STA248 Tutorial 4 - Hypothesis Testing Through Simulation

DUE: April 12, 2021 @ 10 PM EDT

Submission Instructions

For this tutorial only, you may work with **one other classmate** but it is expected that by the end of the activity, each student will be able to complete the activity independently (i.e. learn the skills covered in this tutorial). If you are working with a partner, you will first create a group and submit one copy as a group on Crowdmark. **Note that once you form a group, it cannot be changed.** Read and follow the instructions below carefully.

- There have been some recent changes to how work is submitted on Crowdmark. Make sure to review the submission process HERE, including image size limitations and how to fix them.
- It is expected and required that all problems requiring the extensive use of R beyond simple calculations be completed in an R markdown file with outputs and text cleanly knit to pdf.
- Questions ((a), (b), . . . , etc.) should be clearly labeled in your text. Make use of headers to create eye-catching titles!
- Use `\newpage` to separate your parts onto different pages for easier uploading and reading.
- Your output should include: your code chunks, relevant output values, and written responses using properly displayed LaTeX notation. For problems that involve large simulations, please do not print out the full vector/data frame. Instead, use glimpse() or head() to display the first few rows for your own verification purposes and also for the teaching team to see your results.

Introduction

Hypothesis testing is a common framework to examine whether we have enough evidence to reject a previously held belief - commonly stated as the null hypothesis H_0. In medical sciences, a simple way to test whether a treatment is effective is to measure the difference seen in each patient (e.g. blood pressure) before and after receiving the treatment. Let us denote the true mean difference in blood pressure as μ_d. We can set up a hypothesis testing framework as the following:

$$H_0 : \mu_d = 0,\ H_1 : \mu_d \neq 0$$

The null hypothesis states that the true mean difference before and after taking the treatment is 0 - the treatment has absolutely no effect on average; while the alternative hypothesis states otherwise. Note that the alternative hypothesis H_1 covers positive and negative values for μ_d - the null hypothesis can be rejected regardless of whether the treatment *benefits* or *harms* the patient. This is called a **two-sided test** under the hypothesis testing framework.

One may ask why we even bother to reject H_0 if the treatment is harmful to the patients, and you are exactly right. In some cases, we are only interested in changes in one direction. If the treatment aims to reduce blood pressure for the patients who received it, we can perform a **one-sided test** instead:

$$H_0 : \mu_d = 0,\ H_1 : \mu_d < 0$$

The alternative hypothesis now only contains the range $\mu_d < 0$ when the treatment benefits the patients (reduced blood pressure, on average). Thus, we would never reject H_0 if the observed mean difference is above 0. Beware of the differences between one-sided / two-sided tests and the consequences whenever you encounter a hypothesis test.

Aim of the Tutorial

In this tutorial, you will guided through how to perform hypothesis testing using simulation. Often times, statisticians rely on asymptotic properties to perform hypothesis testing. However, those properties can be difficult to derive; require additional assumptions, or are inadequate for small sample sizes. Thus, hypothesis testing through simulation serves as a powerful alternative. We will focus on the simple case **when we know the distribution from which the data is drawn**.

Part 1 - Hypothesis Testing Through Simulation

A manufacturing company claims that the average lifetime of the machinery it produces lasts 3 years, however, some users are skeptical of the claim and believe that the true average lifetime of these machines is possibly lower. Your task now is to investigate the company's claim. Due to budgetary constraints, you only randomly sampled 30 machines produced by the company and observed that the average lifespan was 2.1 years. Assume that you know the lifespan of these machines follow an *exponential* distribution. How can you test the hypothesis through a simulation-based approach?

Recall that there are four steps for testing a hypothesis through simulation:

1. Simulating random samples of size n from the distribution under the assumption that the null hypothesis is true.
2. For each simulated dataset, compute your sample statistic and test statistic.
3. Repeat many times to simulate the distribution of the test statistic under H_0.
4. Estimate the p-value by finding the proportion of test statistics as extreme as the test statistic for your original data.

