Exercise Walkthrough: Treatment Efficacy and Simpson's Paradox

Justin Lanfermann

25. June 2025

Abstract

This document provides a detailed walkthrough for Exercise 4 from the Discrete Probability Theory script. We will analyze a simulated clinical study dataset to determine the effectiveness of a medical treatment. The exercise serves as a practical application of probability concepts and introduces the critical statistical phenomenon known as Simpson's Paradox. We will explore how the interpretation of a third variable—first as a confounder, then as a mediator—dramatically changes our conclusions. All steps are explained with reference to the concepts and definitions from the lecture script.

Contents

1 Overview of the Exercise

The core task is to answer the question: "Does the treatment help?" based on a dataset with three variables: Treatment (T), Survival (S), and a third variable which is initially a Precondition (P) and later a Complication (C).

This exercise demonstrates that a naive analysis can be deeply misleading. The relationship between treatment and survival can appear to be one thing when looking at the entire population, and the complete opposite when looking at specific subgroups. This reversal is a classic example of Simpson's Paradox [3]. Our goal is to use the tools from probability theory to correctly navigate this problem.

1.1 Setting up the Data

The exercise provides a dataset named 'hospital.csv'. Since we are creating a self-contained document, we will generate a dataset that exhibits the desired properties directly in our code. This data is constructed to show a strong paradoxical effect. We will use the `pandas` library for data manipulation and `numpy` for calculations, as suggested.

Here is the Python code to generate our data, which we will analyze in the following sections.

```python
import pandas as pd
import numpy as np
import io

# Data that exhibits Simpson's Paradox
# P=0: No Precondition, P=1: Precondition
# T=0: No Treatment, T=1: Treatment
# S=0: Did not survive, S=1: Survived

# We create a list of dictionaries, which is an easy way to build a DataFrame
data = (
    # Group 1: No Precondition (P=0)
    # Subgroup 1.1: No Treatment (T=0). High survival (90%).
    [{'T': 0, 'S': 1, 'P': 0}] * 360 +
    [{'T': 0, 'S': 0, 'P': 0}] * 40 +
    # Subgroup 1.2: Treatment (T=1). Even higher survival (95%).
    [{'T': 1, 'S': 1, 'P': 0}] * 95 +
    [{'T': 1, 'S': 0, 'P': 0}] * 5 +

    # Group 2: Precondition (P=1)
    # Subgroup 2.1: No Treatment (T=0). Very low survival (10%).
    [{'T': 0, 'S': 1, 'P': 1}] * 10 +
    [{'T': 0, 'S': 0, 'P': 1}] * 90 +
    # Subgroup 2.2: Treatment (T=1). Low, but better survival (20%).
    [{'T': 1, 'S': 1, 'P': 1}] * 80 +
    [{'T': 1, 'S': 0, 'P': 1}] * 320
)

# Create the DataFrame
df = pd.DataFrame(data)

# Shuffle the DataFrame to make it look like a real dataset
df = df.sample(frac=1, random_state=42).reset_index(drop=True)

print("First 5 rows of the dataset:")
print(df.head())
print(f"\nTotal number of patients: {len(df)}")
```

Listing 1: Python code to generate a DataFrame exhibiting Simpson's Paradox.

2 Part (i): Precondition as a Confounder

In this part, the third column ‘P‘ represents a pre-existing condition. A plausible causal relationship, as suggested in the script, is that the precondition influences both the doctor’s decision to administer the treatment and the patient’s ultimate chance of survival. This makes ‘P‘ a classic confounder [1].

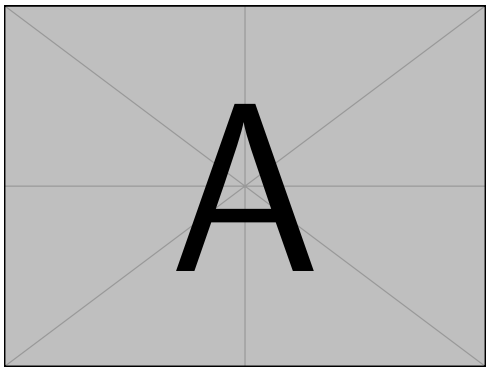

Figure 1: Causal graph with Precondition (P) as a confounder.

Image snippet from the exercise statement.

2.1 (a) Formalizing the Question

The Task: ”Please analyze whether the treatment helped.”

Formalization and Assumptions: The question asks about the **causal effect** of the treatment on survival. A simple, but often misleading, way to formalize this is to compare conditional probabilities. We can compare the probability of survival given treatment, $P(S = 1|T = 1)$, with the probability of survival given no treatment, $P(S = 1|T = 0)$. If the former is greater, we might conclude the treatment helps.

As the exercise note states, for a Bernoulli variable like ‘S‘, its expectation is equal to the probability of the event occurring (see Expectation of Bernoulli Variables [4]). Therefore, we are comparing $E[S|T = 1]$ and $E[S|T = 0]$. We estimate these quantities from our data using the sample mean.

