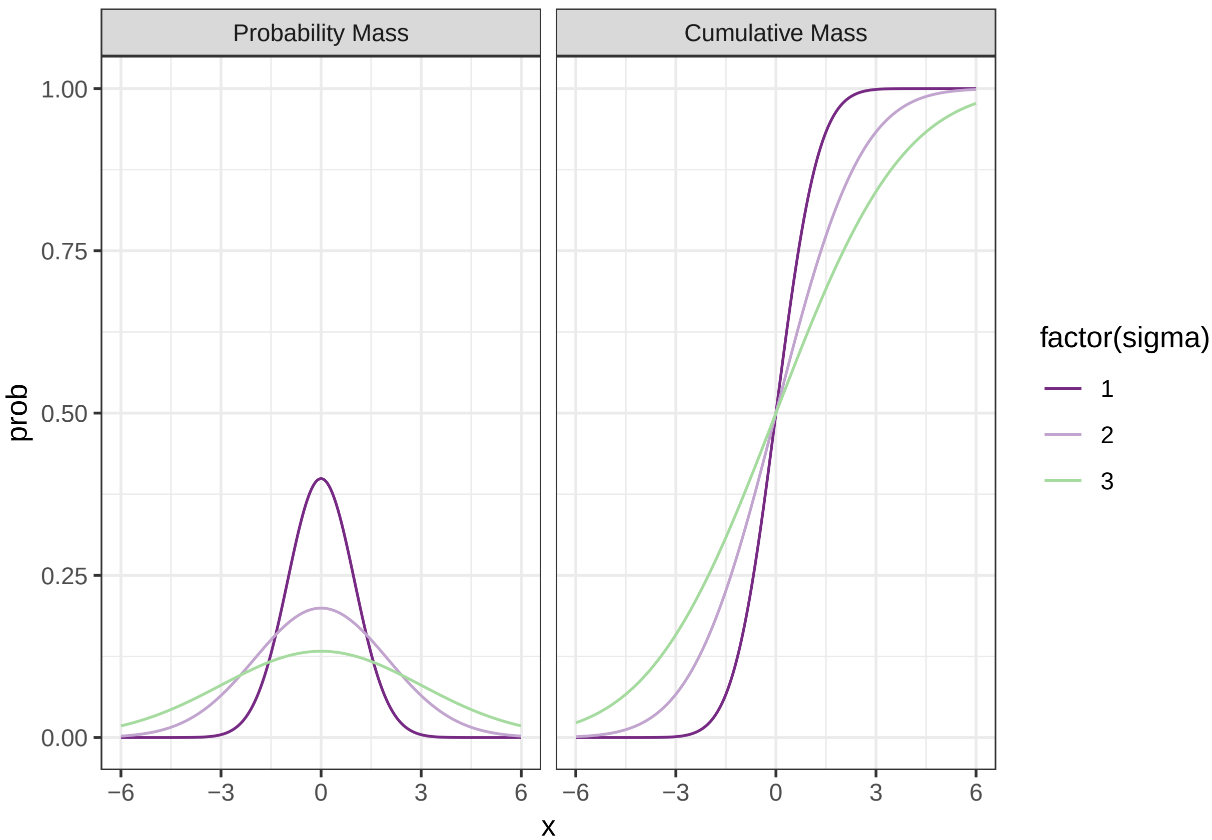

Figure 29.1: The graphs on the left show probability distribution functions, and the graphs on the right show cumulative distribution functions. The effect of changing the variance, a vertical rescaling of the distribution, is evident

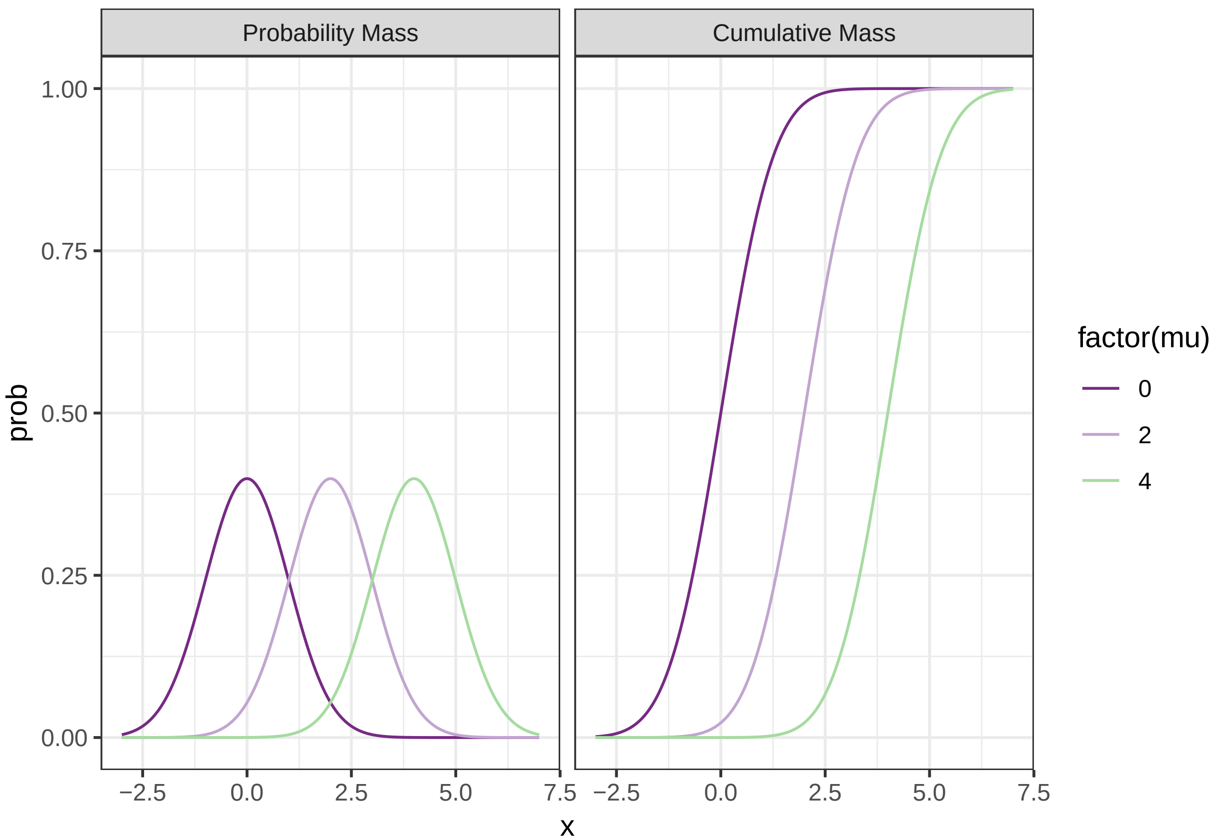

Figure 29.2: The graphs on the left show probability distribution functions, and the graphs on the right show cumulative distribution functions. The effect of changing the mean, a horizontal shift in location, is illustrated

29.3 Expected Values

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big\} \\
{}^{[1]} &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp\Big\{-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big\} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (z\sigma+\mu) \exp\Big\{-\frac{1}{2}z^2\Big\}\sigma dz \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} z\sigma^2 \exp\Big\{-\frac{1}{2}z^2\Big\} + \mu\sigma \exp\Big\{-\frac{1}{2}z^2\Big\} dz \\
&= \frac{1}{\sqrt{2\pi}\sigma} \Bigg[\int_{-\infty}^{\infty} z\sigma^2 \exp\Big\{-\frac{1}{2}z^2\Big\} dz + \int_{-\infty}^{\infty} \mu\sigma \exp\Big\{-\frac{1}{2}z^2\Big\} dz\Bigg] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \Bigg[-\sigma^2 \exp\Big\{-\frac{1}{2}z^2\Big\}\Bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \mu\sigma \exp\Big\{-\frac{1}{2}z^2\Big\} dz\Bigg] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \Bigg[-0+0+\int_{-\infty}^{\infty} \mu\sigma \exp\Big\{-\frac{1}{2}z^2\Big\} dz\Bigg] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \Bigg[\mu\sigma \int_{-\infty}^{\infty} \exp\Big\{-\frac{1}{2}z^2\Big\} dz\Bigg] \\
{}^{[2]} &= \frac{1}{\sqrt{2\pi}\sigma} \Bigg[2\mu\sigma \int_{0}^{\infty} \exp\Big\{-\frac{1}{2}z^2\Big\} dz\Bigg] \\
&= \frac{2\mu\sigma}{\sqrt{2\pi}\sigma} \Bigg[\int_{0}^{\infty} \exp\Big\{-\frac{1}{2}z^2\Big\} dz\Bigg] \\
{}^{[3]} &= \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{\infty} \frac{1}{2} u^{-\frac{1}{2}} e^{-\frac{u}{2}} du \\
&= \frac{2\mu}{2\sqrt{2\pi}} \int_{0}^{\infty} u^{-\frac{1}{2}} e^{-\frac{u}{2}} du \\
&= \frac{\mu}{\sqrt{2\pi}} \int_{0}^{\infty} u^{\frac{1}{2}-1} e^{-\frac{u}{2}} du \\
{}^{[4]} &= \frac{\mu}{\sqrt{2\pi}} 2^{\frac{1}{2}} \Gamma\Big(\frac{1}{2}\Big) \\
{}^{[5]} &= \frac{\mu\sqrt{2\pi}}{\sqrt{2\pi}} \\
&= \mu
\end{aligned}
$$

1. Let $z = \frac{x-\mu}{\sigma}$
   $\Rightarrow x = z\sigma + \mu \Rightarrow dx = \sigma dz$

2. This change can be made because the function being integrated is an even function. See Theorem **??**
3. Let $u = z^2 \Rightarrow z = u^{\frac{1}{2}} \Rightarrow dz = \frac{1}{2}u^{-\frac{1}{2}}$
4. $\int\limits_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \beta^\alpha\Gamma(\alpha)$
5. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$\begin{aligned}
E(X^2) &= \int\limits_{-\infty}^{\infty} x^2\frac{1}{\sqrt{2\pi}\sigma}\exp\Big\{\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big\}dx \\
{}^{[1]} &= \int\limits_{-\infty}^{\infty} \frac{(z\sigma+\mu)^2}{\sqrt{2\pi}\sigma}\exp\Big\{\frac{z^2}{2}\Big\}\sigma dz \\
&= \int\limits_{-\infty}^{\infty} \frac{(z\sigma+\mu)^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz \\
&= \int\limits_{-\infty}^{\infty} \frac{z^2\sigma^2+2z\sigma\mu+\mu^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz \\
&= \int\limits_{-\infty}^{\infty} \Bigg[\frac{z^2\sigma^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\} + \frac{2z\sigma\mu}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\} + \frac{\mu^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}\Bigg]dz \\
&= \int\limits_{-\infty}^{\infty} \frac{z^2\sigma^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + \int\limits_{-\infty}^{\infty} \frac{2z\sigma\mu}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + \int\limits_{-\infty}^{\infty} \frac{\mu^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz \\
&= \sigma^2\int\limits_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + 2\sigma\mu\int\limits_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + \mu^2\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz \\
{}^{[2]} &= \sigma^2\int\limits_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + 2\sigma\mu\int\limits_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + \mu^2\cdot 1 \\
{}^{[3]} &= \sigma^2\int\limits_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + 2\sigma\mu E(Z) + \mu^2 \\
&= \sigma^2\int\limits_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}}\exp\Big\{\frac{z^2}{2}\Big\}dz + 2\sigma\mu\cdot 0 + \mu^2 \\
{}^{[4]} &= \sigma^2\Bigg(\frac{z}{\sqrt{2\pi}\cdot -\exp\Big\{-\frac{z^2}{2}\Big\}} - \int\limits_{-\infty}^{\infty} -\exp\Big\{-\frac{z^2}{2}\frac{1}{\sqrt{2\pi}}dz\Bigg) + \mu^2 \\
&= \sigma^2\Bigg(-0+0+\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{z^2}{2}\Big\}dz\Bigg) + \mu^2 \\
{}^{[5]} &= \sigma^2(1) + \mu^2 \\
&= \sigma^2 + \mu^2
\end{aligned}$$

1. $z = \frac{x-\mu}{\sigma}$
   $\Rightarrow x = z\sigma + \mu \Rightarrow dx = \sigma dz$
2. Theorems **??** and **??** with $\mu = 0$ and $\sigma^2 = 1$

3. Using the previous result with $\mu = 0$ and $\sigma^2 = 1$, $E(Z) = 0$.
4. Integration by parts: $\int_a^b u \; dv = [u \cdot v]_a^b - \int_a^b v \; du$ with $u = \frac{z}{\sqrt{2\pi}} \Rightarrow du = \frac{1}{\sqrt{2\pi}} dz\backslash$ and $dv = z \exp\{-\frac{z^2}{2}\} \Rightarrow v = -\exp\{-\frac{z^2}{2}\}$
5. See footnote 2.

$$\begin{aligned} \mu &= E(X) \\ &= \mu \end{aligned}$$

$$\begin{aligned} \sigma^2 &= E(X^2) - E(X)^2 \\ &= \sigma^2 + \mu^2 - \mu^2 \\ &= \sigma^2 \end{aligned}$$

29.4 Moment Generating Function

$$
\begin{aligned}
E(e^{tX}) &= \int\limits_{-\infty}^{\infty} \exp\{tx\} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{\frac{1}{2\sigma^2} - (x-\mu)^2\Big\} dx \\
&= \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{tx - \frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big\} dx \\
{}^{[1]} &= \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{t(z\cdot\sigma + \mu - \frac{1}{2}z^2\Big\} \sigma dz \\
&= \int\limits_{-\infty}^{\infty} \frac{\sigma}{\sqrt{2\pi}\sigma} \exp\Big\{t(z\cdot\sigma + \mu) - \frac{1}{2}z^2\Big\} dz \\
&= \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\{\mu t\} \exp\Big\{z\sigma t - \frac{1}{2}z^2\Big\} dz \\
&= \exp\{\mu t\} \int\limits_{-\infty}^{\infty} \frac{\sigma}{\sqrt{2\pi}\sigma} \exp\Big\{z\sigma t - \frac{1}{2}z^2\Big\} dz \\
&= \exp\{\mu t\} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{z\sigma t - \frac{1}{2}z^2 - \frac{1}{2}t^2\sigma^2 + \frac{1}{2}t^2\sigma^2\Big\} dz \\
&= \exp\{\mu t\} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{ - \frac{1}{2}z^2 + z\sigma t - \frac{1}{2}t^2\sigma^2\Big\} \exp\Big\{\frac{1}{2}t^2\sigma^2\Big\} dz \\
&= \exp\{\mu t\} \exp\Big\{\frac{1}{2}t^2\sigma^2\Big\} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{z^2 - 2z\sigma t + t^2\sigma^2\Big\} dz \\
&= \exp\Big\{\mu t + \frac{t^2\sigma^2}{2}\Big\} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{\frac{1}{2}(z-\sigma t)^2\Big\} dz \\
{}^{[2]} &= \exp\Big\{\mu t + \frac{t^2\sigma^2}{2}\Big\} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{\frac{1}{2}(u)^2\Big\} du \\
&= \exp\Big\{\mu t + \frac{t^2\sigma^2}{2}\Big\} \cdot 1 \\
&= \exp\Big\{\mu t + \frac{t^2\sigma^2}{2}\Big\}
\end{aligned}
$$

1. $z = \frac{x-\mu}{\sigma}$
   $\Rightarrow x = z\cdot\sigma + \mu \Rightarrow dx = \sigma dz$
2. $u = z - \sigma t$
   $\Rightarrow z = u + \sigma t \Rightarrow dz = du$
3. See Theorems **??** and **??**

$$M_X^{(1)}(t) = \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \cdot (\mu + t\sigma^2)$$

$$\begin{aligned}
M_X^{(2)}(t) &= \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \cdot \sigma^2 + (\mu + t\sigma^2)\exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\}(\mu + t\sigma^2) \\
&= M_X^{(2)}(t) = \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \cdot \sigma^2 + (\mu + t\sigma^2)^2 \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\}
\end{aligned}$$

$$\begin{aligned}
E(X) &= M_X^{(1)}(0) = \exp\left\{\mu \cdot 0 + \frac{0^2\sigma^2}{2}\right\} \cdot (\mu + 0 \cdot \sigma) \\
&= e^0 \cdot (\mu + 0) \\
&= 1 \cdot \mu \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= M_X^{(2)}(0) \\
&= \exp\left\{\mu \cdot 0 + \frac{0^2\sigma^2}{2}\right\} \cdot \sigma^2 + (\mu + 0 \cdot \sigma^2)^2 \exp\left\{\mu \cdot 0 + \frac{0^2\sigma^2}{2}\right\} \\
&= e^0\sigma^2 + (\mu + 0)^2 e^0 \\
&= 1 \cdot \sigma^2 + \mu^2 \cdot 1 \\
&= \sigma^2 + \mu^2
\end{aligned}$$

$$\begin{aligned}
\mu &= E(X) \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= E(X^2) - E(X) \\
&= \sigma^2 + \mu^2 - \mu^2 \\
&= \sigma^2
\end{aligned}$$

29.5 Maximum Likelihood Estimators

Let $x_1, x_2, \ldots, x_n$ be a random sample from a Normal distribution with mean μ and variance σ^2.

29.5.1 Likelihood Function

$$
\begin{aligned}
L(x_1, x_2, \ldots, x_n|\theta) &= f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta) \\
&= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x_1-\mu)^2}{2\sigma^2}\right\}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x_2-\mu)^2}{2\sigma^2}\right\}\cdots\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x_n-\mu)^2}{2\sigma^2}\right\} \\
&= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\prod_{i=1}^{n}\exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{\sum_{i=1}^{n}\frac{-(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}-(x_i-\mu)^2\right\}
\end{aligned}
$$

29.5.2 Log-likelihood Function

$$
\begin{aligned}
\ell(\theta) &= \ln\big(L(\theta)\big) \\
&= \ln\left(\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}-(x_i-\mu)^2\right\}\right) \\
&= \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n + \ln\left(\exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}-(x_i-\mu)^2\right\}\right) \\
&= n\ln\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}-(x_i-\mu)^2 \\
&= n\ln\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2) \\
&= n\ln\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}x_i^2 - 2\mu\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\mu^2\right] \\
&= n\ln\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{2\mu}{2\sigma^2}\sum_{i=1}^{n}x_i - \frac{1}{2\sigma^2}n\mu^2 \\
&= n\ln\left(\frac{1}{\sqrt{2\pi}}\sigma^{-1}\right) - \frac{\sigma^{-2}}{2}\sum_{i=1}^{n}x_i^2 + \mu\sigma^{-2}\sum_{i=1}^{n}x_i - \frac{n\mu^2\sigma^{-2}}{2}
\end{aligned}
$$

29.5.3 MLE for μ

$$\begin{aligned}\frac{d\ell}{d\mu} &= 0 - 0 + \sigma^{-2}\sum_{i=1}^{n} - \frac{2n\mu\sigma^{-2}}{2} \\ &= \sigma^{-2}\sum_{i=1}^{n} - n\mu\sigma^{-2}\end{aligned}$$

$$\begin{aligned}0 &= \sigma^{-2}\sum_{i=1}^{n} - n\mu\sigma^{-2} \\ \Rightarrow n\mu\sigma^{-2} &= \sigma^{-2}\sum_{i=1}^{n} \\ \Rightarrow n\mu\ &= \sum_{i=1}^{n} \\ \Rightarrow \mu &= \frac{1}{n}\sum_{i=1}^{n}\end{aligned}$$

So $\hat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n}$ is the maximum likelihood estimator for μ.

29.5.4 MLE for σ^2

The work in deriving the MLE for σ^2 is greatly reduced if we use

$$\begin{aligned}
\ell(\theta) &= n \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{n} -(x_i-\mu)^2 \\
&= n \ln \left(\frac{1}{\sqrt{2\pi}}\sigma^{-1}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} -(x_i-\mu)^2
\end{aligned}$$

$$\begin{aligned}
\frac{d\ell}{d\sigma} &= n\left(\frac{1}{\frac{1}{\sqrt{2\pi}}\sigma^{-1}}\right)\left(\frac{-1}{\sqrt{2\pi}\sigma^{-2}}\right) - \frac{-2}{2}\sigma^{-3}\sum_{i=1}^{n} -(x_i-\mu)^2 \\
&= \frac{-n\sqrt{2\pi}\sigma}{\sqrt{2\pi}\sigma^2} + \frac{1}{\sigma^3}\sum_{i=1}^{n} -(x_i-\mu)^2 \\
&= \frac{-n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} -(x_i-\mu)^2
\end{aligned}$$

$$\begin{aligned}
0 &= \frac{-n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} -(x_i-\mu)^2 \\
\Rightarrow \frac{n}{\sigma} &= \frac{1}{\sigma}\sum_{i=1}^{n} -(x_i-\mu)^2 \\
\Rightarrow \frac{\sigma^3}{\sigma} &= \frac{1}{n}\sum_{i=1}^{n} -(x_i-\mu)^2 \\
\Rightarrow \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} -(x_i-\mu)^2
\end{aligned}$$

So $\hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n} -(x_i-\mu)^2$ is the maximum likelihood estimator for σ^2. Notice however that this MLE is a biased estimator[1].

29.6 Theorems for the Normal Distribution

29.6.1 Validity of the Distribution (Polar Coordinates)

$$\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 1$$

Proof:

Let $A = \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$.

By using the identity transformation $y = x$, we may also write

[1] See **??**

$$
\begin{aligned}
A &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy \\
\Rightarrow A^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dx\ dy \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2} - \frac{(y-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu)^2+(y-\mu)^2]} dx\ dy \\
[1] &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} t^2+u^2} dt\ du \\
[2] &= \frac{1}{2\pi\sigma^2} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2\sigma^2}} r\ dr\ d\theta \\
&= \frac{1}{2\pi\sigma^2} \int_{0}^{2\pi} -\sigma^2 e^{-\frac{r^2}{2\sigma^2}} \Big|_0^{\infty} d\theta \\
&= \frac{1}{2\pi\sigma^2} \int_{0}^{2\pi} 0 + \sigma^2\ d\theta \\
&= \frac{1}{2\pi\sigma^2} \int_{0}^{2\pi} \sigma^2\ d\theta \\
&= \frac{1}{2\pi\sigma^2} \cdot \left(\sigma^2\theta \Big|_0^{2\pi} \right) \\
&= \frac{1}{2\pi\sigma^2} [2\pi\sigma^2 - 0] \\
&= \frac{2\pi\sigma^2}{2\pi\sigma^2} \\
&= 1 \\
\Rightarrow A^2 &= 1 \\
\Rightarrow A &= 1
\end{aligned}
$$

1. Substitute $t = x - \mu$ and $u = y - \mu$
2. $t = r\cos\theta$ and $u = r\sin\theta$.\ Thus $t^2 + u^2 = r^2\cos^2\theta + r^2\sin^2\theta = r^2(\cos^2\theta + \sin^2\theta) = r^2$

Thus $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 1$

29.6.2 Validity of the Distribution (Gamma Function)

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 1$$

Proof:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
[1] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\frac{-(y)^2}{2\sigma^2}} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy \\
[2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{u}{2}} \frac{1}{2}\sigma u^{-\frac{1}{2}} du \\
&= \frac{\sigma}{2\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} u^{-\frac{1}{2}} e^{-\frac{u}{2}} du \\
[3] &= \frac{2\sigma}{2\sqrt{2\pi}\sigma} \int_{0}^{\infty} u^{\frac{1}{2}-1} e^{-\frac{u}{2}} du \\
[4] &= \frac{1}{\sqrt{2\pi}} 2^{\frac{1}{2}} \Gamma\big(\frac{1}{2}\big) \\
[5] &= \frac{\sqrt{2\pi}}{\sqrt{2\pi}} \\
&= 1
\end{aligned}$$

1. $y = x - \mu \;\Rightarrow x = y + \mu \;\Rightarrow dx = dy$
2. $u = (\frac{y}{\sigma})^2 \;\Rightarrow y = \sigma u^{\frac{1}{2}} \;\Rightarrow dy = \frac{1}{2}\sigma u^{-\frac{1}{2}}$
3. If $f(x)$ is an even function then $\int_{-\infty}^{\infty} f(x)dx = 2\int_{0}^{\infty} f(x)dx$ (Theorem **??**).
4. $\int_{0}^{\infty} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^{\alpha}\Gamma(\alpha)$
5. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

29.6.3 Multiple of a Normal Random Variable

Let X be a Normal random variable with parameters μ and σ^2, and let c be a constant. If $Y = cX$, then $Y \sim Normal(c\mu,\ c^2\sigma^2)$.

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{tcX}) \\
&= \exp\left\{\mu tc + \frac{t^2c^2\sigma^2}{2}\right\} \\
&= \exp\left\{c\mu t + \frac{t^2c^2\sigma^2}{2}\right\}
\end{aligned}$$

Which is the Moment Generating Function of a Normal random variable with mean $c\mu$ and variance $c^2\sigma^2$. Thus $Y \sim Normal(c\mu,\ c^2\sigma^2)$.

29.6.4 Weighted Sum of Normal Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent random variables from Normal distributions with parameters μ_i and σ_i^2, i.e. $X_i \sim Normal(\mu_i, \sigma_i^2),\ i = 1, 2, \ldots, n$, and let $a_1, a_2, \ldots, a_n$ be constants.

