8.4.2 General Setting and Definitions

Example 8.22 provided a basic introduction to hypothesis testing. Here, we would like to provide a general setting for problems of hypothesis testing and formally define the terminology that is used in hypothesis testing. Although there are several new phrases such as null hypothesis, type I error, significance level, etc., there are not many new concepts or tools here. Thus, after going through a few examples, the concepts should become clear.

Suppose that θ is an unknown parameter. A hypothesis is a statement such as $\theta = 1$, $\theta > 1.3$, $\theta \neq 0.5$, etc. In hypothesis testing problems, we need to decide between two contradictory hypotheses. More precisely, let S be the set of possible values for θ. Suppose that we can partition S into two disjoint sets S_0 and S_1. Let H_0 be the hypothesis that $\theta \in S_0$, and let H_1 be the hypothesis that $\theta \in S_1$.

H_0 (the **null** hypothesis): $\theta \in S_0$.

H_1 (the **alternative** hypothesis): $\theta \in S_1$.

In Example 8.22, $S = [0, 1]$, $S_0 = \{\frac{1}{2}\}$, and $S_1 = [0, 1] - \{\frac{1}{2}\}$. Here, H_0 is an example of a **simple** hypothesis because S_0 contains only one value of θ. On the other hand, H_1 is an example of **composite** hypothesis since S_1 contains more than one element. It is often the case that the null hypothesis is chosen to be a simple hypothesis.
Often, to decide between H_0 and H_1, we look at a function of the observed data. For instance, in Example 8.22, we looked at the random variable Y, defined as

$$Y = \frac{X - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}},$$

where X was the total number of heads. Here, X is a function of the observed data (sequence of heads and tails), and thus Y is a function of the observed data. We call Y a *statistic*.

Definition 8.3. Let X_1, X_2, $\cdots$, X_n be a random sample of interest. A **statistic** is a real-valued function of the data. For example, the sample mean, defined as

$$W(X_1, X_2, \cdots, X_n) = \frac{X_1 + X_2 + \ldots + X_n}{n},$$

is a statistic. A **test statistic** is a statistic based on which we build our test.

To decide whether to choose H_0 or H_1, we choose a test statistic, $W = W(X_1, X_2, \cdots, X_n)$. Now, assuming H_0, we can define the set $A \subset \mathbb{R}$ as the set of possible values of W for which we would accept H_0. The set A is called the **acceptance region**, while the set $R = \mathbb{R} - A$ is said to be the **rejection region**. In Example 8.22, the acceptance region was found to be the set $A = [-1.96, 1.96]$, and the set $R = (-\infty, -1.96) \cup (1.96, \infty)$ was the rejection region.

There are two possible errors that we can make. We define **type I error** as the event that we reject H_0 when H_0 is true. Note that the probability of type I error in general depends on the real value of θ. More specifically,

$$\begin{aligned} P(\text{type I error} \mid \theta) &= P(\text{Reject } H_0 \mid \theta) \\ &= P(W \in R \mid \theta), \quad \text{for } \theta \in S_0. \end{aligned}$$

If the probability of type I error satisfies

$$P(\text{type I error}) \leq \alpha, \quad \text{for all } \theta \in S_0,$$

then we say the test has **significance level** α or simply the test is a **level** α test. Note that it is often the case that the null hypothesis is a simple hypothesis, so S_0 has only one element (as in Example 8.22). The second possible error that we can make is to accept H_0 when H_0 is false. This is called the **type II error**. Since the alternative hypothesis, H_1, is usually a composite hypothesis (so it includes more than one value of θ), the probability of type II error is usually a function of θ. The probability of type II error is usually shown by β:

$$\beta(\theta) = P(\text{Accept } H_0 \mid \theta), \quad \text{for } \theta \in S_1.$$

We now go through an example to practice the above concepts.

Example 8.23
Consider a radar system that uses radio waves to detect aircraft. The system receives a signal and, based on the received signal, it needs to decide whether an aircraft is present or not. Let X be the received signal. Suppose that we know

$$X = W, \qquad \text{if no aircraft is present.}$$

$$X = 1 + W, \qquad \text{if an aircraft is present.}$$

where $W \sim N(0, \sigma^2 = \frac{1}{9})$. Thus, we can write $X = \theta + W$, where $\theta = 0$ if there is no aircraft, and $\theta = 1$ if there is an aircraft. Suppose that we define H_0 and H_1 as follows:

H_0 (null hypothesis): No aircraft is present.

H_1 (alternative hypothesis): An aircraft is present.

a. Write the null hypothesis, H_0, and the alternative hypothesis, H_1, in terms of possible values of θ.
b. Design a level 0.05 test ($\alpha = 0.05$) to decide between H_0 and H_1.
c. Find the probability of type II error, β, for the above test. Note that this is the probability of missing a present aircraft.
d. If we observe $X = 0.6$, is there enough evidence to reject H_0 at significance level $\alpha = 0.01$?
e. If we would like the probability of missing a present aircraft to be less than 5%, what is the smallest significance level that we can achieve?