Assume that we know the the lifespan of these machines, L, follows an *exponential* distribution with parameter θ.

$$L \sim Exp(\theta)$$

Now, if what the company claims is true, θ would be 3 years since the mean of an exponential distribution is θ. Therefore, our null hypothesis would be $\theta = 3$ and the alternative hypothesis otherwise.

$$H_0 : \theta = 3,\ H_1 : \theta \neq 3$$

Our end goal will be to conduct a two-sided hypothesis test, but we'll start with a simpler case for now:

Recall that our observed average lifespan is 2.1 years. For now, suppose we suspect prior to data collection that the mean lifespan is less than three years, or $\theta < 3$. Let's begin by testing via simulation the following hypotheses:

$$H_0 : \theta = 3,\ H_1 : \theta < 3$$

Exercise (12 points)

We know we can recover probabilities via simulating enough observations from a distribution and looking at relative frequencies. Let's first determine which probabilities to find in this hypothesis test:

a) (2 points) What sample statistic would be used to estimate θ? Do we necessarily know the sampling distribution of the resulting estimator $\hat{\theta}$ for any sample size n?

b) (1 point) If this were a one-sided test with $H_1 : \theta < 3$, what event of your $\hat{\theta}$ would you use in (a) to compute the p-value?

c) (2 points) Briefly explain what it would mean to simulate data from the null hypothesis in this context. Following the steps above, detail how you could generate many $\hat{\theta}$ under H_0.

d) (3 points) Since the p-value represents how likely it is to observe random data as extreme as was observed when the null hypothesis is true, use your steps in (c) with $B = 100,000$ simulations to estimate the p-value. Report your p-value. *Note: Recall that the parameterization used by default in R is the rate, λ. Use the fact that $\lambda = \frac{1}{\theta}$ when simulating your data.*

e) (1 point) Returning to the original question, we should have conducted a two-sided test instead of a one-sided test. We can potentially find evidence to reject the null hypothesis if the true average lifespan is greater *or* lower than the company's claim. What event of your $\hat{\theta}$ would you use in a two-sided test to compute the p-value?

f) (3 points) Using the same simulations produced in (d), estimate the p-value for a two-sided test. *Hint: You should get a larger p-value and it is harder to reject H_0 with a two-sided test.*

   In general, is it easier or more difficult to reject the null hypothesis using a one-sided test? Briefly explain why.

Part 2 - Power Through Simulation

Now let's change our perspective and pretend you are part of the advisory board for the company. The company informs you that the *true average lifespan* is actually 2.7 years and they are worried that the investigators would reject H_0 and accuse the company of false advertising. Your task is to investigate **the probability that the investigators are able to correctly reject the null hypothesis**, also known as **power** in the hypothesis testing framework.

As you probably have noticed, you need to know the *true average lifespan* in order to calculate power - Rejecting the null hypothesis is much easier when the true average lifespan is 2 years compared to 2.9 years.

The first step is to determine the range of the observed average lifespan that the investigators in part 1 would reject the null hypothesis. Suppose that the company intends to reject the null hypothesis if the sample average lifespan is less than 2 years or longer than 4 years.

The question of power now becomes finding the probability that the investigators would observe an average lifespan in this range and reject the null hypothesis if **the true average lifespan is 2.7 years**.

Exercise (12 points)

Continuing with the two-sided test:

$$H_0 : \theta = 3; \; H_1 : \theta \neq 3$$

a) (1 point) Find the significance level of the test based on the company's rejection criteria.

b) (3 points) Estimate the power of this test using the rejection criteria and the definition of power to simulate data from the appropriate distribution. The estimated power is approximately 7.6%, indicating that it is relatively difficult to reject the company's claim that the average lifespan is 3 years using only 30 randomly sampled machines.

c) (4 points) Increasing the number of machines (samples) helps increase power to correctly reject H_0 while also reducing the chance of both type I and type II error. Using the rejection criteria by the company, define:

   - Type I error
   - Type II error

   In this context. Using your simulations in 2(b) and 1(e) to estimate the two error rates.

d) (4 points) Repeat your simulations using a sample size of $n = 100$ and find the estimated power and error rates. How has the sample size change affected these measures?

BONUS - (3 points): Use your simulation in 1(c) to estimate the rejection region for a 2% significance level in a **one-sided test**. Use it to estimate the power of your test at $n = 30$.

Submit separately from part 2 on Crowdmark. If successful, bonus points are applied to your tutorial 4 point total without exceeding 24 points.