If $Y = \sum\limits_{i=1}^{n} a_i X_i$, then $Y \sim Normal\big(\sum\limits_{i=1}^{n} a_i\mu_i,\ \sum\limits_{i=1}^{n} a_i^2\sigma_i^2\big)$.

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{t(a_1X_1 + a_2X_2 + \cdots + a_nX_n)}) \\
&= E(e^{ta_1X_1} e^{ta_2X_2} \cdots e^{ta_nX_n}) \\
&= E(e^{ta_1X_1}) E(e^{ta_2X_2}) \cdots E(e^{ta_nX_n}) \\
&= \exp\left\{ta_1\mu_1 + \frac{ta_1^2\sigma_1^2}{2}\right\} \cdot \exp\left\{ta_2\mu_2 + \frac{ta_2^2\sigma_2^2}{2}\right\} \cdot \cdots \cdot \exp\left\{ta_n\mu_n + \frac{ta_n^2\sigma_n^2}{2}\right\} \\
&= \exp\left\{ta_1\mu_1 + \frac{ta_1^2\sigma_1^2}{2} + ta_2\mu_2 + \frac{ta_2^2\sigma_2^2}{2} + \cdots + ta_n\mu_n + \frac{ta_n^2\sigma_n^2}{2}\right\} \\
&= \exp\left\{t\sum_{i=1}^{n} a_i\mu_i + \frac{t\sum\limits_{i=1}^{n} a_i^2\sigma_i^2}{2}\right\}
\end{aligned}$$

Which is the mgf of a Normal random variable with parameters $\sum\limits_{i=1}^{n} a_i\mu_i$ and $\sum\limits_{i=1}^{n} a_i^2\sigma_i^2$. Thus $Y \sim Normal\big(\sum\limits_{i=1}^{n} a_i\mu_i,\ \sum\limits_{i=1}^{n} a_i^2\sigma_i^2\big)$.

29.6.5 Sum of Normal Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent random variables from Normal distributions with parameters μ_i and σ_i^2, i.e. $X_i \sim Normal(\mu_i, \sigma_i^2),\ i = 1, 2, \ldots, n$. Let $Y = \sum\limits_{i=1}^{n} X_i$. Then $Y \sim Normal\big(\sum\limits_{i=1}^{n} \mu_i,\ \sum\limits_{i=1}^{n} \sigma_i^2\big)$.

Proof:

$Y = \sum\limits_{i=1}^{n} X_i$ is a special case of the Weighted Sum of Normal Random Variables where each of the weights is equal to 1. It follows then that $Y \sim Normal\big(\sum\limits_{i=1}^{n} 1\mu_i,\ \sum\limits_{i=1}^{n} 1^2\sigma_i^2\big)$ which is equivalent to saying $Y \sim Normal\big(\sum\limits_{i=1}^{n} \mu_i,\ \sum\limits_{i=1}^{n} \sigma_i^2\big)$.

29.6.6 Standard Normal Transformation

Let X be a Normal random variable with mean μ and variance σ^2. That is, $X \sim$Normal(μ, σ^2). If $Z = \frac{X-\mu}{\sigma}$, then $Z \sim$Normal(0,1), and we say Z has a *Standard Normal* distribution.

Furthermore, if $Z \sim$Normal(0,1), and $X = Z \cdot \sigma + \mu$, then $X \sim Normal(\mu, \sigma^2)$.

Proof:

First, since $X \sim Normal(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$

$$\begin{aligned}
E(e^{tZ}) &= E\Big(\exp\Big\{t\frac{x-\mu}{\sigma}\Big\}\Big) \\
&= E\Big(\exp\Big\{\frac{tx - t\mu}{\sigma}\Big\}\Big) \\
&= E\Big(\exp\Big\{\frac{tx}{\sigma} - \frac{t\mu}{\sigma}\Big\}\Big) \\
&= E\Big(\exp\Big\{\frac{tx}{\sigma}\Big\}\exp\Big\{-\frac{t\mu}{\sigma}\Big\}\Big) \\
&= \exp\Big\{\frac{-t\mu}{\sigma}\Big\}E\Big(\exp\Big\{\frac{t}{\sigma}x\Big\}\Big) \\
&= \exp\Big\{\frac{-t\mu}{\sigma}\Big\}\exp\Big\{\frac{t\mu}{\sigma} + \frac{t^2\sigma^2}{2\sigma^2}\Big\} \\
&= \exp\Big\{\frac{-t\mu}{\sigma} - \frac{t\mu}{\sigma} + \frac{t^2\sigma^2}{2\sigma^2}\Big\} \\
&= \exp\Big\{\frac{t^2\sigma^2}{2\sigma^2}\Big\}
\end{aligned}$$

Which is the Moment Generating Function of a Normal random variable with $\mu = 0$ and $\sigma^2 = 1$. Thus $Z \sim Normal(0, 1)$.

To complete the proof, we start with $Z \sim$Normal(0,1) and let $X = Z \cdot \sigma + \mu$.

$$\begin{aligned}
E(e^{tX}) &= E(e^{t(z\sigma+\mu)}) \\
&= E(e^{tz\sigma}e^{t\mu}) \\
&= e^{t\mu}E(e^{t\sigma z}) \\
&= \exp\{t\mu\}\exp\Big\{\frac{t^2\sigma^2}{2}\Big\} \\
&= \exp\Big\{\mu t + \frac{t^2\sigma^2}{2}\Big\}
\end{aligned}$$

Which is the Moment Generating Funciton of a Normal random variable with mean μ and variance σ^2. Thus $X \sim Normal(\mu, \sigma^2)$.

29.6.7 Distribution of $\bar{X}$

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from a Normal distribution with parameters μ and σ^2, i.e. $X_i \sim Normal(\mu,\ \sigma^2)$. Let $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Then $\bar{X} \sim Normal\big(\mu,\ \frac{\sigma^2}{n}\big)$.

Proof:

$\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ is a special case of the Weighted Sum of Normal Random Variables where $a_i = \frac{1}{n},\ i = 1, 2, \ldots, n$. It follows then that

$$
\begin{aligned}
M_{\bar{X}}(t) &= exp\Big(t\sum_{i=1}^{n}\frac{1}{n}\mu + t\frac{\sum\limits_{i=1}^{n}\frac{1}{n}\sigma^2}{2}\Big) \\
&= exp(t\frac{n\mu}{n} + \frac{t\frac{n\sigma^2}{n}}{2}) \\
&= exp(t\mu + \frac{t\frac{\sigma^2}{n}}{2})
\end{aligned}
$$

Which is the mgf of a Normal random variable with parameters μ and $\frac{\sigma^2}{n}$. Thus $\bar{X} \sim Normal\big(\mu,\ \frac{\sigma^2}{n}\big)$.

29.6.8 Square of a Standard Normal Random Variable

If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2(1)$.

Proof:

$$
\begin{aligned}
M_{Z^2}(t) &= E(e^{tZ^2}) \\
&= \int_{-\infty}^{\infty} e^{tz^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tz^2}e^{-\frac{z^2}{2}}dz \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(-2t+1)}dz \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(1-2t)}dz \\
[1] &= \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{z^2}{2}(1-2t)}dz \\
[2] &= \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-u}\frac{\sqrt{2}u^{-\frac{1}{2}}}{2(1-2t)^{\frac{1}{2}}}du \\
&= \frac{2\sqrt{2}}{2\sqrt{2\pi}(1-2t)^{\frac{1}{2}}}\int_{0}^{\infty} e^{-u}u^{-\frac{1}{2}}du \\
&= \frac{2\sqrt{2}}{2\sqrt{2\pi}(1-2t)^{\frac{1}{2}}}\int_{0}^{\infty} u^{\frac{1}{2}-1}e^{-u}du \\
[3] &= \frac{1}{\sqrt{\pi}(1-2t)^{\frac{1}{2}}}\Gamma(\frac{1}{2}) \\
&= \frac{\sqrt{\pi}}{\sqrt{\pi}(1-2t)^{\frac{1}{2}}} \\
&= \frac{1}{(1-2t)^{\frac{1}{2}}} = (1-2t)^{-\frac{1}{2}}
\end{aligned}
$$

1. $\int\limits_{-\infty}^{\infty} f(x)dx = 2\int\limits_{0}^{\infty} f(x)dx$ when f(x) is an even function (**??**)

2. Let $u = \frac{z^2}{2}(1-2t) \quad \Rightarrow z = \frac{\sqrt{2}u^{\frac{1}{2}}}{(1-2t)^{\frac{1}{2}}}$

   So $dz = \frac{\sqrt{2}u^{-\frac{1}{2}}}{2(1-2t)^{\frac{1}{2}}}$

3. $\int\limits_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx = \beta^\alpha\Gamma(\alpha)$

Which is the mgf of a Chi-Square random variable with 1 degree of freedom. Thus $Z^2 \sim \chi^2(1)$.

Chapter 30

Poisson Distribution

30.1 Probability Mass Function

A random variable is said to have a Poisson distribution with parameter λ if its probability mass function is:

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, ... \\ 0 & otherwise \end{cases}$$

30.2 Cumulative Mass Function

$$P(x) = \begin{cases} e^{-\lambda} \sum\limits_{i=0}^{x} \frac{\lambda^i}{i!}, & x = 1, 2, 3, ... \\ 0 & otherwise \end{cases}$$

A recursive form of the cdf can be derived and has some usefulness in computer applications. With it, one need only initiate the first value and additional cumulative probabilities can be calculated. It is derived as follows:

$$\begin{aligned} P(X = x + 1) &= \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} \\ &= \frac{\lambda}{x+1}\frac{e^{-\lambda}\lambda^x}{x!} \\ &= \frac{\lambda}{x+1}P(X = x) \end{aligned}$$

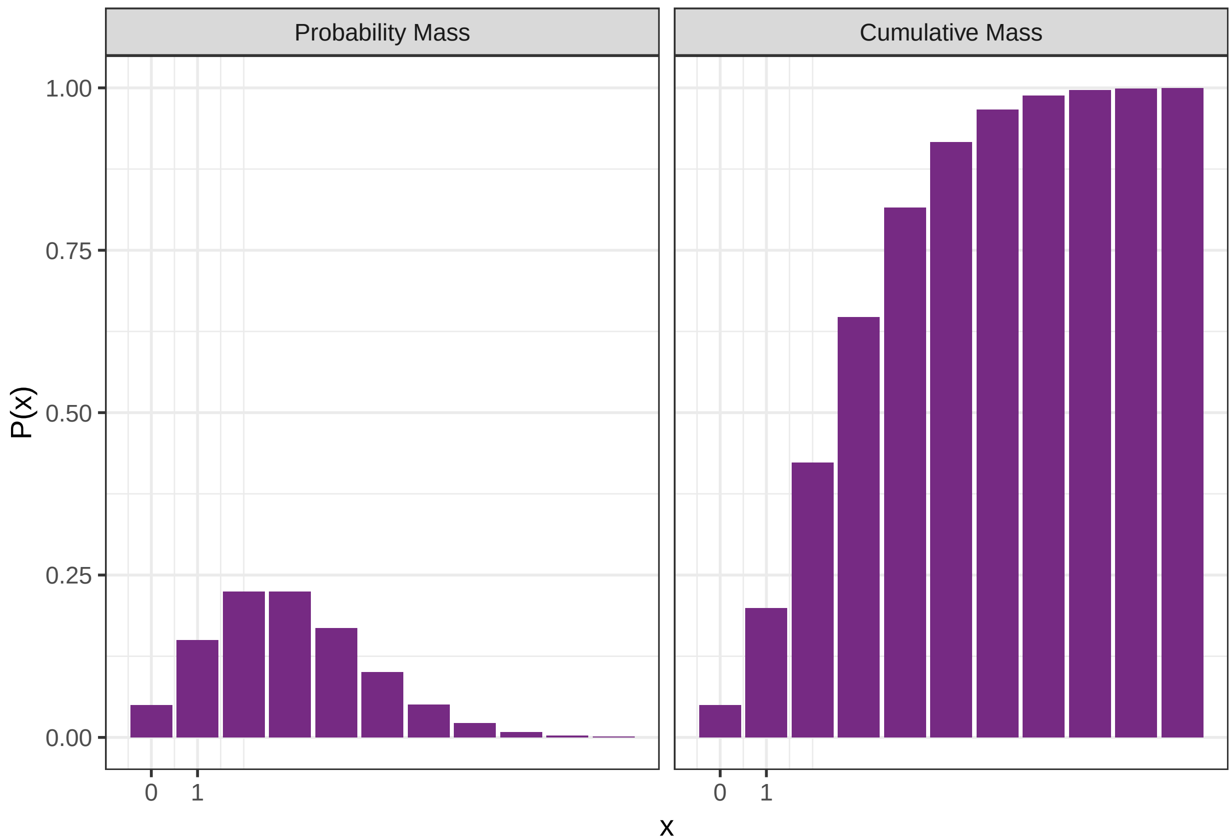

Figure 30.1: The graphs on the left and right show a Poisson probability cumulative distribution function, respectively, for $\lambda = 3$.

30.3 Expected Values

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \Big(0 \frac{\lambda^0}{0!} + 1 \frac{\lambda^1}{1!} + 2 \frac{\lambda^2}{2!} + 3 \frac{\lambda^3}{3!} + \cdots \Big) \\
&= e^{-\lambda} \Big(0 + \lambda^1 + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \cdots \Big) \\
&= \lambda e^{-\lambda} \Big(\lambda^0 + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots \Big) \\
&= \lambda e^{-\lambda} \Big(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots \Big) \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda e^(- \lambda + \lambda) \\
&= \lambda
\end{aligned}
$$

1. Taylor Series Expansion: $e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots$

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \Big(0^2 \frac{\lambda^0}{0!} + 1^2 \frac{\lambda^1}{1!} + 2^2 \frac{\lambda^2}{2!} + 3^2 \frac{\lambda^3}{3!} + \cdots \Big) \\
&= \lambda e^{-\lambda} \Big(\frac{\lambda^0}{1} + 2 \frac{\lambda^1}{1!} + 3 \frac{\lambda^2}{2!} + \cdots \Big) \\
&= \lambda e^{-\lambda} \sum_{x=0}^{\infty} (x+1) \frac{\lambda^x}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \Big(x \frac{\lambda^x}{x!} + \frac{\lambda^x}{x!} \Big) \\
&= \lambda e^{-\lambda} \Big(\sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \Big) \\
&= \lambda \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} + \lambda \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \lambda E(X) + \lambda \\
&= \lambda^2 + \lambda
\end{aligned}
$$

$$
\begin{aligned}
\mu &= E(X) \\
&= \lambda
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= E(X^2) - E(X) \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}
$$

30.4 Moment Generating Function

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=0}^{\infty} \frac{(\lambda e^{tx})^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{tx})^x}{x!} \\
&= e^{-\lambda} \Big[\frac{(\lambda e^t)^0}{0!} + \frac{(\lambda e^t)^1}{1!} + \frac{(\lambda e^t)^2}{2!} + \cdots \Big] \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{(\lambda e^t - \lambda)} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}
$$

1. Taylor Series Expansion: $e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots$

$$
M_X^{(1)}(t) = e^{\lambda(e^t-1)} \lambda e^t = \lambda e^t e^{\lambda(e^t-1)}
$$

$$
\begin{aligned}
M_X^{(2)}(t) &= (e^{\lambda(e^t-1)}(\lambda e^t) + (e^{\lambda(e^t-1)} \lambda e^t)(\lambda e^t) \\
&= \lambda e^t [e^{\lambda e^t - 1)} + e^{\lambda(e^t-1)} \lambda e^t] \\
&= \lambda e^t [e^{\lambda(e^t-1)}(1 + \lambda e^t)]
\end{aligned}
$$

$$
\begin{aligned}
E(X) &= M_X^{(1)}(0) \\
&= \lambda e^0 e^{\lambda(e^0-1)} \lambda e^t = \lambda
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= M_X^{(2)}(0) \\
&= \lambda e^0 [e^{\lambda(e^0-1)}(1 + \lambda e^0)] \\
&= \lambda e^0 [e^{\lambda(e^0-1)}(1 + \lambda e^0] \\
&= \lambda(1 + \lambda) \\
&= \lambda + \lambda^2
\end{aligned}
$$

$$
\begin{aligned}
\mu &= E(X) \\
&= \lambda \\
\sigma^2 &= E(X^2) - E(X) \\
&= \lambda + \lambda^2 - \lambda^2 \\
&= \lambda
\end{aligned}
$$

30.5 Maximum Likelihood Estimator

Let $x_1, x_2, \ldots, x_n$ be a random sample drawn from a Poisson Distribution with parameter λ.

30.5.1 Likelihood Function

$$\begin{aligned}
L(\theta) &= L(x_1, x_2, \ldots, x_n|\theta) \\
&= p(x_1|\theta)p(x_2|\theta)\cdots p(x_n|\theta) \\
&= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdot \ \cdots \ \cdot \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\
&= \frac{e^{-n\lambda}\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_1!}
\end{aligned}$$

30.5.2 Log-likelihood

$$\begin{aligned}
\ell(\lambda) &= \ln(L(\lambda)) \\
&= \ln\Big[\frac{e^{-n\lambda}\lambda^{\sum\limits_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_1!}\Big] \\
&= \ln(e^{-n\lambda}) + \ln\Big(\lambda^{\sum\limits_{i=1}^{n} x_i}\Big) - \ln\Big(\prod_{i=1}^{n} x_1!\Big) \\
&= -n\lambda + \sum_{i=1}^{n} x_i \ln\Big(\lambda\Big) - \ln(\prod_{i=1}^{n} x_1!\Big)
\end{aligned}$$

30.5.3 MLE for λ

$$\begin{aligned}\frac{d\ell}{d\lambda} &= -n - \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} - 0 \\ &= \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} - n\end{aligned}$$

$$\begin{aligned}0 &= \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} - n \\ \Rightarrow \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} &= n \\ \Rightarrow \sum_{i=1}^{n} x_i &= n\lambda \\ \Rightarrow \frac{\sum\limits_{i=1}^{n} x_i}{n} &= \lambda\end{aligned}$$

so $\hat{\lambda} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$ is the Maximum Likelihood Estimator for λ.

30.6 Theorems for the Poisson Distribution

30.6.1 Derivation of the Poisson Distribution

Suppose X is a Binomial random variable in all respects but has an infinite (non-fixed) number of trials, each with probability of success p.

Then the pdf of X is $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Proof:

For an infinite number of trials we take $\lim\limits_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x}$.

By rewriting p in terms of μ ($\mu = np \Rightarrow p = \frac{\mu}{n}$) we get

$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{x} \frac{\mu}{n}^x (1-\frac{\mu}{n})^{n-x} &= \lim_{n\to\infty} \binom{n}{x} \frac{\mu}{n}^x (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \\
&= \lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-x+1)(n-x)!}{x!(n-x)!} \right) \mu^x \frac{1}{n^x} (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \\
&= \lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-x+1)}{x!} \right) \mu^x \frac{1}{n^x} (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-x+1)}{n^x} \right) \frac{1}{n^x} (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \\
\overset{[1]}{=}& \frac{\mu^x}{x!} \lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-x+1)}{n^x} \right) \lim_{n\to\infty} \frac{1}{n^x} (1-\frac{\mu}{n})^n \lim_{n\to\infty} (1-\frac{\mu}{n})^{-x} \\
&= \frac{\mu^x}{x!} \cdot 1 \cdot e^{-\mu} \cdot 1 \\
&= \frac{e^{-\mu}\mu^x}{x!}
\end{aligned}$$

1. $\lim_{n\to\infty} (1-\frac{x}{n})^n = e^-x$

Traditionally, we use λ in place of μ for the Poisson distribution, giving us the desired result. (Wackerly, III, and Scheaffer 2002, 125)

30.6.2 Validity of the Distribution

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = 1$$

Proof:

$$\begin{aligned}
\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots \right) \\
&= e^{-\lambda} \cdot e^{\lambda} \\
&= e^0 \\
&= 1
\end{aligned}$$

1. Taylor Series Expansion: $e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots$

30.6.3 Sum of Poisson Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent random variables from a Poisson distribution with parameter $\lambda_i,\ i = 1, 2, \ldots, n$; that is, X_i ~Poisson(λ_i).

Let $Y = \sum\limits_{i=1}^{n} X_i$.

Then Y ~Poisson($\sum\limits_{i=1}^{n} \lambda_i$).\

Proof:

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{t(X_1+X_2+\cdots+X_n)}) \\
&= E(e^{tX_1}e^{tX_2}\cdots e^{tX_n}) \\
&= E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n}) \\
&= e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)}\cdots e^{\lambda_n(e^t-1)} \\
&= e^{(\lambda_1+\lambda_2+\cdots+\lambda_n)(e^t-1)} \\
&= e^{(e^{t-1})\sum\limits_{i=1}^{n}\lambda_i}
\end{aligned}$$

Which is the mgf of a Poisson random variable with parameter $\sum\limits_{i=1}^{n}\lambda_i$. Thus $Y \sim$Poisson($\sum\limits_{i=1}^{n}\lambda_i$).

Chapter 31

Probability

31.1 Elementary Probability Concepts

31.1.1 Definition of Probability

Let S be a sample space associated with an experiment. For every event $A \in S$ (ie, A is a subset of S), we assign a number, $P(A)$ - called the *probability* of A - such that the following three axioms hold:

- Axiom 1: $P(A) \geq 0$.
- Axiom 2: $P(S) = 1$.
- Axiom 3: If $A_1, A_2, A_3, ...$ form a sequence of pairwise mutually exclusive events in S (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then
  $P(A_1 \cup A_2 \cup A_3 \cup ...) = \sum_{i=1}^{\infty} P(A_i)$.

31.1.2 Definition: Conditional Probability

The conditional probability of an event A, given that an event B has occured and $P(B) > 0$ is equal to

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

31.1.3 Definition: Independence

Events A and B are said to be independent if any of the following holds

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

31.1.4 Theorem: Multiplicative Law of Probability

The probability of the intersection of two events A and B is

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

Proof:

By the definition of conditional probability

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ \Rightarrow P(A|B) \cdot P(B) &= P(A \cap B) \end{aligned}$$

Likewise

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ \Rightarrow P(B|A) \cdot P(A) &= P(B \cap A) \end{aligned}$$

Since $P(A \cap B) = P(B \cap A)$

$$\begin{aligned} P(A|B) \cdot P(B) &= P(A \cap B) \\ &= P(B \cap A) \\ &= P(B|A) \cdot P(A) \\ \Rightarrow P(A \cap B) &= P(A|B) \cdot P(B) \\ &= P(B|A) \cdot P(A) \end{aligned}$$

31.1.5 Corollary

If A and B are independent, then

$$P(A \cap B) = P(A) \cdot P(B)$$

Proof:

When A and B are independent, by the definition of independence[1],

$$\begin{aligned} P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \\ \Rightarrow = P(A) \cdot P(B) = P(B) \cdot P(A) \end{aligned}$$

31.1.6 Additive Law of Probability

The probability of the union of two events is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof:

$A \cup B = A \cup (A^c \cap B)$ where A and $(A^c \cap B)$ are mutually exclusive. $\Rightarrow P(A \cup B) = P(A) + P(A^c \cap B)$

[1] Definition **??**

$B = (A^c \cap B) \cup (A \cap B)$ where $(A^c \cap B)$ and $(A \cap B)$ are mutually exclusive. $\Rightarrow P(B) = P(A^c \cap B) + P(A \cap B)$
$\Rightarrow P(A^c \cap B) = P(B) - P(A \cap B)$

$P(A \cup B) = P(A) + P(A^c \cap B)$
$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

31.1.7 Corollary

If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

Proof:

When A and B are mutually exclusive, $(A \cap B) = \emptyset$ and $P(A \cap B) = 0$. By Theorem **??**,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ \Rightarrow P(A \cup B) &= P(A) + P(B) - 0 \\ \Rightarrow P(A \cup B) &= P(A) + P(B) \end{aligned}$$

31.1.8 Theorem: Law of Complements

If A is an event, then $P(A) = 1 - P(A^c)$.

Proof:

Let S be the sample space.

$$\begin{aligned} S &= A \cup A^c \\ \Rightarrow P(S) &= P(A \cup A^c) \\ \Rightarrow P(S) &= P(A) + P(A^c) - P(A \cap A^c) \\ \Rightarrow P(S) &= P(A) + P(A^c) - 0 \\ \Rightarrow P(S) &= P(A) + P(A^c) \\ \Rightarrow 1 &= P(A) + P(A^c) \\ \Rightarrow 1 - P(A^c) &= P(A) \\ \Rightarrow P(A) &= 1 - P(A^c) \end{aligned}$$

31.1.9 Definition: Partition of a Sample Space

For some positive integer k, let the sets $B_1, B_2, \ldots, B_k$ be such that

- $S = B_1 \cup B_2 \cup \ldots \cup B_k$.
- $B_i \cap B_j = \emptyset$ for $i \neq j$.

Then the collection of sets $B_1, B_2, \ldots, B_k$ is said to be a *partition* of S.