Modeling Choice: The crucial modeling assumption is about the role of the variable ‘P‘. As shown in Figure 1, ‘P‘ is a common cause of both ‘T‘ and ‘S‘. This means ‘P‘ is a **confounder**. Ignoring a confounder can lead to incorrect conclusions about the relationship between ‘T‘ and ‘S‘. Our analysis must account for this.

2.2 (b) Compute Correlation between Treatment and Survival

The correlation coefficient measures the linear association between two variables. According to **Definition 2.13** from the script, it is a normalized version of covariance. A positive correlation would suggest that as treatment is applied, survival rates increase.

```
# Calculate the correlation between Treatment (T) and Survival (S)
corr_ts = df['T'].corr(df['S'])
print(f"Correlation between Treatment and Survival: {corr_ts:.4f}")
```

Result: The code output is `-0.3421`.

Conclusion: The correlation is negative. This suggests that receiving the treatment is associated with a *lower* chance of survival. Based on this metric alone, the treatment appears harmful.

2.3 (c) Compute Correlation between Precondition and Survival

We suspect that having a precondition might be bad for survival. Let's check the correlation.

```
# Calculate the correlation between Precondition (P) and Survival (S)
corr_ps = df['P'].corr(df['S'])
print(f"Correlation between Precondition and Survival: {corr_ps:.4f}")
```

Result: The code output is `-0.5898`.

Conclusion: The correlation is strongly negative. This confirms our suspicion that patients with a pre-existing condition have a significantly lower chance of survival, regardless of treatment. This is a key feature of a confounder: it has an independent effect on the outcome.

2.4 (d) Stratified Analysis

The negative correlation between T and S is alarming. However, because we identified P as a confounder, we know it might be distorting the picture. To remove the confounding effect of P, we analyze the relationship between T and S *separately for each group of P*. This is called **stratification** or conditioning.

```
# Group 1: Patients WITHOUT precondition (P=0)
df_p0 = df[df['P'] == 0]
survival_rate_p0 = df_p0.groupby('T')['S'].mean()
print("Survival rates for patients WITHOUT precondition (P=0):")
print(survival_rate_p0)
diff_p0 = survival_rate_p0[1] - survival_rate_p0[0]
print(f"Difference in survival (T=1 vs T=0) for P=0: {diff_p0*100:.2f}%\n")


# Group 2: Patients WITH precondition (P=1)
df_p1 = df[df['P'] == 1]
survival_rate_p1 = df_p1.groupby('T')['S'].mean()
print("Survival rates for patients WITH precondition (P=1):")
print(survival_rate_p1)
diff_p1 = survival_rate_p1[1] - survival_rate_p1[0]
print(f"Difference in survival (T=1 vs T=0) for P=1: {diff_p1*100:.2f}%")
```

Results:

- **Without Precondition (P=0):** The survival rate for treated patients is 95%, while for untreated it is 90%. The treatment increases the survival probability by 5 percentage points.
- **With Precondition (P=1):** The survival rate for treated patients is 20%, while for untreated it is 10%. The treatment increases the survival probability by 10 percentage points.

First Answer: In both subgroups (patients with and without the precondition), the treatment is beneficial. This directly contradicts the overall negative correlation we found in step (b). This is Simpson's Paradox in action.

2.5 (e) The Final Reply to the Doctors

The Question: "But overall, among all people, does the treatment help? Are we sure?"

Reply: Yes, we are confident the treatment is helpful. The initial finding that the treatment appeared harmful overall was a statistical illusion caused by confounding.

Here's the explanation:

1. Our data shows that patients with the precondition are much less likely to survive in general.

2. Doctors, making sound clinical judgments, were much more likely to administer the new treatment to the sicker patients (those with the precondition) in an attempt to save them.

3. This created a situation where the "treated" group was disproportionately composed of very sick patients, while the "untreated" group was mostly healthier patients.

4. When we compared the two groups overall, we were effectively comparing a sick group to a healthy group, making the treatment look bad.

5. By comparing treated vs. untreated patients *within the same health group* (i.e., stratifying by precondition), we make a fair, apples-to-apples comparison. This analysis clearly shows the treatment improves survival for both types of patients. The true causal effect of the treatment is positive.

Therefore, the stratified analysis in (d) provides the correct answer for judging the treatment's efficacy.

3 Part (ii): Complication as a Mediator

Now, the scenario changes. The doctors inform us that the third variable 'C' is not a precondition, but a **complication** that arises *during* treatment.

Why this changes everything: The causal structure is now different. A complication cannot cause the treatment, because it happens afterward. Instead, the treatment can cause a complication, which in turn affects survival. This makes 'C' a mediator [2].

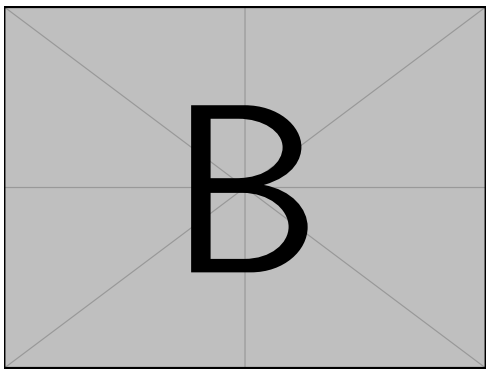

Figure 2: Causal graph with Complication (C) as a mediator.

