

## 6 Type II Errors, Power of a Test and the Neyman-Pearson Lemma

We are returning now to hypothesis testing. Recall that for a target parameter  $\theta$ , we are testing

$$\begin{aligned} H_0 : \theta &= \theta_0, \text{ versus one of} \\ H_1 : \begin{cases} \theta < \theta_0 \\ \theta > \theta_0 \\ \theta \neq \theta_0, \end{cases} \end{aligned} \quad (6.1)$$

The “goodness” of a test is measured by the two probabilities of risk

$$\begin{aligned} \alpha &= P(\text{type I error}) = P(\text{reject } H_0 \mid H_0) \\ \beta &= P(\text{type II error}) = P(\text{not reject } H_0 \mid H_1). \end{aligned}$$

The smaller both of them are, the more reliable the test is. For some problems, a type I error is more dangerous, while for others, a significant type II error is unacceptable. In general,  $\alpha$  is preset, at most 0.05 and the test is designed so that  $\beta$  is also small enough to be acceptable.

### 6.1 Type II Errors and Power of a Test

So far, type II errors were not discussed. That is because the computation of  $\beta$  can be more difficult. The condition that  $H_1$  is true *does not* specify an actual value for the unknown parameter and thus, does not identify a distribution, for which the probability can be computed. The simple condition that a parameter  $\theta$  is less than, greater than or not equal to a value is not enough to help us compute the probability. However, if the alternate  $H_1$  is also a *simple* hypothesis

$$H_1 : \theta = \theta_1,$$

then  $\beta$  can be computed. Thus,  $\beta$ , unlike  $\alpha$ , depends on the value specified in the alternative hypothesis,

$$\beta = \beta(\theta_1).$$

**Example 6.1.** Let us consider again the problem in Example 4.2. in Lecture 11 (or Example 4.4 in Lecture 10): The number of monthly sales at a firm is known to have a mean of 20 and a standard deviation of 4 and all salary, tax and bonus figures are based on these values. However, in times of

economical recession, a sales manager fears that his employees do not average 20 sales per month, but less, which could seriously hurt the company. For a number of 36 randomly selected salespeople, it was found that in one month they averaged 19 sales. At the 5% significance level, does the data confirm or contradict the manager's suspicion?

Now let us find  $\beta$  for the test

$$H_0 : \mu = \mu_0 = 20$$

$$H_1 : \mu = \mu_1 = 18 < 20,$$

i.e. find  $\beta(\mu_1)$ .

**Solution.** We tested a left-tailed alternative for the mean

$$H_0 : \mu = 20$$

$$H_1 : \mu < 20.$$

The population standard deviation was given,  $\sigma = 4$  and for a sample of size  $n = 36$ , the sample mean was  $\bar{X} = 19$ . For the test statistic

$$TS = Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in N(0, 1),$$

the observed value was

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{19 - 20}{\frac{4}{6}} = -1.5.$$

At the significance level  $\alpha = 0.05$ , we have determined the rejection region

$$\begin{aligned} RR &= \left\{ Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z_{0.05} \right\} = \left\{ \frac{\bar{X} - 20}{\frac{4}{6}} \leq -1.645 \right\} \\ &= \left\{ \bar{X} \leq -1.645 \cdot \frac{4}{6} + 20 \right\} = \{ \bar{X} \leq 18.9 \}. \end{aligned}$$

Then, in a similar fashion we compute

$$\beta(\mu_1) = P(\text{not reject } H_0 \mid H_1) = P(\bar{X} > 18.9 \mid \mu = \mu_1).$$

If the true value of  $\mu$  is  $\mu_1$ , then the statistic

$$Z_1 = \frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - 18}{\frac{4}{6}}$$

has a Standard Normal  $N(0, 1)$  distribution. Hence

$$\begin{aligned} \beta(\mu_1) &= P(\bar{X} > 18.9 \mid \mu = \mu_1) \\ &= P\left(\frac{\bar{X} - 18}{\frac{4}{6}} > \frac{18.9 - 18}{\frac{4}{6}} \mid \mu = 18\right) \\ &= P(Z_1 > 1.35 \mid Z_1 \in N(0, 1)) \\ &= 1 - P(Z_1 \leq 1.35 \mid Z_1 \in N(0, 1)) \\ &= 1 - \text{normcdf}(1.35) = 0.0885. \end{aligned}$$

