29 2026-04-17 | Week 13 | Lecture 29

29.0.1 What is a hypothesis test?

Definition 130 (Hypothesis Testing)**.** Given a sample from $X_1, \ldots, X_n$ from a distribution depending on parameter θ. We are interested in comparing two competingk hypotheses: a null hypothesis H_0 and an alternative hypothesis H_a. A ***hypothesis test*** is a procedure or decision rule utilizing a statistic T which specifies for which values of T the hypothesis H_0 is rejected, and for which values H_0 is accepted. The statistic T is called the ***test statistic***.

Associated with a hypothesis test is a ***rejection region*** (RR), the set which specifies the values of the test statistic for which the null hypothesis to be rejected in favor of the alternative hypothesis. That is, if $T \in \text{RR}$, then we reject H_0. If $T \notin \text{RR}$, we accept H_0.

Example 131. Suppose you have a coin but you don't know if it is a fair coin or not. You think maybe the coin is biased in favor of flipping heads. As an experiment, you flip the coin 10 times and you get 8 heads. Is this convincing evidence that the coin is biased?

In this case, the data is an IID sample

$$X_1, \ldots, X_{10}$$

where

$$X_i = \begin{cases} 1 & : \text{ with probability } \theta \\ 0 & : \text{ with probability } 1-\theta \end{cases}$$

The null hypothesis is

$$H_0 : \theta = \frac{1}{2}$$

The alternative hypothesis is that heads is more likely than tails:

$$H_1 : \theta > \frac{1}{2}$$

The p-value is

$$\begin{aligned} p &= \mathbb{P}\left[\text{at least 8 heads when } \theta = \frac{1}{2}\right] \\ &= \binom{10}{8}\theta^8(1-\theta)^2 + \binom{10}{9}\theta^9(1-\theta)^1 + \binom{10}{10}\theta^{10}(1-\theta)^0 \text{ with } \theta = \frac{1}{2} \\ &= \binom{10}{8}\left(\frac{1}{2}\right)^{10} + \binom{10}{9}\left(\frac{1}{2}\right)^{10} + \binom{10}{10}\left(\frac{1}{2}\right)^{10} \\ &= \frac{7}{128} \\ &\approx 0.0546. \end{aligned}$$

In words, if the coin were fair, we would expect it to flip 8 or more heads (out of ten) about 5.46% of the time.

In the previous example, the test statistic was $T =$ (number of heads). A reasonable rejection region might be $[T \geq 8]$ or $[T \geq 9]$.

End of Example 131. □

Example 132. Now, suppose that we flip the coin 100 times and we get 80 of them to be heads. Is this convincing evidence that the coin isn't fair?

To answer this question, we'll compute the p-value. Let $X_1, \ldots, X_{100}$ be our IID random sample, where

$$X_i = \begin{cases} 1 & : \text{ if } i\text{th coin is heads} \\ 0 & : \text{ otherwise} \end{cases}$$

Then the number of heads is the random variable

$$S_n = X_1 + \ldots + X_n.$$

Observe that $\mu = \mathbb{E}[X_i] = \frac{1}{2}$ and Assuming the coin is a fair coin, we have

$$\mu = \mathbb{E}[X_i] = \frac{1}{2}$$

and

$$\sigma^2 = \mathbb{E}\left[X_i^2\right] - \frac{1}{4} = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

so that

$$\sigma = \frac{1}{2}.$$

By the central limit theorem, we know that

$$\frac{S_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}} Z,$$

where Z is a standard normal random variable. For this specific problem, we have:

$$\frac{S_{100}}{100} \approx \frac{1}{2} + \frac{1}{20} Z.$$

We can now compute the p-value:

$$\begin{aligned}
\text{p-value} &= \mathbb{P}[\text{at least 80 heads out of 100 tosses} \mid \text{coin is fair}] \\
&= \mathbb{P}[S_{100} \geq 80 \mid \text{coin is fair}] \\
&= \mathbb{P}\left[\frac{S_{100}}{100} \geq .8 \mid \text{coin is fair}\right] \\
&\approx \mathbb{P}\left[\frac{1}{2} + \frac{Z}{20} \geq .8\right] && \text{Key step. Follows by CLT.} \\
&= \mathbb{P}\left[\frac{Z}{20} \geq .3\right] \\
&= \mathbb{P}[Z \geq 6] \\
&\approx .000000001
\end{aligned}$$

In other words, if the coin were a fair coin, observing heads at least 80 out of 100 times would be extraordinarily unlikely. Therefore we reject H_0.

End of Example 132. □

29.0.2 Type I and Type II errors

There are two types of decision errors in the hypothesis testing framework.

Definition 133 (Decision errors)**.** A ***type I error*** is made if H_0 is rejected when H_0 is true. The probability of a type I error is denoted by α:

$$\alpha = \Pr[\text{Type I error}] = \Pr[\text{rejecting } H_0 \text{ when } H_0 \text{ is true}]$$

The value of α is called the ***significance level*** (or just ***level***) of the test.

Similarly, a ***type II error*** is made if H_0 is accepted when H_a is true. We denote the probability of this happening by β:

$$\beta = \Pr[\text{Type II error}] = \Pr[\text{accepting } H_0 \text{ when } H_a \text{ is true}]$$

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

These types of errors, and their probabilities, can be summarized in the following table:

	Reject H_0	Accept H_0
H_0 true	Type I error level $= \alpha$	Correctly accept H_0 $1 - \alpha$
H_a true	Correctly reject H_0 power $= 1 - \beta$	Type II error β