31.1.10 Definition: Decomposition

If A is any subset of S and $B_1, B_2, \ldots, B_k$ is a partition of S, A can be *decomposed* as follows:

$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$

31.1.11 Theorem: Total Law of Probability

If $B_1, B_2, \ldots, B_k$ is a partition of S such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$, then for any event A

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

Proof:

Any subset A of S can be written as

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_k) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$

Since $B_1, B_2, \ldots, B_k$ is a partition of S, if $i \neq j$,

$(A \cap B_i) \cup (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$. That is, $(A \cap B_i)$ and $(A \cap B_j)$ are mutually exclusive events. Thus,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\ {}^{[1]} \Rightarrow &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k) \\ \Rightarrow &= \sum_{i=1}^{k} P(A|B_i)P(B_i) \end{aligned}$$

1. Theorem ??: Multiplicative Law of Probability

31.1.12 Theorem: Bayes' Rule

If $B_1, B_2, \ldots, B_k$ is a partition of S such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$, then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum\limits_{i=1}^{k} P(A|B_i)P(B_i)}$$

Proof:

$$\begin{aligned} {}^{[1]}P(B_j|A) &= \frac{P(A \cap B)}{P(A)} \\ {}^{[2]} \Rightarrow &= \frac{P(A|B_j)P(B_j)}{P(A)} \\ {}^{[3]} \Rightarrow &= \frac{P(A|B_j)P(B_j)}{\sum\limits_{i=1}^{k} P(A|B_i)P(B_i)} \end{aligned}$$

1. Definition ??, Conditional Probability
2. Definition ??, Conditional Probability
3. Theorem ??, Law of Total Probability

Chapter 32

Real Number System

This chapter was prepared by Steve MacDonald (McDonald 2003–2007).

32.1 Historical Note

The first axiom system known in the history of mathematics was Euclid's aximoatic development of plane geometry. Euclid's treatment began with some *primitive* or *undefined* terms and some assumed statements, which he called *axioms* and *postulates*. For Euclid an *axiom* was a general "self-evident" truth, such as "The whole is greater than any of its parts" or "equals added to equals are equal," whereas a *postulate* was an assumed statement about the relationships among the primitives of his system, such as "Two points determine exactly one line." From these undefined terms and basic assumptions a whole body of other statements, called *theorems* was deduced. For centuries Euclidean Geometry, which was assumed to be the true description of physical reality, remained the only mathematical systems with such an axiomatic foundation.

Then in the nineteenth century, spurred by Lobatchevsky and others who discovered that by modifying the postulates another logically consistent geometry could be constructed, matematicians began to apply this deductive approach to other branches of mathematics. Not only did this work do much to organize and clarify such familiar disciplines as number theory, analysis, and algebra, but it helped develop new areas of mathematics such as topology. Note that today we do not make Euclid's distinction between *axiom* and *postulate*, using the terms synonymously.

As an example of this deductive approach, we now want to give an axiomatic description of the real numbers system and thus place a logical foundation under many of the "rules" you learned in high school algebra.

32.2 The Field of Real Numbers

32.2.1 Definition: The Field of Real Numbers

The Field of Real Numbers is a set $\Re$ of objects called *numbers* together with two well-defined binary operations, called *addition*, denoted by +, and *multiplication*, denoted by $\cdot$ or juxtaposition, satisfying the Field Axioms. (By well-defined, we mean that if $s = s'$ and $t = t'$, then $s + t = s' + t'$.)

32.2.2 Field Axioms

1. (Closure for addition) For each pair $x, y, \in \Re$, there exists a unique object in $\Re$, called the *sum* of x and y denoted by $x + y$.
2. (Associative law for addition) For all $x, y, z \in \Re,\ (x + y) = z = x + (y + z)$.
3. (Additive identity) There exists and object $0 \in \Re$ such that for all $x \in \Re,\ x + 0 = x = 0 + x$.
4. (Additive inverse) For each $x \in \Re$, there exists some $y \in \Re$ such that $x + y = 0 = y + x$. We will usually denote the additive inverse of x by $-x$.
5. (Commutative law of addition) For all $x, y \in \Re,\ x + y = y + x$.
6. (Closure for multiplication) For each pair $, y, \in \Re$, there exists a unique object in $\Re$, called the *product* of x and y and denoted by $x \cdot y$ or xy.
7. (Associative law of multiplication) For all $x, y, z \in \Re,\ x \cdot (y \cdot z) = (x \cdot y) \cdot z$.
8. (Multiplicative identity) There exists an object $1 \in \Re$ such that for all $x \in \Re,\ 1 \cdot x = x = x \cdot 1$.
9. (Multiplicative inverse) For each $x \in \Re$ such that $x \neq 0$, there exists an object $y \in \Re$ such that $x \cdot y = 1 = y \cdot x$. We will usually denote the multiplicative inverse of x by x^{-1}.
10. (Commutative law of multiplication) For all $x, y \in \Re,\ x \cdot y = y \cdot x$.
11. (Distributive law of multiplication over addition) For all $x, y, z \in \Re, x \cdot (y + z) = x \cdot y + x \cdot z$.
12. (Positive Elements) There exists a nonempty subset $\Re^+ \subset \Re$ closed under $+$ and $\cdot$. That is, for all $x, y \in \Re^+,\ x + y \in \Re$ and $x \cdot y \in \Re^+$.
13. (Trichotomy) For any $x \in \Re$, exactly one of these three cases holds: $x \in \Re^+,\ -x \in \Re^+$, or $x = 0$.

32.2.3 Definiton: Less Than (or Equal To)

Let $x, y \in \Re$. We say that x *is less than* y, written $x < y$, provided $y + -x \in \Re^+$. We say that x is *less than or equal* y iff and only iff $x < y$ or $x = y$. (Also, x is said to be greater than y if $y < x$.)

32.2.4 Definition: Bounded Above (and Below)

A set A of real numbers is said to be *bounded above* if there exists some $b \in \Re$ such that $x \leq b,\ \forall x \in A$. In this case, b is called an *upper bound* for A. (Bounded below and lower bound are defined similarly.)

32.2.5 Definition: Least Upper (and Lower) Bound

Let A be a set of real numbers bounded above. An element $\beta \in \Re$ is called the *least upper bound*, often written *lub* for A iff and only if

i. β is an upper bound for A and
ii. $\beta \leq b$ for every b which is an upper bound for A. (A greatest lower bound is defined similarly.)

32.2.6 Completeness Axiom

Every nonempty subset of $\Re$ having an upper bound has a least upper bound.

32.3 Proof that the Field of Rationals is not Complete

Let $A = \{p \in Q^+ | p^2 < 2\}$ and $B = \{p \in Q^+ | p^2 > 2\}$.

We claim that A has no largest element and that B has no smallest element; i.e., given that $p \in A$, we can find some $q \in A$ with $q > p$; and given any $p \in B$, we can find some $q \in B$ with $q < p$.

For any $p \in Q^+$, let

$$\begin{aligned} q &= p - \frac{p^2-2}{p+2} \\ &= \frac{p^2+2p-p^2+2}{p+2} \\ &= \frac{2p+2}{p+2} \end{aligned}$$

Then

$$q^2 - 2 = \frac{4p^2+8p+r-2p^2-8p-8}{(p+2)^2} \quad = \frac{2p^2-4}{(p+2)^2} = \frac{2(p+2)}{(p+2)^2}$$

Now if $p \in A,\ p^2-2<0$,

so $-\frac{p^2-2}{p+2} > 0$,

whence $q = p - \frac{p^2-2}{p+2} > p$.

But $q^2 - 2 = \frac{2(p^2-2)}{(p+2)^2} < 0$

implies that $q^2 < 2$. Thus $q \in A$ and $q > p$.

On the other hand, if $p \in B$, then $p^2 - 2 > 0$,

so $q = p - \frac{p^2-2}{p+2} < p$.

But $q^2 - 2 = \frac{2(p^2-2)}{(p+2)^2} > 0$ implies that $q \in B$.

Here $q \in B$ and $q < p$.

Now it is clear that every member of B is an upper bound for A, and every member of A is a lower bound for set B. It is also clear from the above demonstration that *among the rationals Q,* the nonempty set A has *no least upper bound*; and the nonempty set B has *no greatest lower bound among the rationals* . Therefore we have shown that the ordered field of rational numbers does not satisfy the conditions of the *Completeness Axiom* . Thus it is the *Completeness Axiom* that distinguishes the ordered field of real numbers from the ordered field of rational numbers.

32.4 Preliminary Results in the Field of Real Numbers

32.4.1 Theorem: Uniqueness of Identities

Identity elements are unique.

Proof: Suppose $u \star a = a \star u = a$ and $e \star a = a \star e = a,\ \forall a \in \Re$. Then $u = u \star e = e$.

32.4.2 Theorem 2: Uniqueness of Inverses

If $\star$ is an associative operation, inverse elements for $\star$ are unique.

Proof:

Suppose $a \star a_1 = a_1 \star a = e$ and $a \star a_2 = a_2 \star a = e$, where e is the identity element for the operation $\star$.

Then $a_1 = a_1 \star e = a_1 \star (a \star a_2) = (a_1 \star a) \star a_2) = e \star a_2 = a_2$.

32.4.3 Theorem: Left Cancellation Law

If a has an inverse a' with respect to the associative operation $\star$, and $a \star b = a \star c$, then $b = c$.

Proof:

Suppose $a \star b = a \star c$. Then

$$\begin{aligned} a' \star (a \star b) &= a' \star (a \star c) \\ \Rightarrow (a' \star a) \star b &= (a' \star a) \star c \\ \Rightarrow e \star b &= e \star c \\ \Rightarrow b &= c \end{aligned}$$

32.4.4 Corollary: Right Cancellation

In the field $(\Re, +, \cdot)$, $a + b = a + c \Rightarrow b = c$, and if $a \neq 0$ and $ab = ac$, then $b = c$.

Proof: The Corollary is proved using commutativity and Left Cancellation.

32.4.5 Lemma

$$-0 = 0$$

.

Proof:

By axiom 4 $-0 + 0 = 0$. Because 0 is the additive identity, $-0 + 0 = -0$. Therefore $-0 = -0 + 0 = 0$.

32.4.6 Theorem

$\forall a \in \Re,\ a > 0$ if and only if $a \in \Re^+$.

Proof: Suppose $0 < a$. Then $a - 0 \in \Re^+$, but $a - 0 = a$, so $a \in \Re^+$. Conversely, suppose $a \in \Re^+$. Then $a - 0 = a \in \Re^+$, so $0 < a$.

32.4.7 Theorem

$\forall x \in \Re,\ x \cdot 0 = 0$.

Proof:

$$\begin{aligned} 0 + x \cdot 0 &= x \cdot 0 \\ &= x \cdot (0 + 0) \\ &= x \cdot 0 + x \cdot 0 \\ Rightarrow 0 &= x \cdot 0 \end{aligned}$$

32.4.8 Theorem

$\forall x \in \Re,\ -(-x) = x.$

Proof:

$-(-x) + (-x) = 0$ and $0 = x + (-x)$. So

$$-(-x) + (-x) = x + (-x) \Rightarrow -(-x) \quad = x$$

32.4.9 Theorem

$$\forall x, y \in \Re,\ x \cdot (-y) = -(x \cdot y) = (-x) \cdot y$$

Proof:

$$\begin{aligned} x(-y) + xy &= x(-y + y) \qquad = x \cdot 0 \\ &= 0 \\ &= -(xy) + xy \\ \Rightarrow x(-y) + xy &= -(xy) + xy \\ \Rightarrow x(-y) &= -(xy) \end{aligned}$$

Similarly,

$$\begin{aligned} (-x)y + xy &= (-x + x)y \\ &= 0 \cdot y \\ &= 0 \\ &= -(xy) + xy \\ \Rightarrow (-x)y &= -(xy) \end{aligned}$$

By transitivity, $(-x)y = -(xy) = x(-y)$.

32.4.10 Theorem

$$\forall x \in \Re,\ (-1) \cdot x = -x$$

Proof:

By Theorem 32.4.9

$$(-1)x = -(1x) \quad = -x$$

32.4.11 Corollary

$$\forall x \in \Re,\ (-1) \cdot (-x) = x$$

Proof:

By 32.4.10

$$(-1)(-x) = -(-x) = x$$

32.4.12 Theorem

$$\forall x \in \Re,\ (-x)(-x) = x \cdot x$$

Proof:

By Corollary 32.4.11

$$\begin{aligned}(-x)(-x) &= -(x(-x)) \\ &= -((-x)x) \\ &= -(-(x \cdot x)) \\ &= (x \cdot x) \\ &= x \cdot x\end{aligned}$$

32.4.13 Theorem

Let x and y be any real numbers. Then exactly one of the following is true:

i. $x > y$
ii. $x = y$
iii. $x < y$

Proof:

By Axiom 1, $\Re$ is closed under addition. Thus, since $x, y \in \Re$, $x + (-y) \in \Re$. By Trichotomy, $x + (-y) \in \Re$, $-(x + (-y)) \in \Re$, or $x + (-y) = 0$.

$x + (-y) \in \Re \ \Rightarrow x > y$ (Definition 32.2.3)

$-(x + (-y)) = (-x) + y \in \Re \ \Rightarrow x < y$ (Axiom 4)

$x + (-y) = 0 \ \Rightarrow x = y$ (Definition 32.2.3).

32.4.14 Theorem

$$\forall a, b, c \in \Re,\ (a < b \land b < c) \Rightarrow a < c$$

Proof:

$$\begin{aligned}&a < b \land b < c \\ \Rightarrow\ &b - a > 0 \land c - b > 0 \\ \Rightarrow^{[1]}\ &(b - a) + (c - b) > 0 \\ \Rightarrow\ &(c - b) + (b - a) > 0 \\ \Rightarrow\ &c - b + b - a > 0 \\ \Rightarrow\ &c - a > 0 \\ \Rightarrow\ &a < c\end{aligned}$$

1. Axiom 12

32.4.15 Theorem

$\forall a, b, c \in \Re,\ (a < b \ \Rightarrow a + c < b + c)$.

Proof:

$$\begin{aligned} a &< b \\ \Rightarrow b - a &> 0 \\ \Rightarrow b - c + c - a &> 0 \\ \Rightarrow b + c - a - c &> 0 \\ \Rightarrow (b + c) - (a + c) &> 0 \\ \Rightarrow a + c &< b + c \end{aligned}$$

32.4.16 Theorem

$\forall a, b, c \in \Re,\ (a < b \wedge c > 0) \Rightarrow ac < bc.\backslash$

Proof:

$$\begin{aligned} a &< b \\ \Rightarrow b - a &> 0 \\ {}^{[1]} \Rightarrow c(b - a) &> 0 \\ \Rightarrow bc - ac &> 0 \\ \Rightarrow ac &< bc \end{aligned}$$

1. Axiom 12 states that $\Re^+$ is closed under $\cdot$.

32.4.17 Theorem

$\forall a, b, c \in \Re,\ (a < b \wedge c < 0) \Rightarrow bc < ba$.

Proof:

$$\begin{aligned} a &< b \\ \Rightarrow b - a &> 0 \\ \Rightarrow c(b - a) &< 0 \\ \Rightarrow bc - ba &< 0 \\ \Rightarrow bc &< ba \end{aligned}$$

1. Theorem **??**.

32.4.18 Theorem

If $a > b > 0$ and $c > d > 0$, then $ac > bd$.

Proof:

$$
\begin{aligned}
& a > b \wedge c > d \\
& \Rightarrow a - b > 0 \wedge c - d > 0 \\
& \Rightarrow c(a-b) > 0 \wedge b(c-d) > 0 \\
& \Rightarrow ac - bc > 0 \wedge bc - bd > 0 \\
& \Rightarrow ac - bc + bc - bd > 0 \\
& \Rightarrow ac - bd > 0 \\
& \Rightarrow ac > bd
\end{aligned}
$$

32.4.19 Theorem

In a field containing at least two elements, $1 \in \Re^+$.

Proof:

Since in a field of more than one element, $1 \neq 0$ (Theorem **??**) we may assume that 1 is not 0. Hence, by trichotomy, either $1 \in \Re^+$ or $-1 \in \Re^+$. Suppose that $-1 \in \Re^+$. Then by closure $(-1)(-1) \in \Re^+$, but by Corollary **??**, $(-1)(-1) = 1$, so we now have both $-1 \in \Re^+$ and $1 \in \Re^+$, which is a contradiction, and hence it is impossible that $-1 \in \Re^+$. Therefore, $1 \in \Re^+$.

32.4.20 Theorem

If $x > 0$, then $x^{-1} > 0$.

Proof:

Since x^{-1} has an inverse x, we know that $x^{-1} \neq 0$. Hence, by Axiom 13, either $x^{-1} > 0$ or $x^{-1} < 0$. Suppose $x^{-1} < 0$. Then $-x^{-1} \in \Re^+$, and since $\Re^+$ is closed under multiplication, $(-x^{-1}) \cdot x \in \Re^+$. Now by Theorem **??**, $(-x^{-1}) \cdot x = -(x^{-1} \cdot x) = -1$, so this would imply that $-1 \in \Re^+$, in contradction to Theorem **??**. Since $x^{-1} < 0$ must be be false, we conclude that $x^{-1} > 0$ whenever $x > 0$.

Chapter 33

Rounding

33.1 Floor (Next Lowest Integer)

Rounding to the next lower integer, denoted $\lfloor x \rfloor$, is defined as

$$\lfloor x \rfloor = max\{m \in \mathbb{Z} | m \leq x\}$$

33.2 Ceiling

Rounding to the next larger integer, denoted $\lceil x \rceil$, is defined as

$$\lceil x \rceil = min\{m \in \mathbb{Z} | m \geq z\}$$

33.3 Nearest Integer

Rounding to the nearest integer is a common operation without a notation with well established concensus. It may be represented as $[x]$, $\lfloor x \rceil$, $||x||$, $nint(x)$ or $round(x)$. Here, we will use the $nint(x)$ notation so that we may extend the $round(x)$ notation beyond just rounding to the nearest integer..

A mathematical representation of rounding would be really nice to have right here

33.4 Nearest Multiple

Rounding to a nearest multiple may be obtained through a rescaling of the value x into a integer scale based on multiples of m. One available notation for this operation is $mround(x, m)$.

$$mround(x, m) = nint\left(\frac{x}{m}\right) \cdot m$$

33.5 Nearest Place (Base 10)

It is common in mathematical operations to round a value not to an integer, but to a decimal place. This is no different than rounding to a multiple of ten. We define the following conventions.

- Let $x \in \mathbb{R}$
- Let $p \in \mathbb{Z}$ where p represents the negative power of ten of the desired precision of the result.

Thus, when $p = 0$, we with to round to the $10^{-0} = 10^0 = 1$, or ones/integer position. When $p = 1$, we round to the $10^{-1} = \overset{1/10}{=} 0.01$, or tenths position. And when $p = -1$, we round to the $10^1 = 10$, or tens position.

We define the operation $round(x, p)$ to be the operation of rounding to the nearest decimal place.

$$\begin{aligned} round(x,p) &= mround(x, 10^{-p}) \\ &= round\Big(\frac{x}{10^{-p}}, 0\Big) \cdot 10^{-p} \\ &= \frac{round(x \cdot 10^p, 0)}{10^p} \end{aligned}$$

Under this definition, $nint(x)$ is a special case of $round$ where $nint(x) = round(x, 0)$.

33.6 Breaking Ties

33.6.1 Rounding Even

33.6.2 Rounding Odd

33.6.3 Round Away From Zero

33.6.4 Round Toward Zero

33.7 References

(I really need to be better than this)

1. https://en.wikipedia.org/wiki/Floor_and_ceiling_functions
2. https://en.wikipedia.org/wiki/Nearest_integer_function

Chapter 34

Sample Size Estimation

34.1 Solving Group Sample Sizes Using Weights

Let n be the total sample size obtained by adding two groups such that $n = n_1 + n_2$. Let w represent the proportion of n allocated to n_1, referred to as the *weight* of n_1. Then $n_2 = \frac{n_1 \cdot (1-w)}{w}$.

Furthermore, $n = n_1 + \frac{n_1 \cdot (1-w)}{w}$.

Proof:

By the assumptions, we know

$$\begin{aligned} n &= n_1 + n_2 \\ &= w \cdot n + (1 - w) \cdot n \end{aligned}$$

This implies

$$\begin{aligned} n_1 &= w \cdot n \\ n_2 &= (1 - w) \cdot n \end{aligned}$$

We observe the following:

$$\begin{aligned} \frac{n_2}{n_1} &= \frac{(1-w) \cdot n}{w \cdot n} \\ &= \frac{(1-w)}{w} \\ \Rightarrow n_2 &= \frac{n_1 \cdot (1-w)}{w} \end{aligned}$$

This further implies

$$n = n_1 + \frac{n_1 \cdot (1-w)}{w}$$

Notice now that both n and n_2 are defined as functions of n_1 and w. Thus, we may estimate the sample size required in each of two groups by estimating only n_1, provided we know the weight w.

34.1.1 Corollary

For $k \in \mathbb{N}$, let n be the total sample size of k subgroups such that

$$n = n_1 + n_2 + n_3 + ... + n_k$$

Suppose, further, that there exists a vector of weights W that satisfy the following conditions:

1. For each $w_i \in W$, $0 \leq w_i \leq 1$
2. $\sum_{i=1}^{k} w_i = 1$
3. $n = w_1 \cdot n + w_2 \cdot n + ... + w_k \cdot n = \sum_{i=1}^{k} w_i \cdot n$

Let us assign the values $n_1 = w_1 \cdot n$, $n_2 = w_2 \cdot n$, $\ldots$, $n_k = w_k \cdot n$. Then for all $i|i \leq k$,

$$\begin{aligned} \frac{n_i}{n_1} &= \frac{w_i \cdot n}{w_1 \cdot n} \\ &= \frac{w_i}{w_1} \\ \Rightarrow n_i &= \frac{n_1 \cdot w_i}{w_1} \end{aligned}$$

Thus, each n_i may be estimated by estimating the value of n_1 provided W is a fully specified vector of weights for each of the k groups.

Chapter 35

Skew-Normal Distribution

35.1 Preliminary Theorems

35.2 Lemma: A Symmetry Theorem

Suppose the pdf of X, f_X is symmetric about 0. Let $w(\cdot)$ be any odd function. Then the pdf of $Y = w(X)$, f_Y, is also symmetric about 0.

Proof:

Recall that if a pdf is symmetric about zero, it must demonstrate the property $P(T \leq t) = P(T \geq -t)$. Since f_X is symmetric, we know

$$\begin{aligned} P(X \leq x) &= P(X \geq -x) \\ \Rightarrow P\big[w(X) \leq w(x)\big] &= P\big[w(X) \geq w(-x)\big]^{[1]} \quad = P\big[w(X) \geq -w(x)\big] \\ &= P(Y \leq y) \\ &= P(Y \geq -y) \end{aligned}$$

Thus f_Y is symmetric about 0.

1. By the definition of an odd function $f(-x) = -f(x)$.

35.3 Lemma

Let f_0 be a one-dimensional probability density function symmetric over 0.

Also, let G be a one dimensional probability distribution function such that G' exists and is a density function symmetric over 0. Then

$$f(x) = 2f_0(x)G\big(w(x)\big) \quad (-\infty < x < \infty)$$

is a probability density function for any odd fuction $w(\cdot)$.\

Proof:

Let $X \sim f_0$ and $Y \sim G'$.

Now consider the random variable $X - w(Y)|Y$.

When Y is fixed, $X - w(Y)$ is an odd function of X and, by Lemma **??**, $X - w(Y)$ is symmetric over 0. Thus,

$$\begin{aligned}
\frac{1}{2} &= P\big(X - w(Y) \le 0|Y\big) \\
\overset{[1]}{} &= E\Big[P\big(X - w(Y) \le 0|Y\big)\Big] \\
\Rightarrow \frac{1}{2} &= E\big[P\big(X \le w(Y)|Y\big)\big] \\
\Rightarrow \frac{1}{2} &= \int_{-\infty}^{\infty} P\big(X \le w(Y)|Y\big)p(x)dx \\
\Rightarrow \frac{1}{2} &= \int_{-\infty}^{\infty} G\big(w(Y)\big)f_0(x)dx \\
\Rightarrow 1 &= 2\int_{-\infty}^{\infty} f_0(X)G\big(w(Y)\big)dx
\end{aligned}$$

So $f(x) = 2f_0(x)G\big(w(y)\big)$ is a valid density function for all $x,\ x \in \Re$.

1. The expected value of $P(X \le 0) = \frac{1}{2}$ when X is distributed symmetric over 0.

35.4 Expected Values

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \cdot 2f(x)\Phi(\alpha x)dx \\
&= 2\int_{-\infty}^{\infty} xf(x)\Phi(\alpha x)dx \\
&= 2\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{x^2}{2}\Big\}\Bigg[\int_{-\infty}^{\alpha x}\frac{1}{\sqrt{s\pi}}\exp\Big\{-\frac{t^2}{2}\Big\}dt\Bigg]dx \\
&= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\alpha x} x\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{x^2}{2}\Big\}\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{t^2}{2}\Big\}dtdx \\
&= \frac{2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\alpha x} x\exp\Big\{-\frac{x^2}{2}\Big\}\exp\Big\{-\frac{t^2}{2}\Big\}dtdx \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\alpha x} x\exp\Big\{-\frac{x^2}{2}\Big\}\exp\Big\{-\frac{t^2}{2}\Big\}dtdx
\end{aligned}$$

But $\exp\Big\{-\frac{t^2}{2}\Big\}$ cannot be integrated in closed form, so the solution must be found with numerical methods.