■

**Remark 6.2.** Let us take a closer look at the computation of  $\alpha$  and  $\beta$  in the previous example. We used the fact that the variable

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

has a  $N(0, 1)$  distribution. So, when the true value of  $\mu$  is  $\mu_0 = 20$ , then

$$Z_0 = Z(\mu = \mu_0) \in N(0, 1)$$

and when the value is  $\mu_1 = 18$ , then

$$Z_1 = Z(\mu = \mu_1) \in N(0, 1).$$

However, in the end, we expressed the error probabilities  $\alpha$  and  $\beta$ , by looking at the distribution of  $\bar{X}$  *by itself*, not its reduced version. In other words, we used the fact that, when the true value of  $\mu$  is  $\mu_0 = 20$ , then

$$\bar{X} \in N(\mu_0, \sigma/\sqrt{n}) \text{ and } \alpha = P(\bar{X} \leq 18.9),$$

while when the true value is  $\mu_1 = 18$ , then

$$\bar{X} \in N(\mu_1, \sigma/\sqrt{n}) \text{ and } \beta = P(\bar{X} > 18.9).$$

This can be seen graphically in Figure 1.

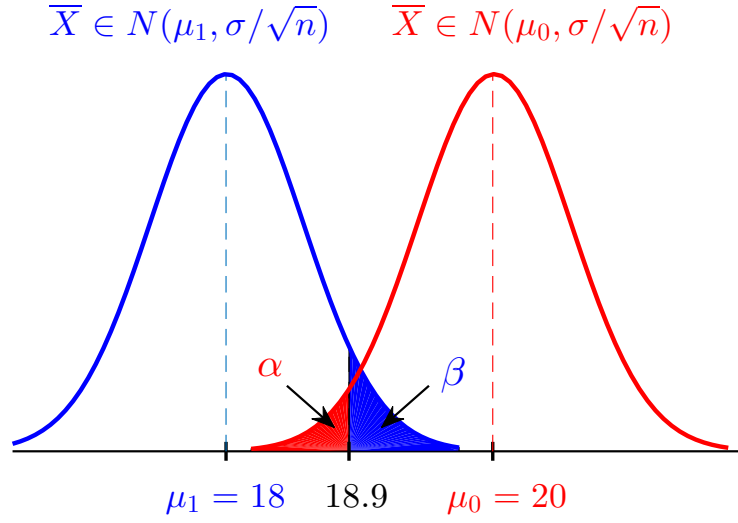


Fig. 1: Type I and type II errors

In order to have a better control over  $\beta$ , we introduce the following notion.

**Definition 6.3.** The **power of a test** on a parameter  $\theta$ , unknown, is the probability of rejecting the null hypothesis

$$\pi(\theta^*) = P(\text{reject } H_0 \mid \theta = \theta^*) = P(TS \in RR \mid \theta = \theta^*), \quad (6.2)$$

when the true value of the parameter is  $\theta = \theta^*$ .

Notice that the power of a test is, usually, a function of the parameter  $\theta$ , because the alternative hypothesis includes a set of parameter values.

Indeed, if the null hypothesis is true, i.e.  $\theta = \theta_0$ , then

$$\pi(\theta_0) = P(TS \in RR \mid \theta = \theta_0) = P(\text{reject } H_0 \mid H_0) = \alpha. \quad (6.3)$$

For any *other* true value (in the alternative hypothesis  $H_1$ )  $\theta = \theta_1 \neq \theta_0$ ,

$$\begin{aligned} \pi(\theta_1) &= P(\text{reject } H_0 \mid \theta = \theta_1) = P(\text{reject } H_0 \mid H_1) \\ &= 1 - P(\text{not reject } H_0 \mid H_1) = 1 - \beta(\theta_1). \end{aligned} \quad (6.4)$$

So, basically, the power of a test is the probability of rejecting a *false* null hypothesis. The larger the

power is, the smaller  $\beta$  is, which is what we want in a test. Then we can state a hypothesis testing problem the following way:

For a parametric test where both hypotheses are simple

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &= \theta_1, \end{aligned}$$

we preset  $\alpha = \pi(\theta_0)$  and we determine a rejection region  $RR$  for which

$$\pi(\theta_1) = 1 - \beta(\theta_1)$$

is *the largest possible*. Such a test is called a **most powerful test**.