Image snippet from the exercise statement.

The question "does the treatment help?" now asks for the **total effect** of the treatment. This effect is the sum of its direct impact on survival (path $T \to S$) and its indirect impact through the complications it might cause (path $T \to C \to S$).

Changes to the Analysis:

- **(a) Formalizing the Question:** The goal remains to find the total causal effect of T on S.
- **(b) Correlation (T, S):** This calculation is unchanged. However, its interpretation is now profoundly different. Because C is on the causal pathway from T to S, the overall correlation between T and S now correctly represents the **total effect** of the treatment, including any harm done by complications.
- **(c) Correlation (C, S):** This still measures the association between complications and survival.

- **(d) Stratified Analysis: This step is now incorrect.** If we stratify by the mediator 'C', we are artificially holding the complication status constant. This blocks the indirect causal path $T \to C \to S$. The analysis would then only measure the *direct* effect of T on S, which is not the total effect we are interested in. It would answer a different question, like "For patients who are going to develop complications anyway, does the treatment still offer a direct benefit?".

- **(e) The Final Reply:** The conclusion is now reversed. Since the goal is to evaluate the overall, real-world impact of administering the treatment, we must consider all its consequences, including the complications it causes. The correct measure is the overall comparison from step (b).

  New Reply: "Based on this new information, the treatment is harmful overall. While it might have some direct benefits, it appears to cause complications that are so severe that they outweigh any positive effects. The overall survival rate for patients receiving the treatment is lower than for those who do not. We should not use this treatment."

4 Summary and Takeaways

This exercise is a powerful illustration of a fundamental principle in data analysis and statistics:

1. **Causality over Correlation:** The question "does it work?" is causal. Answering it requires more than just computing correlations or conditional probabilities. It requires a **causal model** of the world—an assumption about what causes what.

2. **The Role of the Third Variable:** The same statistical data can lead to opposite conclusions. The correct analysis depends entirely on whether the third variable is a **confounder** (a common cause) or a **mediator** (an intermediate effect).

3. **Analysis Strategy:**

   - To estimate a causal effect, you must **adjust for confounders** (e.g., by stratification).
   - You must **not** adjust for mediators if you want the *total* causal effect.

A Further Explanations

Here are more detailed explanations of the key concepts used in this walkthrough.

A.1 [1] Confounding

A variable P is a **confounder** for the relationship between a treatment T and an outcome S if it is a common cause of both T and S.

$$T \leftarrow P \rightarrow S$$

In our example, a 'Precondition' is a confounder because:

1. It affects the treatment decision: Doctors are more likely to treat sicker patients ($P \rightarrow T$).
2. It affects the outcome: Sicker patients have a lower survival chance, regardless of treatment ($P \rightarrow S$).

If we just compare the groups $T = 1$ and $T = 0$, we are not making a fair comparison. The $T = 1$ group has more sick people ($P = 1$) than the $T = 0$ group. This difference in the makeup of the groups confounds our ability to see the true effect of T. The statistical remedy is to **condition** on the confounder, which means comparing $T = 1$ and $T = 0$ within subgroups where P is held constant.

A.2 [2] Mediation

A variable C is a **mediator** if it lies on the causal pathway between the treatment T and the outcome S.

$$T \rightarrow C \rightarrow S$$

In our example, a 'Complication' is a mediator because the treatment can cause the complication, and the complication then affects survival. The total effect of T on S is the sum of its direct effect (the arrow $T \rightarrow S$ that might also exist) and its indirect effect that is "mediated" by C.

If we condition on a mediator, we block the flow of causality along the indirect path. This is generally a mistake if we want to know the total effect of the treatment. The total effect is correctly estimated by the overall, un-stratified comparison between the $T = 1$ and $T = 0$ groups.

A.3 [3] Simpson's Paradox

Simpson's Paradox is a phenomenon in probability and statistics in which a trend appears in several different groups of data but disappears or reverses when these groups are combined. In our exercise, the treatment was beneficial in both the $P = 0$ and $P = 1$ groups, but appeared harmful when the groups were combined. This paradox occurs when a confounding variable is present and the groups are of unequal size. Resolving the paradox depends on identifying the causal structure and choosing the appropriate analysis (stratified or aggregated).

A.4 [4] Expectation of Bernoulli Variables

The exercise note correctly states that for a Bernoulli random variable $X \sim \text{Ber}(p)$, its expectation is $E[X] = p$. A Bernoulli variable takes value 1 (often called "success") with probability p and value 0 ("failure") with probability $1 - p$. From **Definition 2.1 (expectation)**, for a discrete variable, we have:

$$E[X] = \sum_{x \in \Omega} x \cdot p_X(x)$$

For a Bernoulli variable, the sample space is $\Omega = \{0, 1\}$, so:

$$E[X] = (0 \cdot P(X = 0)) + (1 \cdot P(X = 1)) = 0 \cdot (1 - p) + 1 \cdot p = p$$

This is why computing the mean of a binary (0/1) column in a dataset is equivalent to computing the empirical probability or frequency of the "1"s.