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 \cdot 2f(x)\Phi(\alpha x)dx \\
&= 2\int_{-\infty}^{\infty} x^2 f(x)\Phi(\alpha x)dx \\
&= 2\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}\left[\int_{-\infty}^{\alpha x} \frac{1}{\sqrt{s\pi}} \exp\left\{-\frac{t^2}{2}\right\}dt\right]dx \\
&= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\alpha x} x^2 \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\}dtdx \\
&= \frac{2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\alpha x} x^2 \exp\left\{-\frac{x^2}{2}\right\} \exp\left\{-\frac{t^2}{2}\right\}dtdx \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\alpha x} x^2 \exp\left\{-\frac{x^2}{2}\right\} \exp\left\{-\frac{t^2}{2}\right\}dtdx
\end{aligned}
$$

But $\exp\left\{-\frac{t^2}{2}\right\}$ cannot be integrated in closed form, so the solution must be found with numerical methods.

35.5 Estimation of λ

Using the Moment Generating Function, it was shown that the skew of the Skew-Normal distribution can be calculated as

$$
S = sign(\lambda)\left(2 - \frac{\pi}{2}\right)\left(\frac{\lambda^2}{\frac{pi}{2} + (\frac{pi}{2} - 1)\lambda^2}\right)^{\frac{3}{2}}
$$

where S denotes the skew of the distribution. Given a value of skew for the distribution, a link can be made back to λ. We begin by noticing that the following process is identical regardless of the sign of λ. It is presented here as if $\lambda > 0$

$$
\begin{aligned}
S &= \left(2-\frac{\pi}{2}\right)\left(\frac{\lambda^2}{\frac{\pi}{2}+\left(\frac{\pi}{2}-1\right)\lambda^2}\right)^{3/2} \\
\Rightarrow \left(\frac{S}{2-\frac{\pi}{2}}\right) &= \left(\frac{\lambda^2}{\frac{\pi}{2}+\left(\frac{\pi}{2}-1\right)\lambda^2}\right)^{3/2} \\
\Rightarrow \left(\frac{S}{2-\frac{\pi}{2}}\right)^{2/3} &= \left(\frac{\lambda^2}{\frac{\pi}{2}+\left(\frac{\pi}{2}-1\right)\lambda^2}\right) \\
\Rightarrow T &= \left(\frac{\lambda^2}{\frac{\pi}{2}+\left(\frac{\pi}{2}-1\right)\lambda^2}\right) \\
\Rightarrow \left(\frac{\pi}{2}+\left(\frac{\pi}{2}-1\right)\lambda^2\right)T = \lambda^2 \\
\Rightarrow \frac{\pi}{2}T+\left(\frac{\pi}{2}-1\right)\lambda^2 T = \lambda^2 \\
\Rightarrow \frac{\pi}{2}T &= \lambda^2-\left(\frac{\pi}{2}-1\right)\lambda^2 T \\
\Rightarrow \frac{\pi}{2}T &= \lambda^2\left(1-\left(\frac{\pi}{2}-1\right)T\right) \\
\Rightarrow \lambda^2 &= \frac{\frac{\pi}{2}T}{1-\left(\frac{\pi}{2}-1\right)T} \\
\Rightarrow \lambda^2 &= \frac{\frac{\pi}{2}\left(\frac{S}{2-\frac{\pi}{2}}\right)^{2/3}}{1-\left(\frac{\pi}{2}-1\right)\frac{S}{2-\frac{\pi}{2}}\Big)^{2/3}} \\
\Rightarrow \lambda &= \frac{\sqrt{\frac{\pi}{2}}\left(\frac{S}{2-\frac{\pi}{2}}\right)^{1/3}}{\sqrt{1-\left(\frac{\pi}{2}-1\right)\frac{S}{2-\frac{\pi}{2}}\Big)^{2/3}}}
\end{aligned}
$$

1. Let $T=\left(\frac{S}{2-\frac{\pi}{2}}\right)^{2/3}$

This equation is only defined for certain values of S. In particular, S cannot be a number such that the denominator is 0, nor can the that which appears under the radical be negative. These two restrictions can be collapsed, and the equation is defined so long as

$$
\begin{aligned}
1-\left(\frac{\pi}{2}-1\right)\left(\frac{S}{2-\frac{\pi}{2}}\right)^{2/3} &> 0\\
\Rightarrow 1 &> \left(\frac{\pi}{2}-1\right)\left(\frac{S}{2-\frac{\pi}{2}}\right)^{2/3}\\
\Rightarrow \left(\frac{\pi}{2}-1\right)^{-1} &> \left(\frac{S}{2-\frac{\pi}{2}}\right)^{2/3}\\
\Rightarrow \frac{\left(2-\frac{\pi}{2}\right)^{2/3}}{\frac{\pi}{2}-1} &> S^{2/3}\\
\Rightarrow \frac{2-\frac{\pi}{2}}{\left(\frac{\pi}{2}-1\right)^{3/2}} &> S\\
\Rightarrow -.9952 < S &< .9952
\end{aligned}
$$

We notice that the endpoints of this interval are approximations. Ideally, the interval would span from -1 to 1, as most estimators of skew provide a value in that interval–values close to negative one denoting a strong left skew; values close to one denoting a strong right skew; 0 denoting perfect symmetry. Although this relationship is not perfect, it is quite close to what we would like, and can be practically implemented.

Chapter 36

Somers' D

Somers' D has an asymptotically $Normal$ Distribution **??**. It may take any value between -1 and 1. It is used to measure classification agreement between a predictor and outcome variable.

Somers' D is related to a form of a concordance index. Concrodance in measured between 0 and 1 and can effectively be calculated by rescaling Somers' D. The rescaling can be accomplished by:

$$C = \frac{D+1}{2}$$

36.1 Theorems for Somers' D

36.1.1 Theorem: Distribution of Somers' Derived Concordance

Let $D \sim \text{Normal}(\mu, \sigma^2)$. Then $C \sim \text{Normal}\ (\frac{\mu+1}{2}, \frac{\sigma^2}{4})$.

Proof:

$$D \sim Normal(\mu, \sigma^2) \Rightarrow (D{+}1) \sim Normal(\mu{+}1, \sigma^2)(D{+}1) \sim Normal(\mu{+}1, \sigma^2) \Rightarrow \frac{D+1}{2} \sim Normal(\frac{\mu+1}{2}, \frac{\sigma^2}{4})$$

By definition, $C = \frac{D+1}{2}$, so $C \sim Normal(\frac{\mu+1}{2}, \frac{\sigma^2}{4})$.

Note: when the dependent variable is a binary response, the Concordance Index is equal to the area under the Receiver Operator Characteristic (ROC) curve, or AUC.

Chapter 37

Summation

37.1 Theorems of Summation

37.1.1 Theorem

If c is a constant then

$$\sum_{i=1}^{n} c = nc$$

Proof:

$$\sum_{i=1}^{n} c = \underbrace{c + c + \cdots + c}_{n \text{ terms}} = nc$$

37.1.2 Theorem

If $a_1, a_2, \ldots, a_n$ are real numbers and c is a constant, then

$$\sum_{i=1}^{n} ca_i = c\sum_{i=1}^{n} a_i$$

Proof:

$$\begin{aligned}\sum_{i=1}^{n} ca_i &= ca_1 + ca_2 + \cdots + ca_n \\ &= c(a_1 + a_2 + \cdots + a_n) \\ &= c\sum_{i=1}^{n} a_i\end{aligned}$$

37.1.3 Theorem

If $a_{1,2}, \ldots, a_n$ are real numbers and $b_1, b_2, \ldots, b_n$ are real numbers, then

$$\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$$

Proof:

$$\begin{aligned}
\sum_{i=1}^{n}(a_i + b_i) &= a_1 + b_1 + a_2 + b_2 + \cdots + a_n + b_n \\
&= a_1 + a_2 + \cdots + a_n + b_1 + b_2 + \cdots + b_n \\
&= \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i
\end{aligned}$$

37.1.4 Theorem

If a_i and b_j are real numbers for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, then then

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j = a_+ b_+$$

Proof:

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j &= \sum_{i=1}^{n}\left(a_i \sum_{j=1}^{m} b_j\right) \\
&= \sum_{i=1}^{n} a_i b_+ \\
&= b_+ \sum_{i=1}^{n} a_i \\
&= a_+ b_+
\end{aligned}$$

37.1.5 Theorem

If a_i is a real number for $i = 1, 2, \ldots, n$ and b is a real number, then

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b = m a_+ b$$

Proof:

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b &= \sum_{i=1}^{n} m a_i b \\
&= mb \sum_{i=1}^{n} a_i \\
&= m a_+ b
\end{aligned}$$

37.1.6 Theorem

If a_j is a real number for $j = 1, 2, \ldots, m$ and b is a real number, then

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_j b = n a_+ b$$

Proof:

$$\begin{aligned}\sum_{i=1}^{n} \sum_{j=1}^{m} a_j b &= \sum_{i=1}^{n} \left(b \sum_{j=1}^{m} a_j \right) \\ &= \sum_{i=1}^{n} a_+ b \\ &= n a_+ b\end{aligned}$$

37.1.7 Theorem

If a_i and b_{ij} are real numbers for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, then

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_{ij} = \sum_{i=1}^{n} a_i b_{i+}$$

Proof:

$$\begin{aligned}\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_{ij} &= \sum_{i=1}^{n} \left(a_i \sum_{j=1}^{m} b_{ij} \right) \\ &= \sum_{i=1}^{n} a_i b_{i+}\end{aligned}$$

37.1.8 Theorem

If a_j and b_{ij} are real numbers for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, then

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_j b_{ij} = \sum_{i=1}^{n} a_j b_{+j}$$

Proof:

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{m} a_j b_{ij} &= a_1 b_{11} + a_2 b_{12} + \cdots + a_m b_{1m} \\
&\quad + a_1 b_{21} + a_2 b_{22} + \cdots + a_m b_{2m} \\
&\quad \vdots \\
&\quad + a_1 b_{n1} + a_1 b_{n1} + \cdots + a_1 b_{nm} \\
&= a_1 b_{11} + a_1 b_{21} + \cdots + a_1 b_{n1} \\
&\quad + a_2 b_{12} + a_2 b_{22} + \cdots + a_2 b_{n2} \\
&\quad \vdots \\
&\quad + a_m b_{1m} + a_m b_{2m} + \cdots + a_n b_{nm} \\
&= a_1 (b_{11} + b_{21} + \cdots + b_{n1}) \\
&\quad + a_2 (b_{12} + b_{22} + \cdots + b_{n2}) \\
&\quad \vdots \\
&\quad + a_m (b_{1m} + b_{2m} + \cdots + b_{nm}) \\
&= a_1 b_{+1} + a_2 b_{+2} + \cdots + a_m b_{+m} \\
&= \sum_{j=1}^{m} a_j b_{+j}
\end{aligned}
$$

Chapter 38

Analysis of Variance

38.1 One-Way Design

38.1.1 Decomposition of Sums of Squares

$$
\begin{aligned}
SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+} + \bar{x}_{i+} - x_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(x_{ij} - \bar{x}i+) + (\bar{x}_{i+} - \bar{x}_{++})\right]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(\bar{x}_{i+} - \bar{x}_{++}) + (x_{ij} - \bar{x}i+)\right]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(\bar{x}_{i+} - \bar{x}_{++})^2 + 2(\bar{x}_{i+} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+}) + (x_{ij} - \bar{x}_{i+})^2\right] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{x}_{i+} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}2(\bar{x}_{i+} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+}) + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 2\sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+}) + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 2\sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})\left(\sum_{j=1}^{n_i}x_{ij} - \sum_{j=1}^{n_i}\bar{x}_{i+}\right) + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 2\sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})(x_{i+} - n_i\bar{x}_{i+}) + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 2\sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})\big(x_{i+} - n_i\frac{x_{i+}}{n_i}\big) + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 2\sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})(x_{i+} - x_{i+}) + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 2\sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})\cdot 0 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + 0 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 \\
&= \sum_{i=1}^{a}n_i(\bar{x}_{i+} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2
\end{aligned}
$$

The components are commonly referred to as

$$SS_{Factor} = \sum_{i=1}^{a} n_i(\bar{x}_{i+} - \bar{x}_{++})^2$$

and

$$SS_{Error} = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2$$

Notice that SS_{Factor} compares the factor means to the overall mean, and it can be said that SS_{Factor} measures the variation *between* factors. SS_{Error} compares each observation to the overall mean, and can be said to describe the variation *within* factors.

When $n_1 = n_2 = \cdots n_i = n$, the design is said to be balanced.

See (Montgomery 2004, 66)

38.2 Computational Formulas

SS_{Total} and SS_{Factor} can be simplified for convenient computation.

$$\begin{aligned} SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{++})^2 \\ &\overset{[1]}{=} \sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij}^2 - x_{++}\sum_{j=1}^{n_i}\frac{1}{n_i} \end{aligned}$$

1. See Theorem 42.3.1

$$\begin{aligned} SS_{Factor} &= \sum_{i=1}^{a} n_i(\bar{x}_{i+} - \bar{x}_{++})^2 \\ &\overset{[1]}{=} \sum_{i=1}^{a}\frac{\bar{x}_{i+}^2}{n_i} - \bar{x}_{++}\sum_{i=1}^{a}\frac{1}{n_i} \end{aligned}$$

1. See Theorem 42.3.1

SS_{Error} does not simplify to a convenient form, but

$$\begin{aligned} SS_{Total} &= SS_{Factor} + SS_{Error} \\ \Rightarrow SS_{Error} &= SS_{Total} - SS_{Factor} \end{aligned}$$

38.3 Randomized Complete Block Design

Blocking in ANOVA is a method of eliminate the effect of a controllable nuisance variable. To implement this design, suppose we have a treatments we want to compare, and b blocks. We may analyze the data by use of the sums of squares, similar to the one-way design.

38.3.1 Decomposition of Sums of Squares

$$
\begin{aligned}
SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} + \bar{x}_{i+} - \bar{x}_{i+} + \bar{x}_{+j} - \bar{x}_{+j} + \bar{x}_{++} - \bar{x}_{++} - \bar{x}_{++})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++}) + (\bar{x}_{+j} - \bar{x}_{++}) + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})\big]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++})^2 + 2(\bar{x}_{i+} - \bar{x}_{++})(\bar{x}_{+j} - \bar{x}_{++}) \\
&\quad + 2(\bar{x}_{i+} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++}) + (\bar{x}_{+j} - \bar{x}_{++})^2 \\
&\quad + 2(\bar{x}_{+j} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++}) \\
&\quad + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2\big] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++})^2 + (\bar{x}_{+j} - \bar{x}_{++})^2 + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2 \\
&\quad + 2(\bar{x}_{i+} - \bar{x}_{++})(\bar{x}_{+j} - \bar{x}_{++}) + 2(\bar{x}_{i+} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++}) \\
&\quad + 2(\bar{x}_{+j} - \bar{x}_{++})(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})\big] \\
{}^{[1]} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+} - \bar{x}_{++})^2 + (\bar{x}_{+j} - \bar{x}_{++})^2 + (x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2 + 0 + 0 + 0\big] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2 \\
&= b\sum_{i=1}^{a}(\bar{x}_{i+} - \bar{x}_{++})^2 + a\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2
\end{aligned}
$$

1. It is shown that the cross products are equal to zero in Section 38.3.3

These terms are commonly referred to as

$$
\begin{aligned}
SS_{Factor} &= b\sum_{i=1}^{a}(\bar{x}_{i+} - \bar{x}_{++})^2 \\
SS_{Block} &= a\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2 \\
SS_{Error} &= \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{i+} - \bar{x}_{+j} + \bar{x}_{++})^2
\end{aligned}
$$

See (Montgomery 2004, 126)

38.3.2 Computational Formulae

SS_{Total}, SS_{Factor}, and SS_{Block} can all be simplified for convenient computation.

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \bar{x}_{++})^2$$
$$[1] = \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}^2 - \frac{x_{++}}{ab}$$

$$SS_{Factor} = b\sum_{i=1}^{a}(\bar{x}_{i+} - \bar{x}_{++})^2$$
$$[1] = \frac{1}{b}\sum_{i=1}^{a} x_{i+}^2 - \frac{x_{++}^2}{ab}$$

$$SS_{Block} = a\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})^2$$
$$[1] = \frac{1}{a}\sum_{j=1}^{b} x_{+j}^2 - \frac{x_{++}^2}{ab}$$

1. See Theorem 42.3.1

SS_{Error} does not simplify to any convenient form, but may be calculated from the other terms as $SS_{Error} = SS_{Total} - SS_{Factor} - SS_{Block}$

38.3.3 RCBD Cross Products

The cross products of the RCBD design

$$2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+} - \bar{x}_{++})(\bar{x}_{+j} - \bar{x}{++})$$
$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j} - \bar{x}_{++})(x_{ij} + \bar{x}_{i+} + \bar{x}_{+j} - \bar{x}_{++})$$
$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+i} - \bar{x}_{++})(x_{ij} + \bar{x}_{i+} + \bar{x}_{+j} - \bar{x}_{++}) = 0$$

Proof:

$$
\begin{aligned}
& 2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+}-\bar{x}_{++})(\bar{x}_{+j}-\bar{x}{++}) + 2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&\quad + 2\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+i}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&= 2\Bigg(\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{i+}-\bar{x}_{++})(\bar{x}_{+j}-\bar{x}{++}) + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+j}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&\quad + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{+i}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++})\Bigg) \\
&= 2\sum_{i=1}^{a}\sum_{j=1}^{b}\big[(\bar{x}_{i+}-\bar{x}_{++})(\bar{x}_{+j}-\bar{x}{++}) + (\bar{x}_{+j}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++}) \\
&\quad + (\bar{x}_{+i}-\bar{x}_{++})(x_{ij}+\bar{x}_{i+}+\bar{x}_{+j}-\bar{x}_{++})\big] \\
&= 2\sum_{i=1}^{a}\sum_{j=1}^{b}\big[\bar{x}_{i+}\bar{x}_{+j}-\bar{x}_{i+}\bar{x}_{++}-\bar{x}_{+j}\bar{x}_{++}+\bar{x}_{++}^2 \\
&\quad + x_{ij}\bar{x}_{+j}-\bar{x}_{i+}\bar{x}_{+j}-\bar{x}_{+j}^2+\bar{x}_{+j}\bar{x}_{++}-x_{ij}\bar{x}_{++}+\bar{x}_{i+}\bar{x}_{++}+\bar{x}_{+j}\bar{x}_{++}-\bar{x}_{++}^2 \\
&\quad + x_{ij}\bar{x}_{+j}-\bar{x}_{i+}^2-\bar{x}_{i+}\bar{x}_{+j}+\bar{x}_{+j}\bar{x}_{++}-x_{ij}\bar{x}_{++}+\bar{x}_{i+}\bar{x}_{++}+\bar{x}_{+j}\bar{x}_{++}-\bar{x}_{++}^2\big] \\
&= 2\sum_{i=1}^{a}\sum_{j=1}^{b}(-\bar{x}_{++}^2-\bar{x}_{i+}^2-\bar{x}_{+j}^2+x_{ij}\bar{x}_{i+}+x_{ij}\bar{x}_{+j}-2x_{ij}\bar{x}_{++}-\bar{x}_{i+}\bar{x}_{+j} \\
&\quad + 2\bar{x}_{i+}\bar{x}_{++}+2\bar{x}_{+j}\bar{x}_{++}) \\
&= 2\Bigg(-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{++}^2-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}^2-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{i+}+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{+j} \\
&\quad -\sum_{i=1}^{a}\sum_{j=1}^{b}2x_{ij}\bar{x}_{++}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg) \\
&= 2\Bigg(\frac{ab\bar{x}_{++}^2}{a^2b^2}-\frac{b}{b^2}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{a}{a^2}\sum_{j=1}^{b}\bar{x}_{+j}^2+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{i+}+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{+j} \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg)
\end{aligned}
$$

[1]
$$
\begin{aligned}
&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}\bar{x}_{+j} \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg)
\end{aligned}
$$

[2]
$$
\begin{aligned}
&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{x}_{i+}\bar{x}_{+j}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg)
\end{aligned}
$$

[3]
$$
\begin{aligned}
&= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 \\
&\quad -\frac{2\bar{x}{++}^2}{ab}-\frac{\bar{x}_{++}}{ab}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{i+}\bar{x}_{++}+\sum_{i=1}^{a}\sum_{j=1}^{b}2\bar{x}_{+j}\bar{x}_{++}\Bigg)
\end{aligned}
$$

[4]
$$
= 2\Bigg(\frac{\bar{x}_{++}^2}{ab}-\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2-\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2+\frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2+\frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2
$$

$$-\frac{2\bar{x}{+}{+}^2}{ab} - \frac{\bar{x}_{++}}{ab} + \frac{2\bar{x}_{++}^2}{ab} + \sum_{i=1}^{a}\sum_{j=1}^{b} 2\bar{x}_{+j}\bar{x}_{++}\Bigg) \quad [5] = 2\Bigg(\frac{\bar{x}_{++}^2}{ab} - \frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2 - \frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 + \frac{1}{b}\sum_{i=1}^{a}\bar{x}_{i+}^2 + \frac{1}{a}\sum_{j=1}^{b}\bar{x}_{+j}^2 \quad -\frac{2\bar{x}{+}{+}^2}{ab} - \frac{\bar{x}_{++}}{ab} + \cdot$$

1. See Summation Theorem 37.1.7
2. See Summation Theorem 37.1.8
3. See Summation Theorem 37.1.4
4. See Summation Theorem 37.1.5
5. See Summation Theorem 37.1.6

Using the theorems in Chapter @ref{summation-chapter} it is can be shown that each of the three cross products is equal to zero. However, the physical tedium of reducing each cross product is much greater than the approach taken above.

Chapter 39

Method of Transformations

Suppose we wish to find the distribution funciton for the random variable Y that is either a strictly increasing or strictly decreasing function (Such a function is sure to have an inverse, whereas a function like $Y = X^2$ does not have an inverse).. If we know the distribution of X, we may use the following to determine the cdf of Y.

$$\begin{aligned} P(Y \leq y) &= P(h(X) \leq y) \\ &= P(h^{-1}(h(y)) \leq h^{-1}(x)) \\ &= P(Y \leq h^{-1}(x)) \\ \Rightarrow F_Y(y) &= F_X(h^{-1}(x)) \end{aligned}$$

The pdf can now be found by taking the deriviative of the cdf.

$$\begin{aligned} f_Y(y) &= \frac{d(F_Y(y))}{dy} \\ &= \frac{dF_Y(h^{-1}(y))}{dy} \\ &= f_X(h^{-1}(y))\frac{d(h^{-1}(y))}{dy} \end{aligned}$$

39.1 Example: Cauchy Distribution

Let X have the Uniform pdf

$$f(x) = \begin{cases} \frac{1}{\pi}, & \frac{-\pi}{2} < x < \frac{\pi}{2} \\ 0 & otherwise \end{cases}$$

Let $Y = \tan(X)$. The pdf of Y can be found as follows:

$$\begin{aligned}
h(x) &= \tan(x) \\
\Rightarrow h^{-1}(x) &= \tan^{-1}(x) \\
\Rightarrow \frac{dh^{-1}(x)}{dx} &= \frac{1}{1+x^2} \\
\Rightarrow f_Y(y) &= f_X(h^{-1}(y))\frac{d(h^{-1}(y))}{dy} \\
&= f_X(tan^{-1}(x))\frac{1}{1+y^2} \\
&= \frac{1}{\pi}\frac{1}{1+y^2} \\
&= \frac{1}{\pi(1+y^2)}
\end{aligned}$$

The domain of Y is transformed

$$\frac{-\pi}{2} < x < \frac{-\pi}{2} \Rightarrow \tan\left(\frac{-\pi}{2}\right) < \tan(x) < \tan\left(\frac{-\pi}{2}\right) \Rightarrow -\infty < y < \infty$$

Thus the pdf of the Y, known as the Cauchy distribution, is

$$f_Y(y) = \frac{1}{\pi(1+y^2)}, \quad -\infty < y < \infty$$

Chapter 40

T-test

40.1 One-Sample T-test

The t-test is commonly used to look for evidence that the mean of a normally distributed random variable may differ from a hypothesized (or previously observed) value.

40.1.1 T-Statistic

The t-statistic is a standardized measure of the magnitude of difference between a sample's mean and some known, non-random constant. It is similar to a z-statistic, but differs in that a t-statistic may be calculated without knowledge of the population variance.

40.1.2 Definitions and Terminology

Let $\bar{x}$ be a sample mean from a sample with standard deviation s. Let μ_0 be a constant, and $s_{\bar{x}} = s/\sqrt{n}$ be the standard error of the parameter $\bar{x}$. t is defined:

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

and has $\nu = n - 1$ degrees of freedom.