Solution

a. The null hypothesis corresponds to $\theta = 0$ and the alternative hypothesis corresponds to $\theta = 1$. Thus, we can write

   H_0 (null hypothesis): No aircraft is present: $\theta = 0$.

   H_1 (alternative hypothesis): An aircraft is present: $\theta = 1$.

   Note that here both hypotheses are simple.

b. To decide between H_0 and H_1, we look at the observed data. Here, the situation is relatively simple. The observed data is just the random variable X. Under H_0, $X \sim N(0, \frac{1}{9})$, and under H_1, $X \sim N(1, \frac{1}{9})$. Thus, we can suggest the following test: We choose a threshold c. If the observed value of X is less than c, we choose H_0 (i.e., $\theta = EX = 0$). If the observed value of X is larger than c, we choose H_1 (i.e., $\theta = EX = 1$). To choose c, we use the required α:

$$\begin{aligned} P(\text{type I error}) &= P(\text{Reject } H_0 \mid H_0) \\ &= P(X > c \mid H_0) \\ &= P(W > c) \\ &= 1 - \Phi(3c) \quad \left(\text{since assuming } H_0, X \sim N\left(0, \frac{1}{9}\right)\right). \end{aligned}$$

Letting $P(\text{type I error}) = \alpha$, we obtain

$$c = \frac{1}{3}\Phi^{-1}(1 - \alpha).$$

Letting $\alpha = 0.05$, we obtain

$$c = \frac{1}{3}\Phi^{-1}(0.95) = 0.548$$

c. Note that, here, the alternative hypothesis is a simple hypothesis. That is, it includes only one value of θ (i.e., $\theta = 1$). Thus, we can write

$$\begin{aligned} \beta = P(\text{type II error}) &= P(\text{accept } H_0 \mid H_1) \\ &= P(X < c \mid H_1) \\ &= P(1 + W < c) \\ &= P(W < c - 1) \\ &= \Phi(3(c - 1)). \end{aligned}$$

Since $c = 0.548$, we obtain $\beta = 0.088$.

d. In part (b), we obtained

$$c = \frac{1}{3}\Phi^{-1}(1 - \alpha).$$

For $\alpha = 0.01$, we have $c = \frac{1}{3}\Phi^{-1}(0.99) = 0.775$ which is larger than 0.6. Thus, we cannot reject H_0 at significance level $\alpha = 0.01$.

e. In part (c), we obtained

$$\beta = \Phi(3(c - 1)).$$

To have $\beta = 0.05$, we obtain

$$\begin{aligned} c &= 1 + \frac{1}{3}\Phi^{-1}(\beta) \\ &= 1 + \frac{1}{3}\Phi^{-1}(0.05) \\ &= 0.452 \end{aligned}$$

Thus, we need to have $c \leq 0.452$ to obtain $\beta \leq 0.05$. Therefore,

$$\begin{aligned} P(\text{type I error}) &= 1 - \Phi(3c) \\ &= 1 - \Phi(3 \times 0.452) \\ &= 0.0875, \end{aligned}$$

which means that the smallest significance level that we can achieve is $\alpha = 0.0875$.

Trade-off Between α and β: Since α and β indicate error probabilities, we would ideally like both of them to be small. However, there is in fact a trade-off between α and β. That is, if we want to decrease the probability of type I error (α), then the probability of type II error (β) increases, and vise versa. To see this, we can look at our analysis in Example 8.23. In that example, we found

$$\begin{aligned} \alpha &= 1 - \Phi(3c), \\ \beta &= \Phi(3(c-1)). \end{aligned}$$

Note that $\Phi(x)$ is an increasing function. If we make c larger, α becomes smaller, and β becomes larger. On the other hand, if we make c smaller, α becomes larger, and β becomes smaller. Figure 8.10 shows type I and type II error probabilities for Example 8.23.

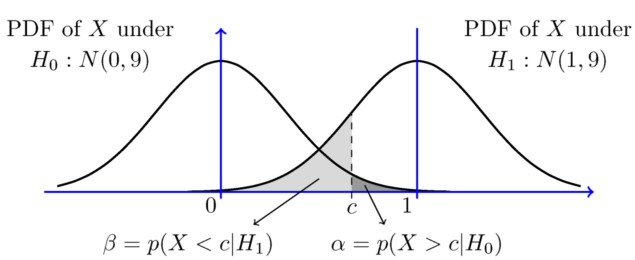

Figure 8.10 - Type I and type II errors in Example 8.23.