Math 40 Probability and Statistical Inference
Winter 2021

Lecture 21 Hypothesis testing (7.1 and 7.2)

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

- We have a parametric model for a density of a random variable X

$$f(x;\theta)$$

- We have a data set, $\{X_i\}_i^n$ IID from $f(x;\theta)$.
- Statistical inference includes
    1. Point estimation (MM, MQ, MLE; chapter 6)
    2. Hypothesis test
    3. Confidence interval
- In chapter 7, we focus on 2 and 3.

- In hypothesis testing, we have a null hypothesis, H_0, and an alternative hypothesis, H_A.
- We use the data to decide whether we reject the null hypothesis or not.
- A value calculated from the data set for the test is called **test statistic** (often denoted as $T = T(\{X_i\})$.
- There is a **rejection region**, R, such that we reject the null hypothesis if the test statistic belongs to the rejection region.
- The rejection region changes based on how we design the test (reliability of the test).
- For the next few pages, we will practice on how to find the test statistic and the rejection region for several examples.

Example 1 A data set $\{X_i\}_i^n$ contains n house prices in Hanover. The NH house price follows a normal distribution with mean $300K$ and variance $(100K)^2$. It is also known that the variance of the Hanover house price is equal to the state variance. We want to test whether the mean house price of Hanover is different from the state mean price.

- Let θ be the mean of the Hanover house price.
- $H_0 : \theta = 300K$, $H_A : \theta \neq 300K$ (two-sided test)
- As we want to check the mean, the **test statistic**, T, is the sample mean

$$T = \overline{X}.$$

- As we are interested in whether the Hanover mean is different from the state mean, it is natural to reject the null hypothesis if the test statistic is far from the state mean.
- The **rejection region** has two parts,

$$R = (0, c_1) \cup (c_2, \infty)$$

  for critical values c_1 and c_2.
- That is, we reject the hypothesis if T is lower than c_1 (that is, the sample mean is too small compared to the state mean) or T is higher than c_2 (that is, the sample mean is too large compared to the state mean).

Example 2 (Cont'd) ,We have another hypothesis test; we want to know whether the Hanover house price is higher than the state mean.

- $H_0 : \theta = 300K$, $H_A : \theta > 300K$ (one-sided test).
- As in the previous case, the test statistics is the sample mean
$$T = \overline{X}.$$
- We are interested in whether the mean house price of Hanover is higher than the state mean.
- Thus, we reject the null hypothesis if the test statistic is higher than a critical value c. Thus, the rejection region is
$$R = (c, \infty).$$

Example 3 (example 7.2 of the textbook) A student answered correctly on 9 out of 15 yes/no questions. Test the hypothesis that the student picked answers at random by computing the maximum number of points under the assumption of random choice.

- We have a binomial random variable with $n = 15$ and an unknown p.
- $H_0 : p = 0.5$ and $H_A : p > 0.5$ (one-sided test)
- The test statistic is the total number of correct answers
$$T = \sum_{i}^{n} X_i.$$
- We reject the null hypothesis if the number of correct answers is larger than a critical value c. Thus, the rejection region is
$$R = [c, 15].$$

- An important question to ask is

How do we find the critical value(s)?

(so that we can find the rejection region explicitly and decide whether to reject the null hypothesis or not)

- Intuitively, the critical values are subjective; when there is a difference, what is a good size to tell the difference is significant?

- Types of errors

	Retain H_0	Reject H_0
H_0 is true		Type I error
H_A is true	Type II error	

- Typically, the probability of making the Type I erroris denoted by α while the probability of the type II error is denoted by β.

	Retain H_0	Reject H_0
H_0 is true	$1-\alpha$	α
H_A is true	β	$1-\beta$

- It is ideal to minimize the two errors simultaneously, but it is not possible.

- (The maximum possible value of) α is called **level of significance** of the test or **size**.

- $1-\beta$ is called the **power** of the test.

Now, we are going to find the rejection regions of the examples (that is, we are going to find the critical values). In the *decision of the rejection region*, we use *the significance level of the test*, α.

- Example 1

  1. The null hypothesis is $\theta = 300K$.
  2. If the null hypothesis is true, the distribution is $\mathcal{N}(300K, (100K)^2)$.
  3. Using the symmetry of the normal distribution, c_1 is the quantile with $p = \alpha/2$ ($c_1 = \texttt{qnorm}(\alpha/2, 300000, 100000)$
  4. Similarly, c_2 is the quantile with $p = 1 - \alpha/2$ ($c_2 = \texttt{qnorm}(1 - \alpha/2, 300000, 100000)$

- Example 2

  1. The null hypothesis is $\theta = 300K$.
  2. If the null hypothesis is true, the distribution is $\mathcal{N}(300K, (100K)^2)$.
  3. As it is a one-sided (to the right), c is the quantile with $p = 1 - \alpha$ ($c = \texttt{qnorm}(1 - \alpha, 300000, 100000)$

- Example 3

  1. The null hypothesis is $p = 0.5$.
  2. If the null hypothesis is true, the distribution is $Binomial(15, 0.5)$.

3. As it is a one-sided (to the right), c is the quantile with $p = 1 - \alpha$
($c = \texttt{qbinom}(1 - \alpha, 15, 0.5)$)

Q In Example 2, what will be the difference if the null hypothesis is $H_0 : \theta \leq 300K$ instead of $H_0 : \theta = 300K$?

In addition to the rejection region approach, we can use the p**-value**.

- The p**-value** is the probability of observing a test statistic at least as extreme as the current test statistic under the assumption that H_0 is true.

- That is, the p-value is the smallest significance level, α, at which we can reject H_0.

- Example 1: If the sample mean is 400K, the p-value is the probability

  $$Pr(X < 200K \text{ or } X > 400K), \quad X \sim \mathcal{N}(300K, (100K)^2)$$

  which is 0.3173.

- Example 2: If the sample mean is 400K, the p-value is the probability

  $$Pr(X > 400K), \quad X \sim \mathcal{N}(300K, (100K)^2)$$

  which is 0.1587.

- Example 3: The test statistic is 9. The p-value is the probability

  $$Pr(X > 9), \quad X \sim Binomial(15, 0.5)$$

  which is 0.1509.

If the p-value is less than α, the test is called **statistically significant**.