40.1.3 Hypotheses

The hypotheses for these test take the forms:

For a two-sided test:

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

For a one-sided test:

$$H_0 : \mu < \mu_0$$
$$H_a : \mu \geq \mu_0$$

or

$$H_0 : \mu > \mu_0$$
$$H_a : \mu \leq \mu_0$$

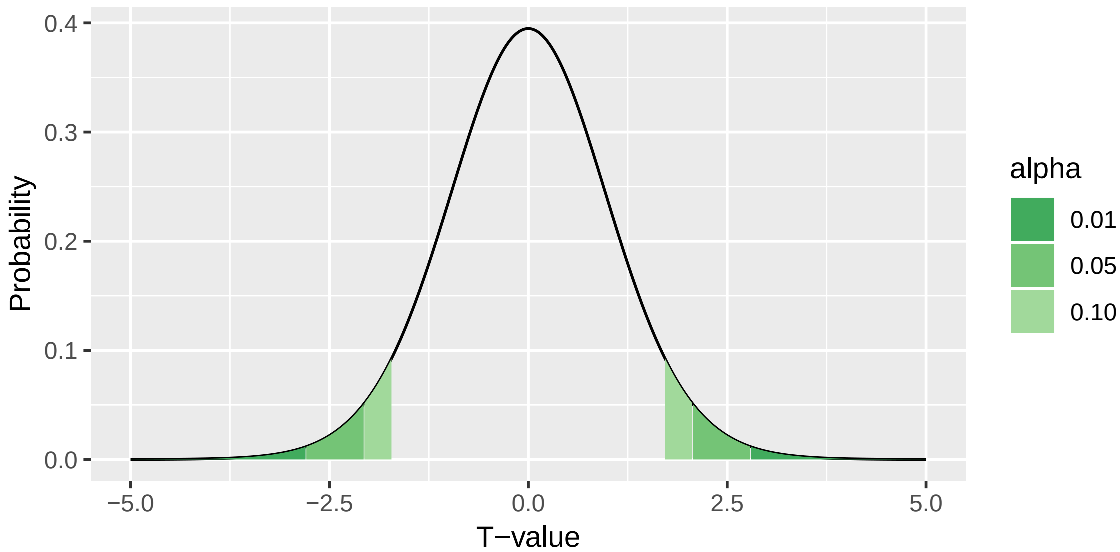

Figure 40.1: The example displayed uses 25 degrees of freedom

To compare a sample $(X_1, \ldots, X_n)$ against the hypothesized value, a T-statistic is calculated in the form:

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Where $\bar{x}$ is the sample mean and s is the sample standard deviation.

40.1.4 Decision Rule

The decision to reject a null hypothesis is made when an observed T-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of shaded areas on both sides equaling the corresponding α. It follows, then, that the decision rule is:

Reject H_0 when $T \leq t_{\alpha/2,\nu}$ or when $T \geq t_{1-\alpha/2,\nu}$.

By taking advantage of the symmetry of the T-distribution, we can simplify the decision rule to:

Reject H_0 when $|T| \geq t_{1-\alpha/2,\nu}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

In the one-sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follows, then, that the decision rule for a lower tailed test is:

Reject H_0 when $T \leq t_{\alpha,\nu}$.

For an upper tailed test, the decision rule is:

Reject H_0 when $T \geq t_{1-\alpha,\nu}$.

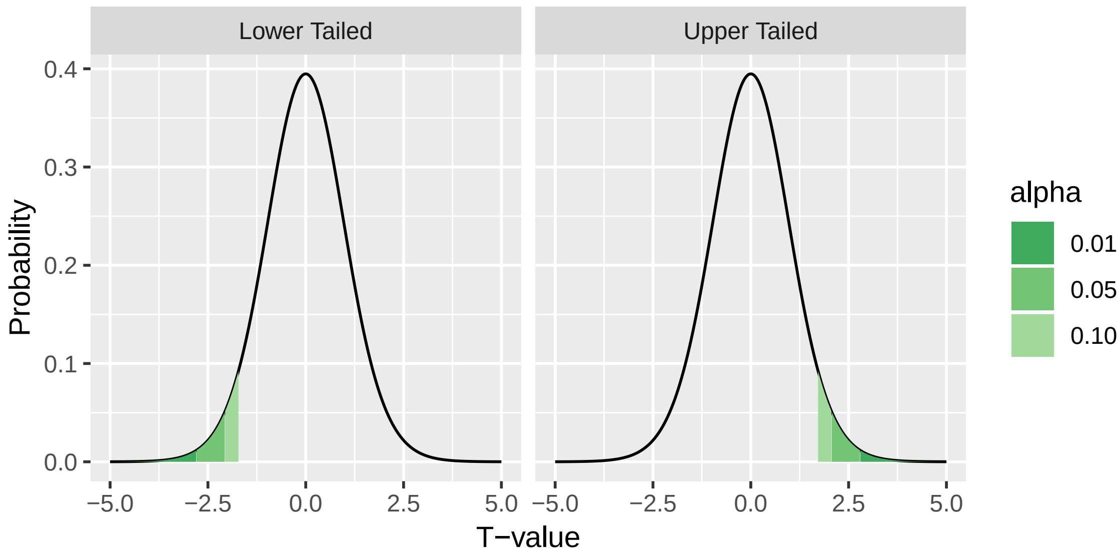

Figure 40.2: The example displayed uses 25 degrees of freedom

Using the symmetry of the T-distribution, we can simplify the decision rule as:

Reject H_0 when $|T| \geq t_{1-\alpha,\nu}$.

```
## Warning: Ignoring unknown aesthetics: ymax

## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of $\bar{x}$:

Reject H_0 when $\bar{x} \leq \mu_0 - t_\alpha \cdot s/\sqrt{n}$ or $\bar{x} \geq \mu_0 + t_\alpha \cdot s/\sqrt{n}$.

This change can be justified by:

$$|T| \geq t_{1-\alpha,\nu}$$
$$\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \geq t_{1-\alpha,\nu}$$

$$\begin{aligned} -\left(\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right) &\geq t_{1-\alpha,\nu} & \frac{\bar{x}-\mu_0}{s/\sqrt{n}} &\geq t_{1-\alpha,\nu} \\ \bar{x}-\mu_0 &\leq -t_{1-\alpha,\nu} \cdot s/\sqrt{n} & \bar{x}-\mu_0 &\geq t_{1-\alpha,\nu} \cdot s/\sqrt{n} \\ \bar{x} &\leq \mu_0 - t_{1-\alpha,\nu} \cdot s/\sqrt{n} & \bar{x} &\geq \mu_0 + t_{1-\alpha,\nu} \cdot s/\sqrt{n} \end{aligned}$$

For a two-sided test, both the conditions apply. The left side condition is used for a left-tailed test, and the right side condition for a right-tailed test.

40.1.5 Power

The derivations below make use of the following symbols:

- $\bar{x}$: The sample mean
- s: The sample standard deviation
- n: The sample size
- μ_0: The value of population mean under the null hypothesis

- μ_a: The value of the population mean under the alternative hypothesis.
- α: The significance level
- $\gamma(\mu)$: The power of the test for the parameter μ.
- $t_{\alpha,\nu}$: A quantile of the central t-distribution for a probability, α and $n-1$ degrees of freedom.
- T: A calculated value to be compared against a t-distribution.
- C: The critical region (rejection region) of the test.

Two-Sided Test

$$\begin{aligned}
\gamma(\mu_a) &= P_{\mu_a}(\bar{x} \in C) \\
&= P_{\mu}\big(\bar{x} \leq \mu_0 - t_{\alpha/2,\nu} \cdot s/\sqrt{n}\big) + P_{\mu_a}\big(\bar{x} \geq \mu_0 + t_{1-\alpha/2,\nu} \cdot s/\sqrt{n}\big) \\
&= P_{\mu_a}\big(\bar{x} - \mu_a \leq \mu_0 - \mu_a - t_{\alpha/2,\nu} \cdot s/\sqrt{n}\big) + P_{\mu_a}\big(\bar{x} - \mu_a \geq \mu_0 - \mu_a + t_{1-\alpha/2,\nu} \cdot s/\sqrt{n}\big) \\
&= P_{\mu_a}\Big(\frac{\bar{x} - \mu}{s/\sqrt{n}} \leq \frac{\mu_0 - \mu_a - t_{\alpha/2,\nu} \cdot s/\sqrt{n}}{s/\sqrt{n}}\Big) + P_{\mu_a}\Big(\frac{\bar{x} - \mu}{s/\sqrt{n}} \geq \frac{\mu_0 - \mu_a + t_{1-\alpha/2,\nu} \cdot s/\sqrt{n}}{s/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(T \leq \frac{\mu_0 - \mu_a}{s/\sqrt{n}} - t_{\alpha/2,\nu}\Big) + P_{\mu_a}\Big(T \geq \frac{\mu_0 - \mu_a}{s/\sqrt{n}} + t_{1-\alpha/2,\nu}\Big) \\
&= P_{\mu_a}\Big(T \leq -t_{\alpha/2,\nu} + \frac{\mu_0 - \mu_a}{s/\sqrt{n}}\Big) + P_{\mu_a}\Big(T \geq t_{1-\alpha/2,\nu} + \frac{\mu_0 - \mu_a}{s/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(T \leq -t_{\alpha/2,\nu} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{s}\Big) + P_{\mu_a}\Big(T \geq t_{1-\alpha/2,\nu} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{s}\Big)
\end{aligned}$$

Both $t_{\alpha/2,\nu}$ and $t_{1-\alpha/2,\nu}$ have non-central T-distributions with non-centrality parameter $\frac{\sqrt{n}(\mu_0 - \mu_a)}{s}$.

One-Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|T| \geq t_{1-\alpha}$. Thus, where T occurs in the derivation below, it may reasonably be replaced with $|T|$.

$$\begin{aligned}
\gamma(\mu_a) &= P_{\mu_a}(\bar{x} \in C) \\
&= P_{\mu_a}\big(\bar{x} \geq \mu_0 + t_{1-\alpha,\nu} \cdot s/\sqrt{n}\big) \\
&= P_{\mu_a}\big(\bar{x} - \mu_a \geq \mu_0 - \mu_a + t_{1-\alpha,\nu} \cdot s/\sqrt{n}\big) \\
&= P_{\mu_a}\Big(\frac{\bar{x} - \mu_a}{s/\sqrt{n}} \geq \frac{\mu_0 - \mu_a + t_{1-\alpha,\nu} \cdot s/\sqrt{n}}{s/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(T \geq \frac{\mu_0 - \mu_a}{s/\sqrt{n}} + t_{1-\alpha,\nu}\Big) \\
&= P_{\mu_a}\Big(T \geq t_{1-\alpha,\nu} + \frac{\mu_0 - \mu_a}{s/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(T \geq t_{1-\alpha,\nu} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{s}\Big)
\end{aligned}$$

Where $t_{1-\alpha,\nu} + \frac{\sqrt{n}(\mu_0 - \mu_a)}{s}$ has a non-central t-distribution with non-centrality parameter $\frac{\sqrt{n}(\mu_0 - \mu_a)}{s}$

40.1.6 Confidence Interval

The confidence interval for θ is written:

$$\bar{x} \pm t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

40.2 Two-Sample T-test

The two sample t-test is commonly used to look for evidence that the mean of one normally distributed random variable may differ from that of another normally distributed random variable. The hypotheses for this test take the forms:

40.2.1 T-Statistic

The t-statistic is a standardize measure of the magnitude of difference between two sample means and some known, non-random difference of population means. It is similar to a two sample z-statistic, but differes in that a t-statistic may be calculated without knowledge of the population variances.

40.2.2 Definitions and Terminology

Let $\bar{x_1}$ and $\bar{x_2}$ be sample means from two independent samples with standard deviations s_1 and s_2. Let μ_1 and μ_2 be constants representing the means of the populations from which $\bar{x_1}$ and $\bar{x_2}$ obtained. t is defined:

$$t = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}$$

Where

$$SE^* = \begin{cases} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, & \sigma_1^2 \neq \sigma_2^2 \\ \sqrt{\frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, & \sigma_1^2 = \sigma_2^2 \end{cases}$$

and the degrees of freedom ν are (by the Welch-Satterthwaite equation)

$$\nu = \begin{cases} \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{s_1^2/n_1}{n_1-1} + \frac{s_2^2/n_2}{n_2-1}}, & \sigma_1^2 \neq \sigma_2^2 \\ n_1 + n_2 - 2, & \sigma_1^2 = \sigma_2^2 \end{cases}$$

40.2.3 Hypotheses

For a two-sided test:

$$H_0 : \mu_1 = \mu_2 H_a : \mu_1 \neq \mu_2$$

For a one-sided test:

$$H_0 : \mu_1 \leq \mu_2 H_a : \mu_1 > \mu_2$$

or

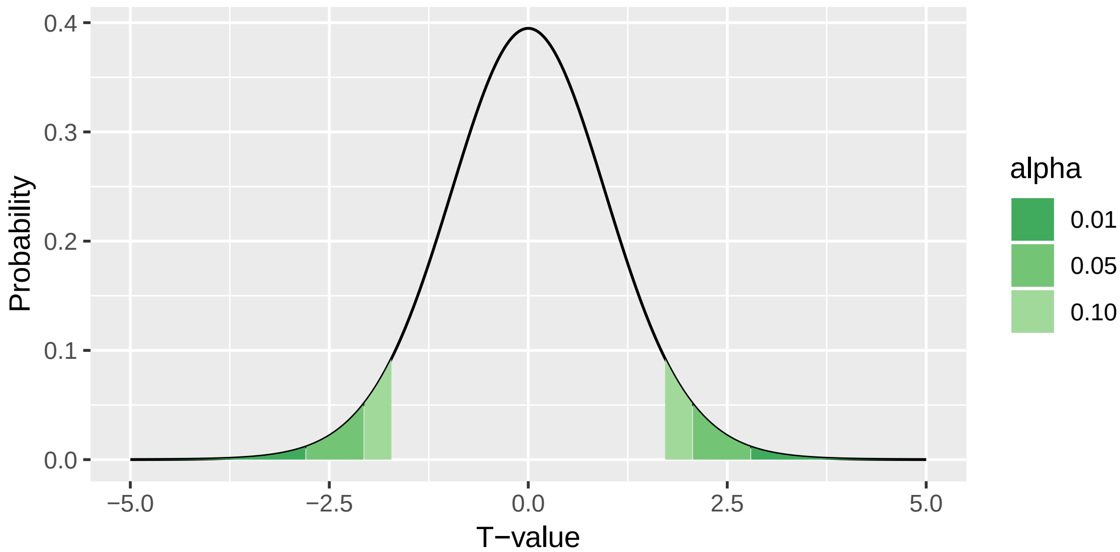

Figure 40.3: The example displayed uses 25 degrees of freedom

$$H_0 : \mu_1 \geq \mu_1 H_a : \mu_1 < \mu_1$$

40.2.4 Decision Rule

The decision to reject a null hypothesis is made when an observed T-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

40.2.4.1 Two Sided Test

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of the shaded areas on both sides equally the corresponding α. It follows then that the decision rule is:

Reject H_0 when $T \leq t_{\alpha/2,\nu}$ or when $T \geq t_{1-\alpha/2,\nu}$.

By taking advantage of the symmetry of the T-distribution, we can simplify the decision rule to:

Reject H_0 when $|T| \geq t_{1-\alpha/2,\nu}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

40.3 One Sided Test

In the one sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follow, then, that the decision rule for a lower tailed test is:

Reject H_0 when $T \leq t_{\alpha,\nu}$.

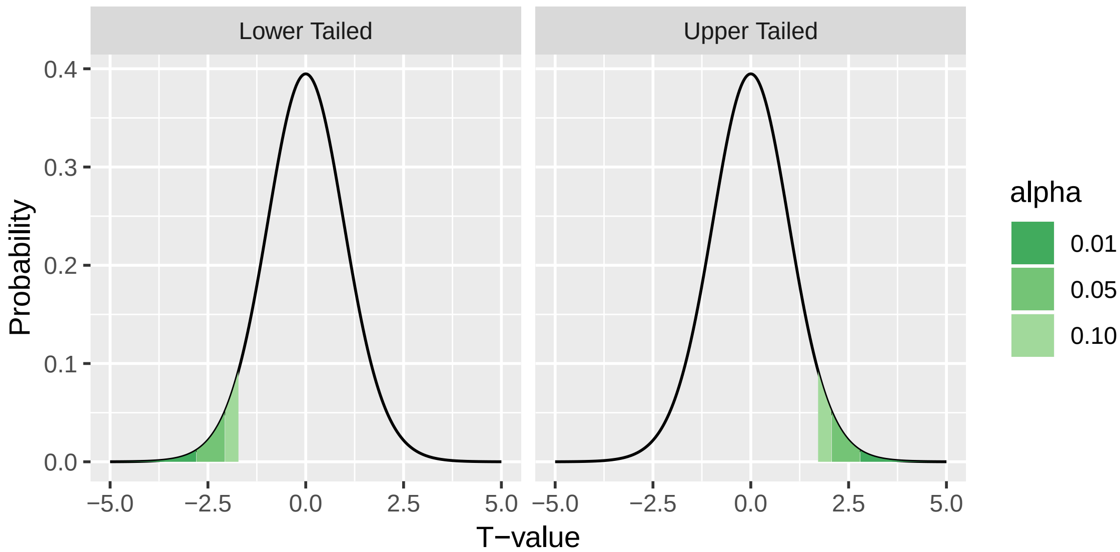

Figure 40.4: The example displayed uses 25 degrees of freedom

For an upper tailed test, the decision rule is:

Reject H_0 when $T \geq t_{1-\alpha,\nu}$.

Using the symmetry of the T-distribution, we can simplify the decision rule as:

Reject H_0 when $|T| \geq t_{1-\alpha,\nu}$.

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of $\bar{x_1}$ and $\bar{x_2}$.

Reject H_0 when $\bar{x_1} - \bar{x_2} \leq (\mu_1 - \mu_2) - t_{\alpha,\nu} \cdot SE^*$ or $\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + t_{\alpha,\nu} \cdot SE^*$

This change can be justified by:

$$
\begin{aligned}
|T| &\geq t_{1-\alpha,\nu} \\
\left|\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right| &\geq t_{1-\alpha,\nu}
\end{aligned}
$$

$$
\begin{aligned}
-\left(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right) &\geq t_{1-\alpha,\nu} \\
(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2) &\leq -t_{1-\alpha,\nu} \cdot SE^* \\
\bar{x_1} - \bar{x_2} &\leq (\mu_1 - \mu_2) - t_{1-\alpha,\nu} \cdot SE^*
\end{aligned}
\qquad
\begin{aligned}
\left(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right) &\geq t_{1-\alpha,\nu} \\
(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2) &\geq t_{1-\alpha,\nu} \cdot SE^* \\
\bar{x_1} - \bar{x_2} &\leq (\mu_1 - \mu_2) + t_{1-\alpha,\nu} \cdot SE^*
\end{aligned}
$$

40.3.1 Power

Two Sided Test

$$
\begin{aligned}
\gamma(\mu_{1a} - \mu_{2a}) &= P_{\mu_{1a}-\mu_{2a}}(\bar{x} \in C) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) \leq (\mu_1 - \mu_2) - t_{\alpha/2,\nu} \cdot SE^*\big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + t_{1-\alpha/2,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a}) \leq (\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) - t_{\alpha/2,\nu} \cdot SE^*\big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a}) \geq (\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha/2,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*} \leq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) - t_{\alpha/2,\nu} \cdot SE^*}{SE^*}\Big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a})}{SE^*} \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha/2,\nu} \cdot SE^*}{SE^*}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \leq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*} - t_{\alpha/2,\nu}\Big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(T \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*} + t_{1-\alpha/2,\nu}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \leq -t_{\alpha/2,\nu} + \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*}\Big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(T \geq t_{1-\alpha/2,\nu} + \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*}\Big)
\end{aligned}
$$

Both $-t_{\alpha/2,\nu} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ and $t_{1-\alpha/2,\nu} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ have non-central T-distributions with non-centrality parameter $\frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$.

One Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|T| \geq t_{1-\alpha}$. Thus, where T occurs in the derivation below, it may reasonably be replaced with $|T|$.

$$
\begin{aligned}
\gamma(\mu_{1a} - \mu_{2a}) &= P_{\mu_{1a}-\mu_{2a}}(\bar{x} \in C) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + t_{1-\alpha,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a}) \geq (\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a})}{SE^*} \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha,\nu} \cdot SE^*}{SE^*}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*} + t_{1-\alpha,\nu}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \geq t_{1-\alpha,\nu} + \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*}\Big)
\end{aligned}
$$

$t_{1-\alpha/2,\nu} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ has a non-central T-distribution with non-centrality parameter $\frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$.

40.3.2 Confidence Interval

The confidence interval for $\mu_1 - \mu_2$ is written:

$$(\bar{x_1} - \bar{x_2}) \pm t_{1-\alpha/2} \cdot SE^*$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = t_{1-\alpha/2} \cdot SE^*$$

40.4 References

1. Wackerly, Mendenhall, Scheaffer, *Mathematical Statistics with Applications*, 6th ed., Duxbury, 2002, ISBN 0-534-37741-6.
2. Daniel, *Biostatistics*, 8th ed., John Wiley & Sons, Inc., 2005, ISBN: 0-471-45654-3.
3. Hogg, McKean, Craig, *Introduction to Mathematical Statistics*, 6th ed., Pearson, 2005, ISBN: 0-13-008507-3
4. Wikipedia, "Student's T test", https://en.wikipedia.org/wiki/Student%27s_t-test
5. Wikipedia, "Welch-Satterthwaite Equation", https://en.wikipedia.org/wiki/Welch%E2%80%93Satterthwaite_equation

Chapter 41

Uniform Distribution

41.1 Probability Density Function

A random variable X is said to have a Uniform Distribution with parameters a and b if its pdf is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & elsewhere \end{cases}$$

41.2 Cumulative Density Function

$$\begin{aligned} F(x) &= \int_a^x \frac{1}{b-a} dt \\ &= \frac{t}{b-a} |_a^x \\ &= \frac{x}{b-a} - \frac{a}{b-a} \\ &= \frac{x-a}{b-a} \end{aligned}$$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1 & elsewhere \end{cases}$$

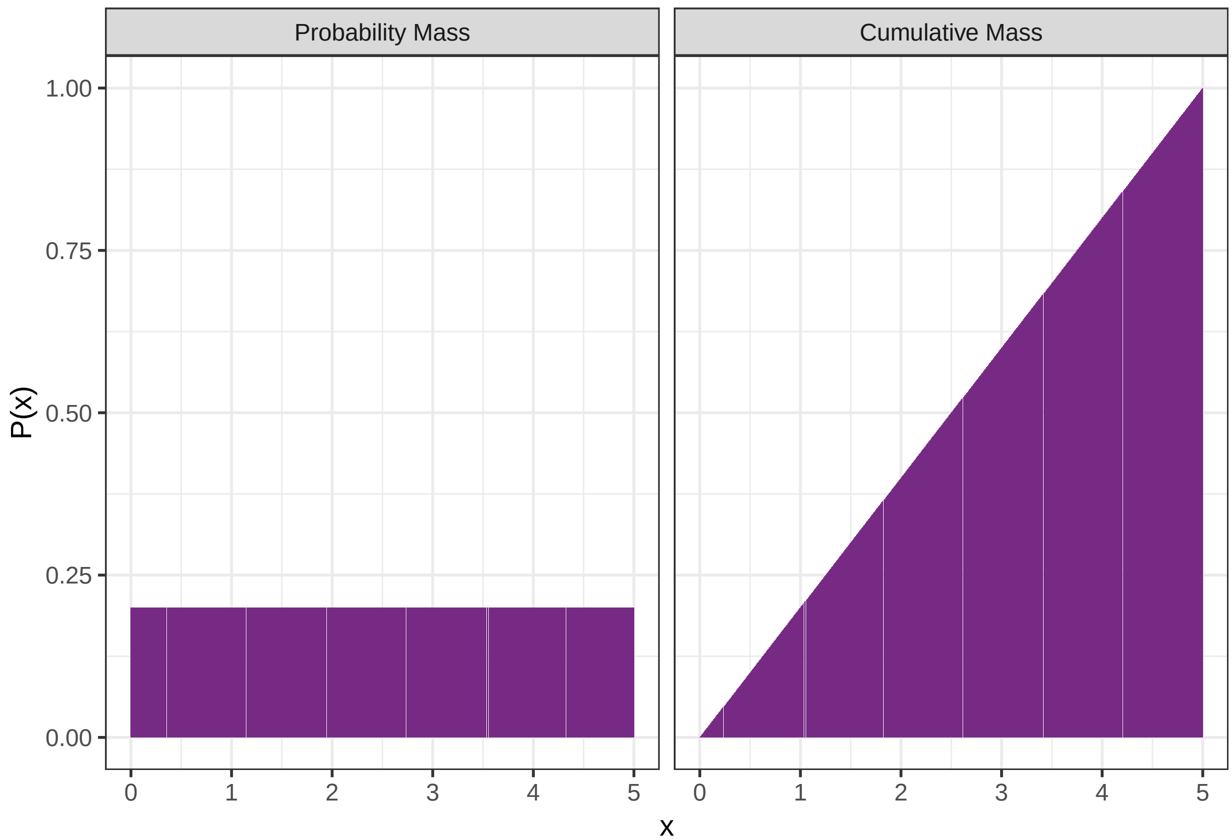

Figure 41.1: The figures on the left and right display the Uniform probability and cumulative distirubtion functions, respectively, for $a = 0, b = 5$.

41.3 Expected Values

$$
\begin{aligned}
E(X) &= \int_a^b x\frac{1}{b-a}dx \\
&= \frac{1}{b-a}\int_a^b x\ dx \\
&= \frac{1}{b-a}\cdot\left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a}\cdot\left[\frac{b^2}{2}-\frac{a^2}{2}\right] \\
&= \frac{1}{b-a}\cdot\frac{b^2-a^2}{2} \\
&= \frac{b^2-a^2}{2(b-a)} \\
&= \frac{(b-a)(b+a)}{2(b-a)} \\
&= \frac{b+a}{2}
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \int_a^b x^2\frac{1}{b-a}dx \\
&= \frac{1}{b-a}\int_a^b x^2\ \frac{1}{b-a}dx \\
&= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^3-a^3}{3}\right] \\
&= \frac{1}{b-a}\left[\frac{(b-a)(b^2+ab+a^2)}{3}\right] \\
&= \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} \\
&= \frac{(b^2+ab+a^2)}{3}
\end{aligned}
$$

$$
\begin{aligned}
\mu &= E(X) \\
&= \frac{b+a}{1}
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= E(X^2)-E(X)^2 \\
&= \frac{b^2+ab+a^2}{3}-\frac{(b-a)^2}{4} \\
&= \frac{4(b^2+ab+a^2)-3(b+a)^2}{12} \\
&= \frac{4(b^2+ab+a^2-3(b^2+2ab+a^2)}{12} \\
& \quad 4b^2+4ab+4a^2-3b^2-6ab-3a^2)
\end{aligned}
$$

41.4 Moment Generating Function

$$
\begin{aligned}
M_X(t) = E(e^{tX}) &= \int_a^b e^{tx}\frac{1}{b-a}dx \\
&= \frac{1}{b-a}\int_a^b e^{tx}dx \\
&= \frac{1}{b-a}\Big[\frac{e^{tb}-e^{ta}}{t}\Big] \\
&= \frac{e^{t(b-a)}}{t(b-a)}
\end{aligned}
$$

$M_X^{(k)}(0)$ will lead to an undefined operation (division by 0). Thus, in the case of the Uniform distribution, we are unable to use the method of moments to identify parameter values.

41.5 Theorems for the Uniform Distribution

41.6 Validity of the Distribution

$$
\int_a^b \frac{1}{b-a} = 1
$$

Proof:

$$
\begin{aligned}
\int_a^b \frac{1}{b-a} &= \frac{x}{b-a}\Big|_a^b \\
&= \frac{b}{b-a} - \frac{a}{b-a} \\
&= \frac{b-a}{b-a} \\
&= 1
\end{aligned}
$$

Chapter 42

Variance Parameter

42.1 Defining Variance With Expected Values

In the case of a discrete random variable, the variance is

$$
\begin{aligned}
\sigma^2 &= \sum_{x=0}^{\infty}(x-\mu)^2 p(x) \\
&= \sum_{x=0}^{\infty}(x^2 - 2\mu x + \mu^2)p(x) \\
&= \sum_{x=0}^{\infty}(x^2 p(x) - 2\mu x \cdot p(x) + \mu^2 p(x)) \\
&= \sum_{x=0}^{\infty} x^2 p(x) - \sum_{x=0}^{\infty} 2\mu x \cdot p(x) + \sum_{x=0}^{\infty} \mu^2 p(x) \\
&= \sum_{x=0}^{\infty} x^2 p(x) - 2\mu \sum_{x=0}^{\infty} x \cdot p(x) + \mu^2 \sum_{x=0}^{\infty} p(x) \\
&= \sum_{x=0}^{\infty} x^2 p(x) - 2\mu \cdot \mu + \mu^2 \\
&= \sum_{x=0}^{\infty} x^2 p(x) - \mu^2 \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

In the case of a continuous random variable, the variance is

$$
\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)dx \\
&= \int_{-\infty}^{\infty} (x^2 f(x) - 2\mu x \cdot f(x) + \mu^2 f(x))dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - \int_{-\infty}^{\infty} 2\mu x \cdot f(x)dx + \int_{-\infty}^{\infty} \mu^2 f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - 2\mu \int_{-\infty}^{\infty} x \cdot f(x)dx + \mu^2 \int_{-\infty}^{\infty} f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - 2\mu \cdot \mu + \mu^2 \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

In general, these results may be summarized as follows:

$$
\begin{aligned}
\sigma^2 &= E[(X-\mu)^2] \\
&= E[(X^2 - 2\mu X + \mu^2)] \\
&= E(X^2) - E(2\mu X) + E(\mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu \cdot \mu + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

42.2 Unbiased Estimator

$$
\begin{aligned}
E\left(\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n}\right) &= \frac{1}{n}E\Big(\sum_{i=1}^{n}(x_i-\bar{x})^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}2\bar{x}x_i+\sum_{i=1}^{n}\bar{x}^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2-2\frac{\sum\limits_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}+n\bar{x}^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2-2\frac{\Big(\sum\limits_{i=1}^{n}x_i\Big)^2}{n}+n\bar{x}^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2-2\frac{n(\sum\limits_{i=1}^{n}x_i)^2}{n^2}+n\bar{x}^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\Big) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2\Big)-E(n\bar{x}^2) \\
&= \frac{1}{n}E\Big(\sum_{i=1}^{n}x_i^2)-nE(\bar{x}^2)\Big) \\
&= \frac{1}{n}\Big[\sum_{i=1}^{n}E(x_i^2)-nE(\bar{x}^2)\Big] \\
[1] &= \frac{1}{n}\Big[\sum_{i=1}^{n}\Big(\sigma^2+\mu^2\Big)-nE(\bar{x}^2)\Big] \\
[2] &= \frac{1}{n}\Big[\sum_{i=1}^{n}\Big(\sigma^2+\mu^2\Big)-n(\frac{\sigma^2}{n}+\mu^2)\Big] \\
\\
&= \frac{1}{n}(n\sigma^2-n\mu^2+\sigma^2-n\mu^2) \\
&= \frac{1}{n}(n\sigma^2-\sigma) \\
&= \frac{1}{n}(n-1)\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

1. $V(X)=E(X^2)-E(X)^2$
   $\Rightarrow E(X^2)=V(X)+E(X)^2=\sigma^2+\mu^2$ $V(\bar{X})=E(\bar{X}^2)-E(\bar{X})^2$
   $\Rightarrow E(\bar{X}^2)=V(\bar{X})+E(\bar{X})^2=\frac{\sigma^2}{n}+\mu^2$

2. By the Central Limit Theorem, $V(\bar{X}) = \frac{\sigma^2}{n}$

Since $E\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}\right) \neq \sigma^2$ it is a biased estimator. Notice, however, that the bias can be eliminated by dividing by $n-1$ instead of by n

$$
\begin{aligned}
E\left(\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right) &= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}(x_i-\bar{x})^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}2\bar{x}x_i+\sum_{i=1}^{n}\bar{x}^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2-2\frac{\sum\limits_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}+n\bar{x}^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2-2\frac{(\sum\limits_{i=1}^{n}x_i)^2}{n}+n\bar{x}^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2-2\frac{n\Big(\sum\limits_{i=1}^{n}x_i\Big)^2}{n^2}+n\bar{x}^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\Big)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2\Big)-E(n\bar{x}^2)\\
&= \frac{1}{n-1}E\Big(\sum_{i=1}^{n}x_i^2\Big)-nE(\bar{x}^2)\\
&= \frac{1}{n-1}\Big[\sum_{i=1}^{n}E(x_i^2)-nE(\bar{x}^2)\Big]\\
[1] &= \frac{1}{n-1}\Big[\sum_{i=1}^{n}(\sigma^2+\mu^2)-nE(\bar{x}^2)\Big]\\
[2] &= \frac{1}{n-1}\Big[\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\Big]\\
&= \frac{1}{n-1}(n\sigma^2-n\mu^2+\sigma^2-n\mu^2)\\
&= \frac{1}{n}(n\sigma^2-\sigma)\\
&= \frac{1}{n-1}(n-1)\sigma^2\\
&= \frac{n-1}{n-1}\sigma^2\\
&= \sigma^2
\end{aligned}
$$

1. $V(X)=E(X^2)-E(X)^2$
   $\Rightarrow E(X^2)=V(X)+E(X)^2=\sigma^2+\mu^2$ $V(\bar{X})=E(\bar{X}^2)-E(\bar{X})^2$
   $\Rightarrow E(\bar{X}^2)=V(\bar{X})+E(\bar{X})^2=\frac{\sigma^2}{n}+\mu^2$
2. By the Central Limit Theorem, $V(\bar{X})=\frac{\sigma^2}{n}$

Thus $E\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right)$ is an unbiased estimator of σ^2, and we define the estimator

$$s^2 = \frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

42.3 Computational Formulae

42.3.1 Computational Formula for σˆ2

$$\begin{aligned}
\sigma^2 &= \frac{\sum\limits_{i=1}^{N}(x_i-\mu)^2}{N} \\
&= \frac{\sum\limits_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}
\end{aligned}$$

Proof:

$$\begin{aligned}
\frac{\sum\limits_{i=1}^{N}(x_i-\mu)^2}{N} &= \frac{\sum\limits_{i=1}^{N}(x_i^2-2\mu x_i+\mu^2)}{N} \\
&= \frac{\sum\limits_{i=1}^{N} x_i^2 - \sum\limits_{i=1}^{N} 2\mu x_i + \sum\limits_{i=1}^{N} \mu^2}{N} \\
&= \frac{\sum\limits_{i=1}^{N} x_i^2 - 2\mu \sum\limits_{i=1}^{N} x_i + N\mu^2}{N} \\
&= \frac{\sum\limits_{i=1}^{N} x_i^2 - 2\frac{\sum_{i=1}^{N} x_i}{N} \sum\limits_{i=1}^{N} x_i + N\left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2}{N} \\
&= \frac{\sum\limits_{i=1}^{N} x_i^2 - 2\frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N} + \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N} \\
&= \frac{\sum\limits_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}
\end{aligned}$$

42.3.2 Computational Formula for s^2

$$
\begin{aligned}
s^2 &= \frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i^2 - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}}{n-1}
\end{aligned}
$$

Proof:

$$
\begin{aligned}
\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n-1} &= \frac{\sum\limits_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)}{n-1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i^2 - \sum\limits_{i=1}^{n} 2\bar{x}x_i + \sum\limits_{i=1}^{n} \bar{x}^2}{n-1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i^2 - 2\bar{x}\sum\limits_{i=1}^{n} x_i + n\bar{x}^2}{n-1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i^2 - 2\frac{\sum\limits_{i=1}^{n} x_i}{n}\sum\limits_{i=1}^{n} x_i + n\left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right)^2}{n-1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i^2 - 2\frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n} + \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}}{n-1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i^2 - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}}{n-1}
\end{aligned}
$$

42.3.3 Corollary: Alternative Computational Formula for s^2

$$
s^2 = \frac{1}{n\cdot(n-1)}\Big(n\cdot\sum_{i=1}^{n} x_i^2 - \big(\sum_{i=1}^{n} x_i\big)^2\Big)
$$

Proof:

Beginning with the result from 42.3.2:

$$
\begin{aligned}
s^2 &= \frac{\sum\limits_{i=1}^{n} x_i^2 - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}}{n-1} \\
&= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}\Big) \\
&= \frac{1}{n-1}\Big(\frac{n \cdot \sum\limits_{i=1}^{n} x_i^2}{n} - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n}\Big) \\
&= \frac{1}{n \cdot (n-1)}\Big(n \cdot \sum_{i=1}^{n} x_i^2 - \big(\sum_{i=1}^{n} x_i\big)^2\Big)
\end{aligned}
$$

42.3.3.1 Application

Both this result and the previous result may be used to optimize calculation of the sample variance in that they permit calculation to occur in one pass of the data. Consider the following sample:

```
x <- c(2, 11, 9, 7, 13, 3, 20)
```

Using the formula

$$
s^2 = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2)}{n-1}
$$

the sample variance is calculated

```
x_bar <- 0
n <- length(x)
for (i in seq_along(x)){
  x_bar <- x_bar + x[i] / n
}

variance <- 0
for (i in seq_along(x)){
  variance <- variance + (x[i] - x_bar)^2 / (n - 1)
}
```

This calculation required the use of two `for` loops. Using the computational formula

$$
s^2 = \frac{1}{n \cdot (n-1)}\Big(n \cdot \sum_{i=1}^{n} x_i^2 - \big(\sum_{i=1}^{n} x_i\big)^2\Big)
$$

the sample variance can be calculated using one `for` loop.

```
sum_square <- 0
sum_i <- 0
n <- length(x)
for(i in seq_along(x)){
  sum_square <- sum_square + x[i]^2
  sum_i <- sum_i + x[i]
```

```
}

variance <- 1 / (n * (n - 1)) * (n * sum_square - sum_i^2)
```

Comparing the processing time of these two approaches shows that the computational formula, utilizing a single `for` loop, requires less than half the processing time of the definitional formula.

```
library(microbenchmark)
# sample of 1,000 values between 1 and 100
x <- sample(1:100, 1000, replace = TRUE)
microbenchmark(
  standard =
  {
    x_bar <- 0
    n <- length(x)
    for (i in seq_along(x)){
      x_bar <- x_bar + x[i] / n
    }

    variance <- 0
    for (i in seq_along(x)){
      variance <- variance + (x[i] - x_bar)^2 / (n - 1)
    }
  },
  computational =
  {
    sum_square <- 0
    sum_i <- 0
    n <- length(x)
    for(i in seq_along(x)){
      sum_square <- sum_square + x[i]^2
      sum_i <- sum_i + x[i]
    }

    variance <- 1 / (n * (n - 1)) * (n * sum_square - sum_i^2)
  }
)
```

```
## Unit: milliseconds
##           expr      min       lq      mean    median        uq      max
##       standard 9.600945 9.903891 10.736589 10.147264 11.170213 17.07874
##  computational 4.497152 4.639826  5.502862  4.730442  5.027518 37.29336
##  neval cld
##    100   b
##    100  a
```

Chapter 43

Weibull Distribution

43.1 Probability Distribution Function

A random variable X is said to have a Weibull Distribution with parameters α and β if its probability density function is:

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, & 0 < x,\ 0 < \alpha,\ 0 < \beta \\ 0 & otherwise \end{cases}$$

43.2 Cumulative Distribution Function

$$\begin{aligned} \int_0^x \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} dt &= \alpha\beta \int_0^x t^{\beta-1} e^{-\alpha t^{\beta}} dt \\ &= \alpha\beta \Big[\frac{-1}{\alpha\beta} e^{-\alpha t^{\beta}} \Big]_0^x \\ &= \alpha\beta \Big[\frac{-1}{\alpha\beta} e^{-\alpha x^{\beta}} + \frac{1}{\alpha\beta} \Big] \\ &= \frac{\alpha\beta}{\alpha\beta} \big(- e^{-\alpha x^{\beta}} + 1 \big) \\ &= 1 - e^{-\alpha x^{\beta}} \end{aligned}$$

Using this result, we can write the Cumulative Distribution Function as

$$F(x) = \begin{cases} 1 - e^{-\alpha x^{\beta}}, & 0 < x,\ 0 < \alpha,\ 0 < \beta \\ 0 & otherwise \end{cases}$$

43.3 Expected Values

$$\begin{aligned}
E(X) &= \int_0^\infty x\alpha\beta x^{\beta-1}e^{-\alpha x^\beta}dx \\
&= \alpha\beta\int_0^\infty x x^{\beta-1}e^{-\alpha x^\beta}dx \\
[1] &= \alpha\beta\int_0^\infty \Big((\frac{y}{\alpha})^{\frac{1}{\beta}}\Big)^\beta e^{-y}\frac{1}{\alpha\beta}\Big(\frac{y}{\alpha}\Big)^{\frac{1}{\beta}-1}dy \\
&= \frac{\alpha\beta}{\alpha\beta}\int_0^\infty \Big(\frac{y}{\alpha}\Big)^{\frac{\beta+1}{\beta}}\Big(\frac{y}{\alpha}\Big)^{\frac{1}{\beta}-1}e^{-y}dy \\
&= \int_0^\infty \Big(\frac{y}{\alpha}\Big)^{\frac{\beta+1}{\beta}-\frac{1}{\beta}-1}e^{-y}dy \\
&= \int_0^\infty \Big(\frac{y}{\alpha}\Big)^{\frac{\beta+1}{\beta}-1}e^{-y}dy \\
&= \frac{1}{\alpha^{\frac{\beta+1}{\beta}-\frac{\beta}{\beta}}}\int_0^\infty y^{\frac{\beta+1}{\beta}-1}e^{-y} \\
&= \alpha^{-\frac{1}{\beta}}\int_0^\infty y^{\frac{\beta+1}{\beta}-1}e^{-y} \\
[2] &= \alpha^{-\frac{1}{\beta}}\Gamma\Big(\frac{\beta+1}{\beta}\Big)
\end{aligned}$$

1. $y = \alpha x^\beta \ \Rightarrow x = (\frac{y}{\alpha})^{\frac{1}{\beta}} \ \Rightarrow dx = \frac{1}{\alpha\beta}(\frac{y}{\alpha})^{\frac{1}{\beta}-1}$
2. $\int_0^\infty x^{\alpha-1}e^{-x}dx = \Gamma(\alpha)$

$$
\begin{aligned}
E(X^2) &= \alpha\beta \int_0^\infty x^2 x^{\beta-1} e^{-\alpha x^\beta} dx \\
&= \alpha\beta \int_0^\infty x^{\beta+1} e^{-\alpha x^\beta} dx \\
[1] &= \alpha\beta \int_0^\infty \Big((\frac{y}{\alpha})^{\frac{1}{\beta}} \Big)^{\beta+1} e^{-y} \frac{1}{\alpha\beta} (\frac{y}{\alpha})^{\frac{1}{\beta}-1} dy \\
&= \frac{\alpha\beta}{\alpha\beta} \int_0^\infty \Big(\frac{y}{\alpha}\Big)^{\frac{\beta+1}{\beta}} \Big(\frac{y}{\alpha}\Big)^{\frac{1}{\beta}-1} e^{-y} dy \\
&= \int_0^\infty \Big(\frac{y}{\alpha}\Big)^{\frac{\beta+1}{\beta}+\frac{1}{\beta}-1} e^{-y} dy \\
&= \frac{1}{\alpha^{\frac{\beta+2}{\beta}-\frac{\beta}{\beta}}} \int_0^\infty y^{\frac{\beta+2}{\beta}-1} e^{-y} dy \\
&= \alpha^{-\frac{2}{\beta}} \int_0^\infty y^{\frac{\beta+2}{\beta}-1} e^{-y} dy \\
[2] &= \alpha^{-\frac{2}{\beta}} \Gamma\Big(\frac{\beta+2}{\beta}\Big)
\end{aligned}
$$

1. $y = \alpha x^\beta \ \Rightarrow x = (\frac{y}{\alpha})^{\frac{1}{\beta}} \ \Rightarrow dx = \frac{1}{\alpha\beta}(\frac{y}{\alpha})^{\frac{1}{\beta}-1}$
2. $\int_0^\infty x^{\alpha-1} e^{-x} dx = \Gamma(\alpha)$

$$
\begin{aligned}
\mu &= E(X) \\
&= \alpha^{-\frac{1}{\beta}} \Gamma\Big(\frac{\beta+1}{\beta}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= E(X^2) - E(X)^2 \\
&= \alpha^{-\frac{2}{\beta}} \Gamma\Big(\frac{\beta+2}{\beta}\Big) - \alpha^{-\frac{2}{\beta}} \Gamma\Big(\frac{\beta+1}{\beta}\Big)^2 \\
&= \alpha^{-\frac{2}{\beta}} \Big[\Gamma\Big(\frac{\beta+2}{\beta}\Big) - \Gamma\Big(\frac{\beta+1}{\beta}\Big)^2 \Big]
\end{aligned}
$$

43.4 Theorems for the Weibull Distribution

43.4.1 Validity of the Distribution

$$
\int_0^\infty \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} dx = 1
$$

Proof:

$$
\begin{aligned}
\int_0^\infty \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} dx &= \alpha\beta \int_0^\infty x^{\beta-1} e^{-\alpha x^\beta} dx \\
{}^{[1]} &= \alpha\beta \int_0^\infty \Big(\big(\frac{y}{\alpha}\big)^{\frac{1}{\beta}} \Big)^{\beta-1} e^{-y} \big(\frac{y}{\alpha}\big)^{\frac{1}{\beta}-1} \frac{1}{\alpha\beta} dy \quad = \frac{\alpha\beta}{\alpha\beta} \int_0^\infty \big(\frac{y}{\alpha}\big)^{\frac{\beta}{-1}} \big(\frac{y}{\alpha}\big)^{\frac{1}{\beta}-1} e^{-y} dy \\
&= \int_0^\infty \big(\frac{y}{\alpha}\big)^{\frac{\beta-1}{\beta}+\frac{1-\beta}{\beta}} e^{-y} dy \\
&= \int_0^\infty \frac{y^0}{\alpha^0} e^{-y} dy \\
&= \int_0^\infty y^{1-1} e^{-y} dy \\
{}^{[2]} &= \Gamma(1) = 1
\end{aligned}
$$

1. $y = \alpha x^\beta \ \Rightarrow x = (\frac{y}{\alpha})^{\frac{1}{\beta}} \ \Rightarrow dx = \frac{1}{\alpha\beta}(\frac{y}{\alpha})^{\frac{1}{\beta}-1}$
2. $\int_0^\infty x^{\alpha-1} e^{-x} dx = \Gamma(\alpha)$

Chapter 44

Z-test of Sample Proportions

44.1 One-Sample Z-test

The z-test of proportions is one approach used to look for evidence that the proportion of a sample may differ from a hypothesized (or previously observed) value. It assumes a normal distribution approximation to a Binomial distribution.

44.1.1 Z-Statistic

The z-statistic is a standardized measure of the magnitude of difference between a sample's proportion and some known, non-random constant.

44.1.2 Definitions and Terminology

Let p be a sample proportion from a sample. Let π_0 be a constant. z is defined:

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}$$

44.1.3 Hypotheses

The hypotheses for these test take the forms:

For a two-sided test:

$$H_0 : \pi = \pi_0$$
$$H_a : \pi \neq \pi_0$$

For a one-sided test:

$$H_0 : \pi < \pi_0$$
$$H_a : \pi \geq \pi_0$$

or

$$H_0 : \pi > \pi_0$$
$$H_a : \pi \leq \pi_0$$

To compare a sample $(X_1, \ldots, X_n)$ against the hypothesized value, a Z-statistic is calculated in the form:

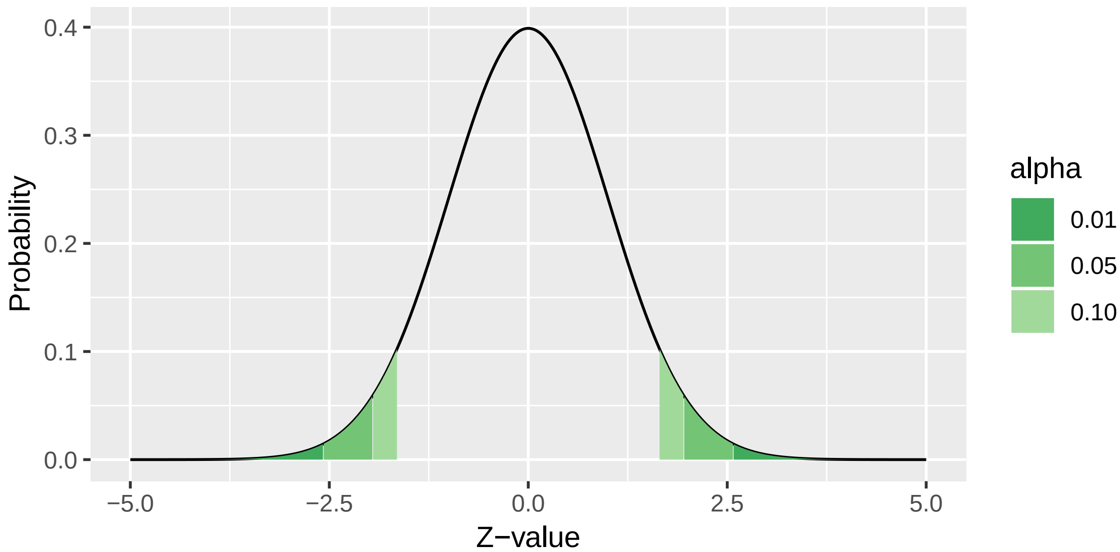

Figure 44.1: Rejection regions for the Z-test of proportions

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}$$

Where p is the sample proportion.

44.1.4 Decision Rule

The decision to reject a null hypothesis is made when an observed Z-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of shaded areas on both sides equaling the corresponding α. It follows, then, that the decision rule is:

Reject H_0 when $Z \leq z_{\alpha/2}$ or when $Z \geq z_{1-\alpha/2}$.

By taking advantage of the symmetry of the Z-distribution, we can simplify the decision rule to:

Reject H_0 when $|Z| \geq z_{1-\alpha/2}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

In the one-sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follows, then, that the decision rule for a lower tailed test is:

Reject H_0 when $Z \leq z_{\alpha,\nu}$.

For an upper tailed test, the decision rule is:

Reject H_0 when $Z \geq z_{1-\alpha,\nu}$.

Using the symmetry of the Z-distribution, we can simplify the decision rule as:

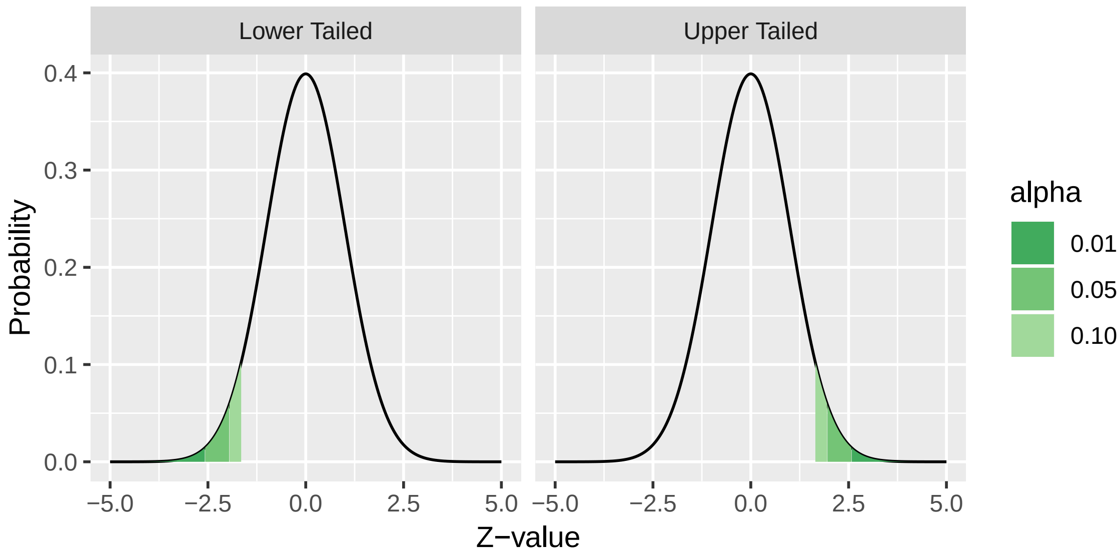

Figure 44.2: Rejection regions for one-tailed Z-test

Reject H_0 when $|Z| \geq z_{1-\alpha}$.

```
## Warning: Ignoring unknown aesthetics: ymax

## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of p:

Reject H_0 when $p \leq \pi_0 - z_\alpha \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}$ or $p \geq \pi_0 + z_\alpha \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}$.

This change can be justified by:

$$|Z| \geq z_{1-\alpha}$$
$$\left|\frac{p-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}\right| \geq z_{1-\alpha}$$

$$-\left(\frac{p-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}\right) \geq z_{1-\alpha} \qquad \frac{p-\pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}} \geq z_{1-\alpha}$$
$$p-\pi_0 \leq -z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}} \qquad p-\pi_0 \geq z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}$$
$$p \leq \pi_0 - z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}} \qquad p \geq \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}$$

For a two-sided test, both the conditions apply. The left side condition is used for a left-tailed test, and the right side condition for a right-tailed test.

44.1.5 Power

The derivations below make use of the following symbols:

- p: The sample proportion
- n: The sample size

- π_0: The value of population mean under the null hypothesis
- π_a: The value of the population mean under the alternative hypothesis.
- α: The significance level
- $\gamma(\mu)$: The power of the test for the parameter μ.
- z_α: A quantile of the Standard Normal distribution for a probability, α.
- Z: A calculated value to be compared against a Standard Normal distribution.
- C: The critical region (rejection region) of the test.

Two-Sided Test

$$\begin{aligned}
\gamma(\pi_a) &= P_{\pi_a}(p \in C) \\
&= P_\mu\Big(p \leq \pi_0 - z_{\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}\Big) + P_{\pi_a}\Big(p \geq \pi_0 + z_{1-\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}\Big) \\
&= P_{\pi_a}\Big(p - \pi_a \leq \pi_0 - \pi_a - z_{\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}\Big) + \\
&\quad P_{\pi_a}\Big(p - \pi_a \geq \pi_0 - \pi_a + z_{1-\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}\Big) \\
&= P_{\pi_a}\Big(\frac{p - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}} \leq \frac{\pi_0 - \pi_a - z_{\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big) + \\
&\quad P_{\pi_a}\Big(\frac{p - \mu}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}} \geq \frac{\pi_0 - \pi_a + z_{1-\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big) \\
&= P_{\pi_a}\Big(Z \leq \frac{\pi_0 - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}} - z_{\alpha/2}\Big) + P_{\pi_a}\Big(Z \geq \frac{\pi_0 - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}} + z_{1-\alpha/2}\Big) \\
&= P_{\pi_a}\Big(Z \leq -z_{\alpha/2} + \frac{\pi_0 - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big) + P_{\pi_a}\Big(Z \geq z_{1-\alpha/2} + \frac{\pi_0 - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big) \\
&= P_{\pi_a}\Big(Z \leq -z_{\alpha/2} + \frac{\sqrt{n} \cdot (\pi_0 - \pi_a)}{\sqrt{\pi_a \cdot (1-\pi_a)}}\Big) + P_{\pi_a}\Big(Z \geq z_{1-\alpha/2} + \frac{\sqrt{n} \cdot (\pi_0 - \pi_a)}{\sqrt{\pi_a \cdot (1-\pi_a)}}\Big)
\end{aligned}$$

Both $z_{\alpha/2}$ and $z_{1-\alpha/2}$ have Standard Normal distributions.

One-Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|Z| \geq z_{1-\alpha}$. Thus, where Z occurs in the derivation below, it may reasonably be replaced with $|Z|$.

$$\begin{aligned}
\gamma(\pi_a) &= P_{\pi_a}(p \in C) \\
&= P_{\pi_a}\big(p \geq \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}\big) \\
&= P_{\pi_a}\big(p - \pi_a \geq \pi_0 - \pi_a + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}\big) \\
&= P_{\pi_a}\Big(\frac{p-\pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}} \geq \frac{\pi_0 - \pi_a + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big) \\
&= P_{\pi_a}\Big(Z \geq \frac{\pi_0 - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}} + z_{1-\alpha}\Big) \\
&= P_{\pi_a}\Big(Z \geq z_{1-\alpha} + \frac{\pi_0 - \pi_a}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big) \\
&= P_{\pi_a}\Big(Z \geq z_{1-\alpha} + \frac{\sqrt{n} \cdot (\pi_0 - \pi_a)}{\sqrt{\frac{\pi_a \cdot (1-\pi_a)}{n}}}\Big)
\end{aligned}$$

Where $z_{1-\alpha}$ has a Standard Normal distribution.

44.1.6 Confidence Interval

The confidence interval for θ is written:

$$p \pm z_{1-\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = z_{1-\alpha/2} \cdot \sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}$$

44.2 References

1. Wackerly, Mendenhall, Scheaffer, *Mathematical Statistics with Applications*, 6th ed., Duxbury, 2002, ISBN 0-534-37741-6.
2. Daniel, *Biostatistics*, 8th ed., John Wiley & Sons, Inc., 2005, ISBN: 0-471-45654-3.
3. Hogg, McKean, Craig, *Introduction to Mathematical Statistics*, 6th ed., Pearson, 2005, ISBN: 0-13-008507-3

Chapter 45

Z-test

45.1 One-Sample Z-test

The t-test is commonly used to look for evidence that the mean of a normally distributed random variable may differ from a hypothesized (or previously observed) value.

45.1.1 Z-Statistic

The z-statistic is a standardized measure of the magnitude of difference between a sample's mean and some known, non-random constant. Calculation of the z-statistic assumes knowledge of the population variance.

45.1.2 Definitions and Terminology

Let $\bar{x}$ be a sample mean from a sample with standard deviation σ. Let μ_0 be a constant, and σ be the population standard deviation. z is defined:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

45.1.3 Hypotheses

The hypotheses for these test take the forms:

For a two-sided test:

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

For a one-sided test:

$$H_0 : \mu \leq \mu_0$$
$$H_a : \mu > \mu_0$$

or

$$H_0 : \mu \geq \mu_0$$
$$H_a : \mu < \mu_0$$

To compare a sample $(X_1, \ldots, X_n)$ against the hypothesized value, a Z-statistic is calculated in the form:

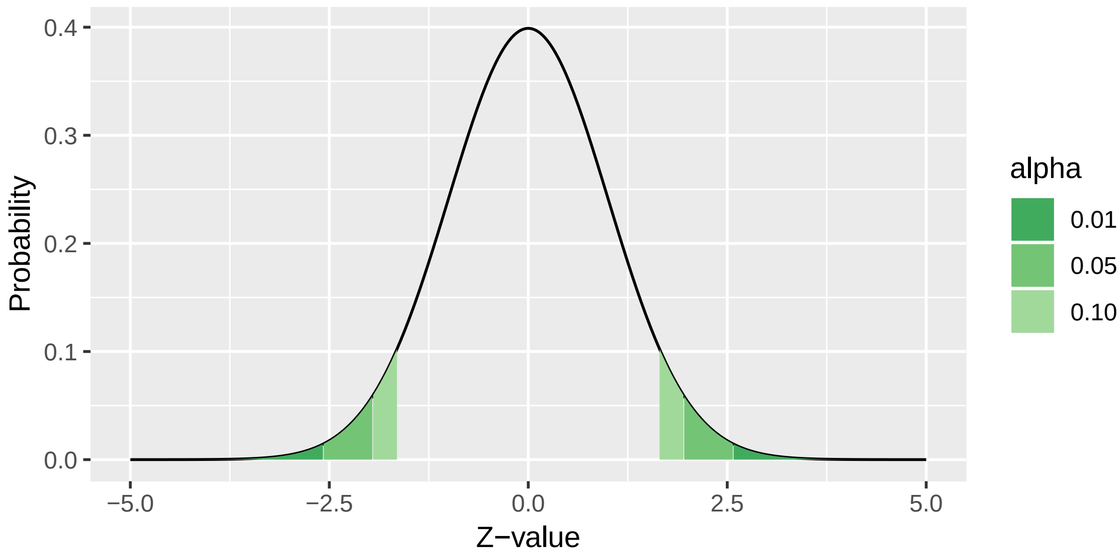

Figure 45.1: Critical regions for the two-sided Z-Test

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Where $\bar{x}$ is the sample mean and σ is the population standard deviation.

45.1.4 Decision Rule

The decision to reject a null hypothesis is made when an observed Z-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of shaded areas on both sides equaling the corresponding α. It follows, then, that the decision rule is:

Reject H_0 when $Z \leq z_{\alpha/2}$ or when $Z \geq z_{1-\alpha/2}$.

By taking advantage of the symmetry of the Z-distribution, we can simplify the decision rule to:

Reject H_0 when $|Z| \geq z_{1-\alpha/2}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

In the one-sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follows, then, that the decision rule for a lower tailed test is:

Reject H_0 when $Z \leq z_{\alpha}$.

For an upper tailed test, the decision rule is:

Reject H_0 when $Z \geq z_{1-\alpha}$.

Using the symmetry of the Z-distribution, we can simplify the decision rule as:

Reject H_0 when $|Z| \geq z_{1-\alpha}$.

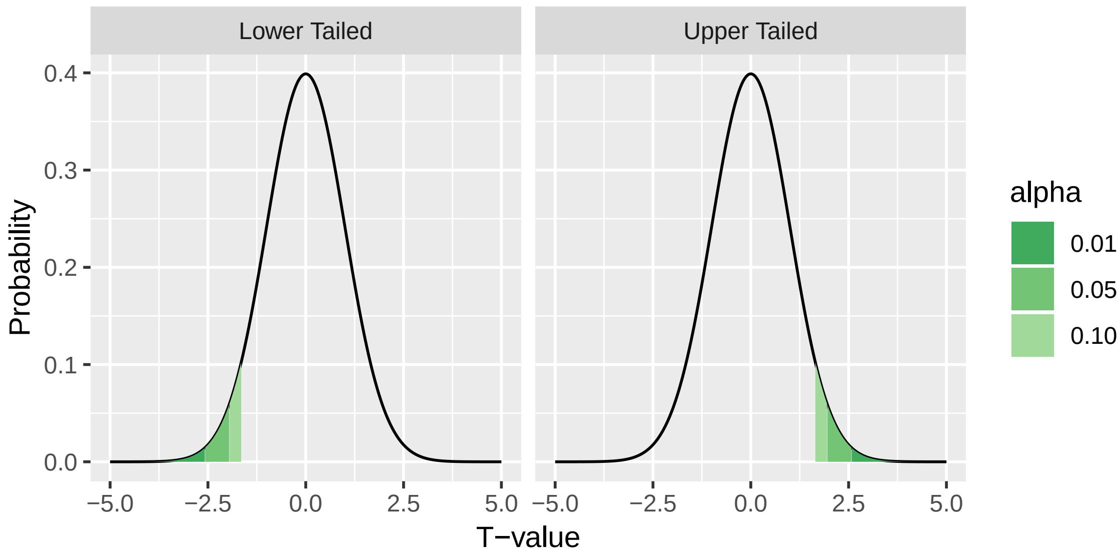

Figure 45.2: Critical regions for the two-sided Z-Test

```
## Warning: Ignoring unknown aesthetics: ymax

## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of $\bar{x}$:

Reject H_0 when $\bar{x} \leq \mu_0 - z_\alpha \cdot \sigma/\sqrt{n}$ or $\bar{x} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$.

This change can be justified by:

$$
\begin{aligned}
|Z| &\geq z_{1-\alpha} \\
\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| &\geq z_{1-\alpha}
\end{aligned}
$$

$$
\begin{aligned}
-\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) &\geq z_{1-\alpha} & \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} &\geq z_{1-\alpha} \\
\bar{x}-\mu_0 &\leq -z_{1-\alpha} \cdot \sigma/\sqrt{n} & \bar{x}-\mu_0 &\geq z_{1-\alpha} \cdot \sigma/\sqrt{n} \\
\bar{x} &\leq \mu_0 - z_{1-\alpha} \cdot \sigma/\sqrt{n} & \bar{x} &\geq \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}
\end{aligned}
$$

For a two-sided test, both the conditions apply. The left side condition is used for a left-tailed test, and the right side condition for a right-tailed test.

45.1.5 Power

The derivations below make use of the following symbols:

- $\bar{x}$: The sample mean
- σ: The population standard deviation
- n: The sample size
- μ_0: The value of population mean under the null hypothesis
- μ_a: The value of the population mean under the alternative hypothesis.
- α: The significance level
- $\gamma(\mu)$: The power of the test for the parameter μ.

- $t_{\alpha,\nu}$: A quantile of the central t-distribution for a probability, α and $n-1$ degrees of freedom.
- T: A calculated value to be compared against a t-distribution.
- C: The critical region (rejection region) of the test.

Two-Sided Test

$$\begin{aligned}
\gamma(\mu_a) &= P_{\mu_a}(\bar{x} \in C) \\
&= P_{\mu}\big(\bar{x} \leq \mu_0 - z_{\alpha/2} \cdot \sigma/\sqrt{n}\big) + P_{\mu_a}\big(\bar{x} \geq \mu_0 + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\big(\bar{x} - \mu_a \leq \mu_0 - \mu_a - z_{\alpha/2} \cdot \sigma/\sqrt{n}\big) + P_{\mu_a}\big(\bar{x} - \mu_a \geq \mu_0 - \mu_a + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\Big(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu_a - z_{\alpha/2} \cdot \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\Big) + P_{\mu_a}\Big(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu_a + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \leq \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} - z_{\alpha/2}\Big) + P_{\mu_a}\Big(Z \geq \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} + z_{1-\alpha/2}\Big) \\
&= P_{\mu_a}\Big(Z \leq -z_{\alpha/2} + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}\Big) + P_{\mu_a}\Big(Z \geq z_{1-\alpha/2} + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \leq -z_{\alpha/2} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{\sigma}\Big) + P_{\mu_a}\Big(Z \geq z_{1-\alpha/2} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{\sigma}\Big)
\end{aligned}$$

** This sentence needs work ** Both $z_{\alpha/2}$ and $z_{1-\alpha/2}$ have normal distributions with non-centrality parameter $\frac{\sqrt{n}(\mu_0 - \mu_a)}{\sigma}$.

One-Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|Z| \geq z_{1-\alpha}$. Thus, where Z occurs in the derivation below, it may reasonably be replaced with $|Z|$.

$$\begin{aligned}
\gamma(\mu_a) &= P_{\mu_a}(\bar{x} \in C) \\
&= P_{\mu_a}\big(\bar{x} \geq \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\big(\bar{x} - \mu_a \geq \mu_0 - \mu_a + z_{1-\alpha} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\Big(\frac{\bar{x} - \mu_a}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu_a + z_{1-\alpha} \cdot \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \geq \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} + z_{1-\alpha}\Big) \\
&= P_{\mu_a}\Big(Z \geq z_{1-\alpha} + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \geq z_{1-\alpha} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{\sigma}\Big)
\end{aligned}$$

** This sentence is not accurate ** Where $z_{1-\alpha} + \frac{\sqrt{n}(\mu_0 - \mu_a)}{\sigma}$ has a non-central t-distribution with non-centrality parameter $\frac{\sqrt{n}(\mu_0 - \mu_a)}{\sigma}$

45.1.6 Confidence Interval

The confidence interval for θ is written:

$$\bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

45.2 Two-Sample T-test

The two sample t-test is commonly used to look for evidence that the mean of one normally distributed random variable may differ from that of another normally distributed random variable. The hypotheses for this test take the forms:

45.2.1 T-Statistic

The t-statistic is a standardize measure of the magnitude of difference between two sample means and some known, non-random difference of population means. It is similar to a two sample z-statistic, but differes in that a t-statistic may be calculated without knowledge of the population variances.

45.2.2 Definitions and Terminology

Let $\bar{x_1}$ and $\bar{x_2}$ be sample means from two independent samples with standard deviations s_1 and s_2. Let μ_1 and μ_2 be constants representing the means of the populations from which $\bar{x_1}$ and $\bar{x_2}$ obtained. z is defined:

$$z = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}$$

Where

$$SE^* = \begin{cases} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, & \sigma_1^2 \neq \sigma_2^2 \\ \sqrt{\frac{(n_1-1)\cdot\sigma_1^2+(n_2-1)\cdot\sigma_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, & \sigma_1^2 = \sigma_2^2 \end{cases}$$

45.2.3 Hypotheses

For a two-sided test:

$$H_0 : \mu_1 = \mu_2 H_a : \mu_1 \neq \mu_2$$

For a one-sided test:

$$H_0 : \mu_1 \leq \mu_2 H_a : \mu_1 > \mu_2$$

or

$$H_0 : \mu_1 \geq \mu_1 H_a : \mu_1 < \mu_1$$

45.2.4 Decision Rule

The decision to reject a null hypothesis is made when an observed T-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

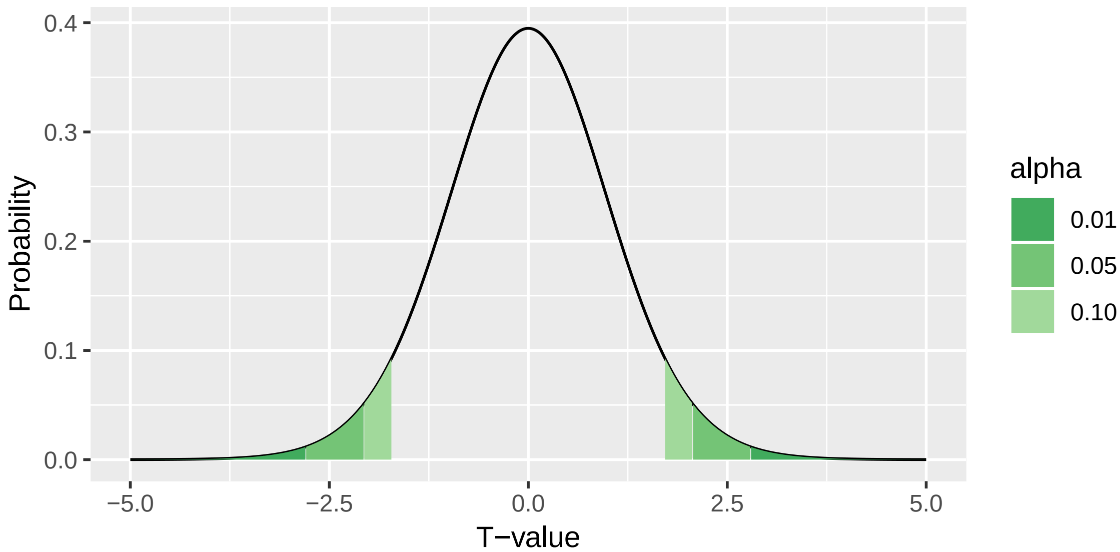

Figure 45.3: The example displayed uses 25 degrees of freedom

45.2.4.1 Two Sided Test

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of the shaded areas on both sides equally the corresponding α. It follows then that the decision rule is:

Reject H_0 when $Z \leq z_{\alpha/2}$ or when $Z \geq z_{1-\alpha/2}$.

By taking advantage of the symmetry of the Z-distribution, we can simplify the decision rule to:

Reject H_0 when $|Z| \geq z_{1-\alpha/2}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

45.3 One Sided Test

In the one sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follow, then, that the decision rule for a lower tailed test is:

Reject H_0 when $T \leq t_{\alpha,\nu}$.

For an upper tailed test, the decision rule is:

Reject H_0 when $T \geq t_{1-\alpha,\nu}$.

Using the symmetry of the T-distribution, we can simplify the decision rule as:

Reject H_0 when $|T| \geq t_{1-\alpha,\nu}$.

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of $\bar{x_1}$ and $\bar{x_2}$.

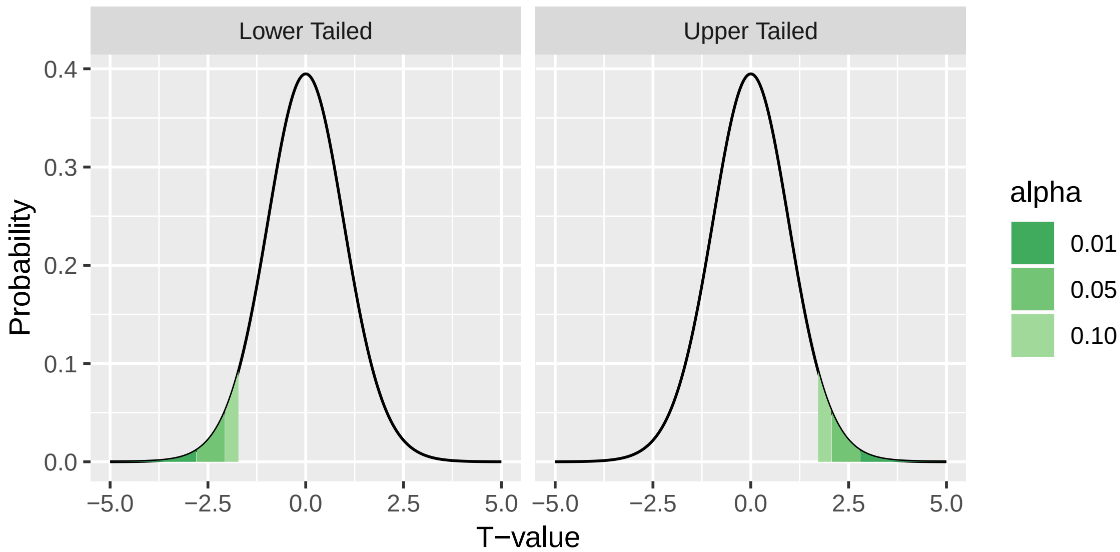

Figure 45.4: The example displayed uses 25 degrees of freedom

Reject H_0 when $\bar{x_1} - \bar{x_2} \leq (\mu_1 - \mu_2) - t_{\alpha,\nu} \cdot SE^*$ or $\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + t_{\alpha,\nu} \cdot SE^*$

This change can be justified by:

$$
\begin{aligned}
|T| &\geq t_{1-\alpha,\nu} \\
\left|\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right| &\geq t_{1-\alpha,\nu}
\end{aligned}
$$

$$
\begin{aligned}
-\left(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right) &\geq t_{1-\alpha,\nu} \\
(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2) &\leq -t_{1-\alpha,\nu} \cdot SE^* \\
\bar{x_1} - \bar{x_2} &\leq (\mu_1 - \mu_2) - t_{1-\alpha,\nu} \cdot SE^*
\end{aligned}
\qquad
\begin{aligned}
\left(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right) &\geq t_{1-\alpha,\nu} \\
(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2) &\geq t_{1-\alpha,\nu} \cdot SE^* \\
\bar{x_1} - \bar{x_2} &\leq (\mu_1 - \mu_2) + t_{1-\alpha,\nu} \cdot SE^*
\end{aligned}
$$

45.3.1 Power

Two Sided Test

$$\begin{aligned}
\gamma(\mu_{1a} - \mu_{2a}) &= P_{\mu_{1a}-\mu_{2a}}(\bar{x} \in C) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) \leq (\mu_1 - \mu_2) - t_{\alpha/2,\nu} \cdot SE^*\big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + t_{1-\alpha/2,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a}) \leq (\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) - t_{\alpha/2,\nu} \cdot SE^*\big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a}) \geq (\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha/2,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*} \leq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) - t_{\alpha/2,\nu} \cdot SE^*}{SE^*}\Big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a})}{SE^*} \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha/2,\nu} \cdot SE^*}{SE^*}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \leq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*} - t_{\alpha/2,\nu}\Big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(T \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*} + t_{1-\alpha/2,\nu}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \leq -t_{\alpha/2,\nu} + \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*}\Big) + \\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(T \geq t_{1-\alpha/2,\nu} + \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*}\Big)
\end{aligned}$$

Both $-t_{\alpha/2,\nu} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ and $t_{1-\alpha/2,\nu} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ have non-central T-distributions with non-centrality parameter $\frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$.

One Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|T| \geq t_{1-\alpha}$. Thus, where T occurs in the derivation below, it may reasonably be replaced with $|T|$.

$$\begin{aligned}
\gamma(\mu_{1a} - \mu_{2a}) &= P_{\mu_{1a}-\mu_{2a}}(\bar{x} \in C) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + t_{1-\alpha,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a}) \geq (\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha,\nu} \cdot SE^*\big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_{1a} - \mu_{2a})}{SE^*} \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a}) + t_{1-\alpha,\nu} \cdot SE^*}{SE^*}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \geq \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*} + t_{1-\alpha,\nu}\Big) \\
&= P_{\mu_{1a}-\mu_{2a}}\Big(T \geq t_{1-\alpha,\nu} + \frac{(\mu_1 - \mu_2) - (\mu_{1a} - \mu_{2a})}{SE^*}\Big)
\end{aligned}$$

$t_{1-\alpha/2,\nu} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ has a non-central T-distribution with non-centrality parameter $\frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$.

45.3.2 Confidence Interval

The confidence interval for $\mu_1 - \mu_2$ is written:

$$(\bar{x_1} - \bar{x_2}) \pm t_{1-\alpha/2} \cdot SE^*$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = t_{1-\alpha/2} \cdot SE^*$$

45.4 References

1. Wackerly, Mendenhall, Scheaffer, *Mathematical Statistics with Applications*, 6th ed., Duxbury, 2002, ISBN 0-534-37741-6.
2. Daniel, *Biostatistics*, 8th ed., John Wiley & Sons, Inc., 2005, ISBN: 0-471-45654-3.
3. Hogg, McKean, Craig, *Introduction to Mathematical Statistics*, 6th ed., Pearson, 2005, ISBN: 0-13-008507-3
4. Wikipedia, "Student's T test", https://en.wikipedia.org/wiki/Student%27s_t-test
5. Wikipedia, "Welch-Satterthwaite Equation", https://en.wikipedia.org/wiki/Welch%E2%80%93Satterthwaite_equation

Chapter 46

Z-test of Sample Means

46.1 One-Sample Z-test

The z-test is commonly used to look for evidence that the mean of a normally distributed random variable may differ from a hypothesized (or previously observed) value. It's utility is limited, as it assume knowledge of the population standard deviation, which is rarely known.

46.1.1 Z-Statistic

The z-statistic is a standardized measure of the magnitude of difference between a sample's mean and some known, non-random constant.

46.1.2 Definitions and Terminology

Let $\bar{x}$ be a sample mean from a sample with known population standard deviation σ. Let μ_0 be a constant, and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ be the standard error of the parameter $\bar{x}$. z is defined:

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

46.1.3 Hypotheses

The hypotheses for these test take the forms:

For a two-sided test:

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

For a one-sided test:

$$H_0 : \mu < \mu_0$$
$$H_a : \mu \geq \mu_0$$

or

$$H_0 : \mu > \mu_0$$
$$H_a : \mu \leq \mu_0$$

To compare a sample $(X_1, \ldots, X_n)$ against the hypothesized value, a T-statistic is calculated in the form:

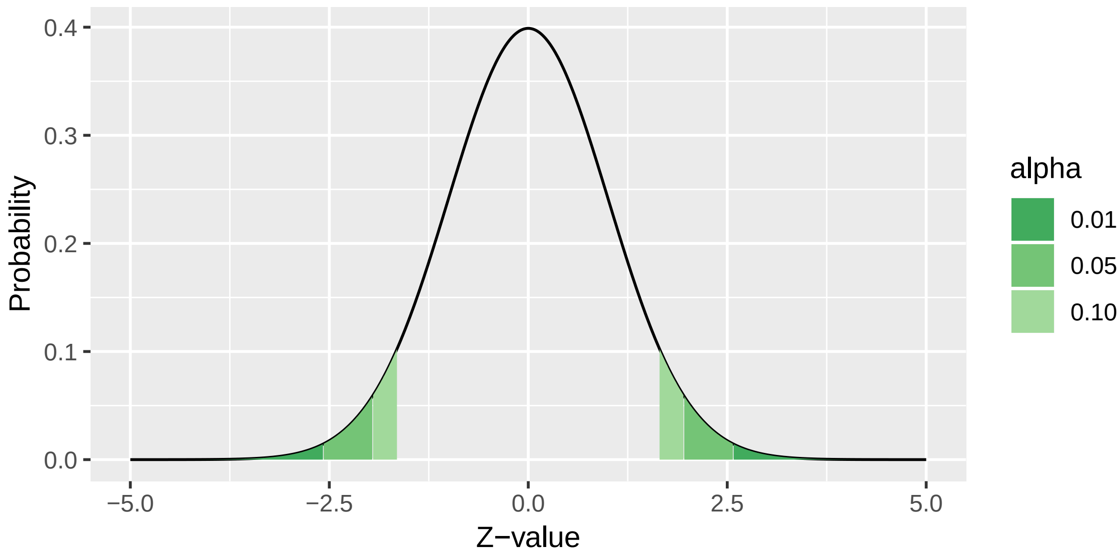

Figure 46.1: Rejection regions for the Z-test

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Where $\bar{x}$ is the sample mean and σ is the population standard deviation.

46.1.4 Decision Rule

The decision to reject a null hypothesis is made when an observed T-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of shaded areas on both sides equaling the corresponding α. It follows, then, that the decision rule is:

Reject H_0 when $Z \leq z_{\alpha/2}$ or when $Z \geq z_{1-\alpha/2}$.

By taking advantage of the symmetry of the Z-distribution, we can simplify the decision rule to:

Reject H_0 when $|Z| \geq z_{1-\alpha/2}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

In the one-sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follows, then, that the decision rule for a lower tailed test is:

Reject H_0 when $Z \leq z_{\alpha,\nu}$.

For an upper tailed test, the decision rule is:

Reject H_0 when $Z \geq z_{1-\alpha,\nu}$.

Using the symmetry of the Z-distribution, we can simplify the decision rule as:

Reject H_0 when $|Z| \geq z_{1-\alpha}$.

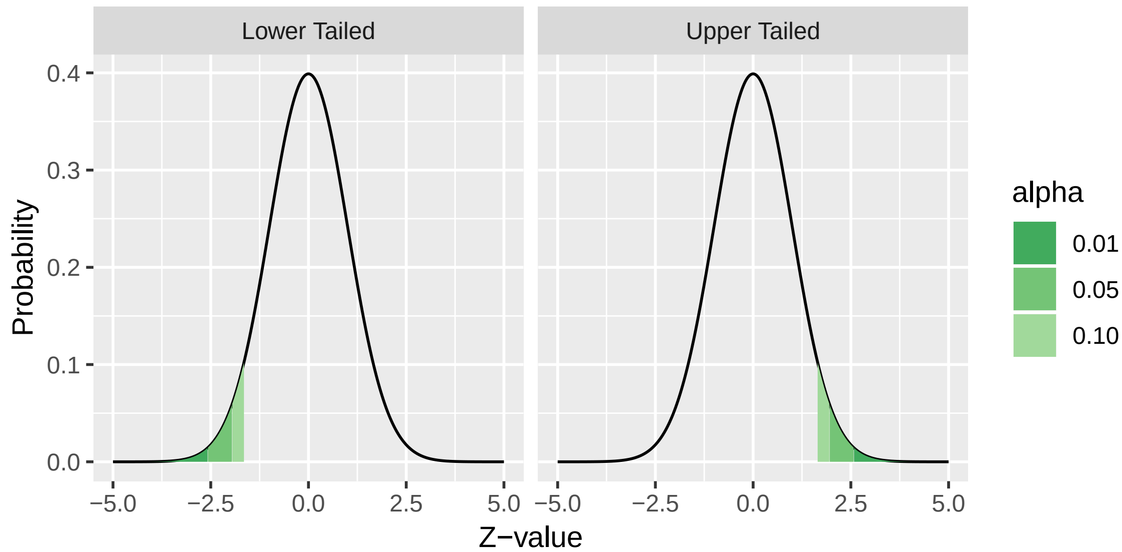

Figure 46.2: Rejection regions for one-tailed Z-test

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of $\bar{x}$:

Reject H_0 when $\bar{x} \leq \mu_0 - z_\alpha \cdot \sigma/\sqrt{n}$ or $\bar{x} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$.

This change can be justified by:

$$|Z| \geq z_{1-\alpha}$$

$$\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| \geq z_{1-\alpha}$$

$$\begin{aligned} -\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) &\geq z_{1-\alpha} & \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} &\geq z_{1-\alpha} \\ \bar{x} - \mu_0 &\leq -z_{1-\alpha} \cdot \sigma/\sqrt{n} & \bar{x} - \mu_0 &\geq z_{1-\alpha} \cdot \sigma/\sqrt{n} \\ \bar{x} &\leq \mu_0 - z_{1-\alpha} \cdot \sigma/\sqrt{n} & \bar{x} &\geq \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n} \end{aligned}$$

For a two-sided test, both the conditions apply. The left side condition is used for a left-tailed test, and the right side condition for a right-tailed test.

46.1.5 Power

The derivations below make use of the following symbols:

- $\bar{x}$: The sample mean
- s: The sample standard deviation
- n: The sample size
- μ_0: The value of population mean under the null hypothesis
- μ_a: The value of the population mean under the alternative hypothesis.
- α: The significance level
- $\gamma(\mu)$: The power of the test for the parameter μ.

- z_α: A quantile of the Standard Normal distribution for a probability, α.
- Z: A calculated value to be compared against a Standard Normal distribution.
- C: The critical region (rejection region) of the test.

Two-Sided Test

$$\begin{aligned}
\gamma(\mu_a) &= P_{\mu_a}(\bar{x} \in C) \\
&= P_\mu\big(\bar{x} \leq \mu_0 - z_{\alpha/2} \cdot \sigma/\sqrt{n}\big) + P_{\mu_a}\big(\bar{x} \geq \mu_0 + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\big(\bar{x} - \mu_a \leq \mu_0 - \mu_a - z_{\alpha/2} \cdot \sigma/\sqrt{n}\big) + P_{\mu_a}\big(\bar{x} - \mu_a \geq \mu_0 - \mu_a + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\Big(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu_a - z_{\alpha/2} \cdot \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\Big) + P_{\mu_a}\Big(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu_a + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \leq \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} - z_{\alpha/2}\Big) + P_{\mu_a}\Big(Z \geq \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} + z_{1-\alpha/2}\Big) \\
&= P_{\mu_a}\Big(Z \leq -z_{\alpha/2} + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}\Big) + P_{\mu_a}\Big(Z \geq z_{1-\alpha/2} + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \leq -z_{\alpha/2} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{\sigma}\Big) + P_{\mu_a}\Big(Z \geq z_{1-\alpha/2} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{\sigma}\Big)
\end{aligned}$$

Both $z_{\alpha/2}$ and $z_{1-\alpha/2}$ have Standard Normal distributions.

One-Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|Z| \geq z_{1-\alpha}$. Thus, where Z occurs in the derivation below, it may reasonably be replaced with $|Z|$.

$$\begin{aligned}
\gamma(\mu_a) &= P_{\mu_a}(\bar{x} \in C) \\
&= P_{\mu_a}\big(\bar{x} \geq \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\big(\bar{x} - \mu_a \geq \mu_0 - \mu_a + z_{1-\alpha} \cdot \sigma/\sqrt{n}\big) \\
&= P_{\mu_a}\Big(\frac{\bar{x} - \mu_a}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu_a + z_{1-\alpha} \cdot \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \geq \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}} + z_{1-\alpha}\Big) \\
&= P_{\mu_a}\Big(Z \geq z_{1-\alpha} + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}\Big) \\
&= P_{\mu_a}\Big(Z \geq t_{1-\alpha} + \frac{\sqrt{n} \cdot (\mu_0 - \mu_a)}{\sigma}\Big)
\end{aligned}$$

Where $z_{1-\alpha}$ has a Standard Normal distribution.

46.1.6 Confidence Interval

The confidence interval for θ is written:

$$\bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

46.2 Two-Sample Z-test

The two sample z-test is commonly used to look for evidence that the mean of one normally distributed random variable may differ from that of another normally distributed random variable. The hypotheses for this test take the forms:

46.2.1 Z-Statistic

The z-statistic is a standardized measure of the magnitude of difference between two sample means and some known, non-random difference of population means. It is similar to a two sample t-statistic, but differs in that a z-statistic may not be calculated without knowledge of the population variances.

46.2.2 Definitions and Terminology

Let $\bar{x_1}$ and $\bar{x_2}$ be sample means from two independent samples with standard deviations σ_1 and σ_2. Let μ_1 and μ_2 be constants representing the means of the populations from which $\bar{x_1}$ and $\bar{x_2}$ obtained. z is defined:

$$z = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}$$

Where

$$SE^* = \begin{cases} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, & \sigma_1^2 \neq \sigma_2^2 \\ \sqrt{\frac{(n_1-1)\cdot\sigma_1^2 + (n_2-1)\cdot\sigma_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, & \sigma_1^2 = \sigma_2^2 \end{cases}$$

46.2.3 Hypotheses

For a two-sided test:

$$H_0 : \mu_1 = \mu_2 H_a : \mu_1 \neq \mu_2$$

For a one-sided test:

$$H_0 : \mu_1 \leq \mu_2 H_a : \mu_1 > \mu_2$$

or

$$H_0 : \mu_1 \geq \mu_1 H_a : \mu_1 < \mu_1$$

46.2.4 Decision Rule

The decision to reject a null hypothesis is made when an observed Z-value lies in a critical region that suggests the probability of that observation is low. We define the critical region as the upper bound we are willing to accept for α, the Type I Error.

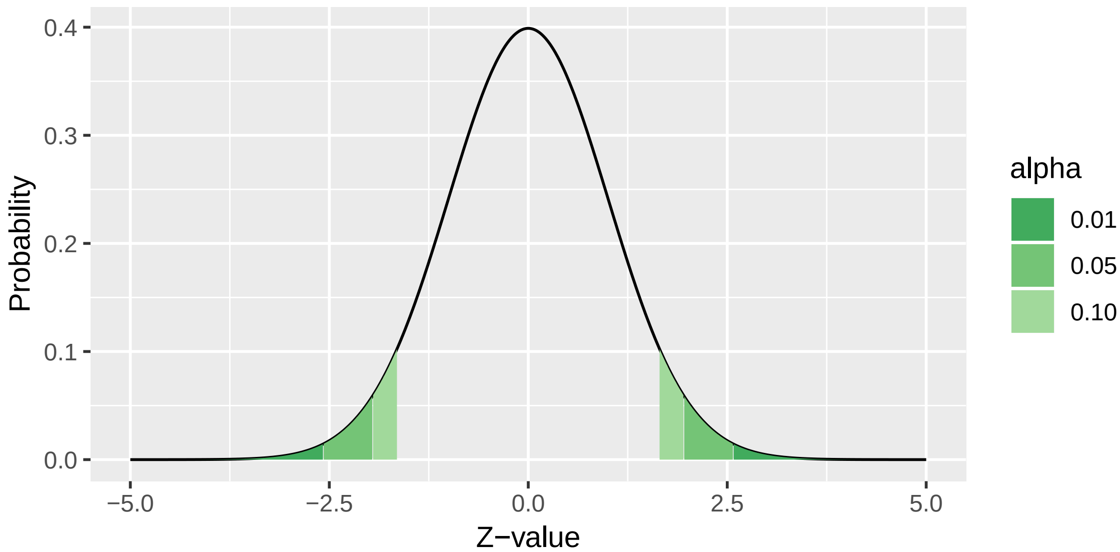

Figure 46.3: Rejection region for a two-tailed test

46.2.4.1 Two Sided Test

In the two-sided test, α is shared equally in both tails. The rejection regions for the most common values of α are depicted in the figure below, with the sum of the shaded areas on both sides equally the corresponding α. It follows then that the decision rule is:

Reject H_0 when $Z \leq z_{\alpha/2}$ or when $Z \geq z_{1-\alpha/2}$.

By taking advantage of the symmetry of the Z-distribution, we can simplify the decision rule to:

Reject H_0 when $|Z| \geq z_{1-\alpha/2}$

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

46.3 One Sided Test

In the one sided test, α is placed in only one tail. The rejection regions for the most common values of α are depicted in the figure below. In each case, α is the area in the tail of the figure. It follow, then, that the decision rule for a lower tailed test is:

Reject H_0 when $Z \leq z_{\alpha}$.

For an upper tailed test, the decision rule is:

Reject H_0 when $Z \geq z_{1-\alpha}$.

Using the symmetry of the Z-distribution, we can simplify the decision rule as:

Reject H_0 when $|Z| \geq z_{1-\alpha}$.

```
## Warning: Ignoring unknown aesthetics: ymax
```

```
## Warning: Ignoring unknown aesthetics: ymax
```

The decision rule can also be written in terms of $\bar{x_1}$ and $\bar{x_2}$.

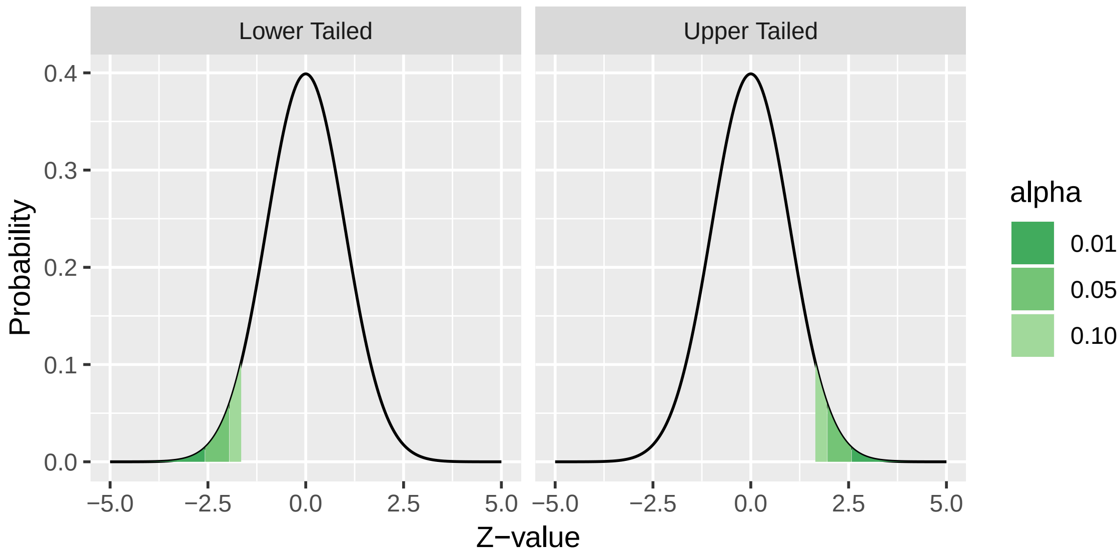

Figure 46.4: Rejection regions for one-tailed tests

Reject H_0 when $\bar{x_1} - \bar{x_2} \leq (\mu_1 - \mu_2) - z_\alpha \cdot SE^*$ or $\bar{x_1} - \bar{x_2} \geq (\mu_1 - \mu_2) + z_\alpha \cdot SE^*$

This change can be justified by:

$$
\begin{aligned}
|Z| &\geq z_{1-\alpha} \\
\left|\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\right| &\geq z_{1-\alpha}
\end{aligned}
$$

$$
\begin{aligned}
-\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\Big) &\geq z_{1-\alpha} \\
(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2) &\leq -z_{1-\alpha} \cdot SE^* \\
\bar{x_1} - \bar{x_2} &\leq (\mu_1 - \mu_2) - z_{1-\alpha} \cdot SE^*
\end{aligned}
\qquad
\begin{aligned}
\Big(\frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{SE^*}\Big) &\geq z_{1-\alpha} \\
(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2) &\geq z_{1-\alpha} \cdot SE^* \\
\bar{x_1} - \bar{x_2} &\leq (\mu_1 - \mu_2) + z_{1-\alpha} \cdot SE^*
\end{aligned}
$$

46.3.1 Power

Two Sided Test

$$\begin{aligned}
\gamma(\mu_{1a}-\mu_{2a}) &= P_{\mu_{1a}-\mu_{2a}}(\bar{x}\in C)\\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1}-\bar{x_2}) \leq (\mu_1-\mu_2) - z_{\alpha/2}\cdot SE^*\big)+\\
&\quad P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1}-\bar{x_2} \geq (\mu_1-\mu_2) + z_{1-\alpha/2}\cdot SE^*\big)\\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1}-\bar{x_2}) - (\mu_{1a}-\mu_{2a}) \leq (\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a}) - z_{\alpha/2}\cdot SE^*\big)+\\
&\quad P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1}-\bar{x_2}) - (\mu_{1a}-\mu_{2a}) \geq (\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a}) + z_{1-\alpha/2}\cdot SE^*\big)\\
&= P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1}-\bar{x_2}) - (\mu_1-\mu_2)}{SE^*} \leq \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a}) - z_{\alpha/2}\cdot SE^*}{SE^*}\Big)+\\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1}-\bar{x_2}) - (\mu_{1a}-\mu_{2a})}{SE^*} \geq \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a}) + z_{1-\alpha/2}\cdot SE^*}{SE^*}\Big)\\
&= P_{\mu_{1a}-\mu_{2a}}\Big(Z \leq \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a})}{SE^*} - z_{\alpha/2}\Big)+\\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(Z \geq \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a})}{SE^*} + z_{1-\alpha/2}\Big)\\
&= P_{\mu_{1a}-\mu_{2a}}\Big(Z \leq -z_{\alpha/2} + \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a})}{SE^*}\Big)+\\
&\quad P_{\mu_{1a}-\mu_{2a}}\Big(Z \geq z_{1-\alpha/2} + \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a})}{SE^*}\Big)
\end{aligned}$$

Both $-z_{\alpha/2} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ and $z_{1-\alpha/2} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ have Standard Normal distributions.

One Sided Test

For convenience, the power for only the upper tailed test is derived here.
Recall that the symmetry of the t-test allows us to use the decision rule: Reject H_0 when $|Z| \geq t_{1-\alpha}$. Thus, where Z occurs in the derivation below, it may reasonably be replaced with $|Z|$.

$$\begin{aligned}
\gamma(\mu_{1a}-\mu_{2a}) &= P_{\mu_{1a}-\mu_{2a}}(\bar{x}\in C)\\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1}-\bar{x_2} \geq (\mu_1-\mu_2) + z_{1-\alpha}\cdot SE^*\big)\\
&= P_{\mu_{1a}-\mu_{2a}}\big((\bar{x_1}-\bar{x_2}) - (\mu_{1a}-\mu_{2a}) \geq (\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a}) + z_{1-\alpha}\cdot SE^*\big)\\
&= P_{\mu_{1a}-\mu_{2a}}\Big(\frac{(\bar{x_1}-\bar{x_2}) - (\mu_{1a}-\mu_{2a})}{SE^*} \geq \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a}) + z_{1-\alpha}\cdot SE^*}{SE^*}\Big)\\
&= P_{\mu_{1a}-\mu_{2a}}\Big(Z \geq \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a})}{SE^*} + z_{1-\alpha}\Big)\\
&= P_{\mu_{1a}-\mu_{2a}}\Big(Z \geq z_{1-\alpha} + \frac{(\mu_1-\mu_2) - (\mu_{1a}-\mu_{2a})}{SE^*}\Big)
\end{aligned}$$

$z_{1-\alpha/2} + \frac{(\mu_1-\mu_2)-(\mu_{1a}-\mu_{2a})}{SE^*}$ has a Normal Distribution.

46.3.2 Confidence Interval

The confidence interval for $\mu_1 - \mu_2$ is written:

$$(\bar{x_1}-\bar{x_2}) \pm z_{1-\alpha/2}\cdot SE^*$$

The value of the expression on the right is often referred to as the *margin of error*, and we will refer to this value as

$$E = z_{1-\alpha/2}\cdot SE^*$$

46.4 References

1. Wackerly, Mendenhall, Scheaffer, *Mathematical Statistics with Applications*, 6th ed., Duxbury, 2002, ISBN 0-534-37741-6.
2. Daniel, *Biostatistics*, 8th ed., John Wiley & Sons, Inc., 2005, ISBN: 0-471-45654-3.
3. Hogg, McKean, Craig, *Introduction to Mathematical Statistics*, 6th ed., Pearson, 2005, ISBN: 0-13-008507-3

Chapter 47

References

Agresti, Alan. 1996. *An Introduction to Categorical Data Analysis.* John Wiley; Sons.

Brunette, John. 2003–2007a. "A Binomial Expansion Theorem with the Binomial Theorem as a Corrolary." Lecture Notes.

———. 2003–2007b. "Chebychev's Theorem for Absolute Deviation." Lecture Notes.

Hogg, Robert V., Joseph W. McKean, and Allen T. Craig. 2005. *Introduction to Mathematical Statistics.* 6th ed. Pearson.

Law, Averill M., and W. David Kelton. 2000. *Simulation Modeling and Analysis.* 3rd ed. McGraw-Hill.

Mantel, Nathan. 1963. "Chi-Square Tests with One Degree of Freedom; Extensions of the Mantel- Haenszel Procedure." *Journal of the American Statistical Association* 58 (303): 690–700.

McDonald, Stephen. 2003–2007. "An Axiomatic Approach to the Real Number System." Lecture Notes.

Montgomery, Douglas C. 2004. *Design and Analysis of Experiments.* 5th ed. John Wiley & Sons, Inc.

Wackerly, Dennis D., William Mendenhall III, and Richard L. Scheaffer. 2002. *Mathematical Statistics with Applications.* 6th ed. Duxbury.