## 6.2 The Neyman-Pearson Lemma (NPL)

Most powerful tests cannot always be found. The following result gives a procedure for finding a most powerful test, when both hypotheses tested are simple.

**Lemma 6.4** (Neyman-Pearson (NPL)). *Let  $X$  be a characteristic with pdf  $f(x; \theta)$ , with  $\theta \in A \subset \mathbb{R}$ , unknown. Suppose we test on  $\theta$  the simple hypotheses*

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &= \theta_1, \end{aligned}$$

*based on a random sample  $X_1, \dots, X_n$ . Let  $L(\theta) = L(X_1, \dots, X_n; \theta)$  denote the likelihood function of this sample. Then for a fixed  $\alpha \in (0, 1)$ , a most powerful test is the test with rejection region given by*

$$RR = \left\{ \frac{L(\theta_1)}{L(\theta_0)} \geq k_\alpha \right\}, \quad (6.5)$$

*where the constant  $k_\alpha > 0$  depends only on  $\alpha$  and the sample variables.*

**Example 6.5.** Suppose  $X_1$  represents a single observation from a probability density given by

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1}, & \text{if } x \in (0, 1) \\ 0, & \text{otherwise.} \end{cases}$$

Find the NPL most powerful test that at the 5% significance level tests

$$\begin{aligned}H_0 : \theta &= 1 \quad (= \theta_0) \\H_1 : \theta &= 30 \quad (= \theta_1).\end{aligned}$$

Also, find  $\beta$  for that test.

**Solution.** Since our sample has size 1, we have

$$\frac{L(\theta_1)}{L(\theta_0)} = \frac{f(X_1; \theta_1)}{f(X_1; \theta_0)} = \frac{30X_1^{29}}{1} = 30X_1^{29}.$$

So the rejection region given by the NPL is

$$RR = \{30X_1^{29} \geq k_\alpha\} = \{X_1 \geq K_\alpha\},$$

where  $K_\alpha = \left(\frac{1}{30}k_\alpha\right)^{1/29}$ .

We find the value of  $K_\alpha$  from

$$\begin{aligned}\alpha &= P(X_1 \in RR \mid H_0) = P(X_1 \geq K_\alpha \mid \theta = 1) \\&= \int_{K_\alpha}^1 dx = 1 - K_\alpha,\end{aligned}$$

i.e.  $K_\alpha = 1 - \alpha = 0.95$ .

So, of all tests for testing  $H_0$  versus  $H_1$ , based on a sample of size 1, the observation  $X_1$ , at the significance level  $\alpha = 0.05$ , the most powerful test has rejection region

$$RR = \{X_1 \geq 0.95\}.$$

For this test,

$$\begin{aligned}\beta(\theta_1) &= P(X_1 < K_\alpha \mid \theta = 30) = \int_0^{K_\alpha} 30x^{29} dx \\&= x^{30} \Big|_0^{K_\alpha} = (K_\alpha)^{30} = (1 - \alpha)^{30} = 0.166\end{aligned}$$

and the power is

$$\pi(\theta_1) = 1 - \beta(\theta_1) = 0.834.$$

Note that the error probability  $\beta$  that we obtained is *unacceptably large*, but considering that the estimation was based on a sample of size *one*, we cannot expect too much accuracy. ■

**Remark 6.6.** Notice that the rejection region and, hence, the most powerful test we found in Example 6.5, depend on the value stated in  $H_1$ . For a different value of  $\theta_1$ , we would have found a *different* rejection region. That is usually the case. However, sometimes, a test obtained with the NPL actually maximizes the power for *every* value in  $H_1$ , i.e. even if  $H_1$  is not a simple hypothesis. Such a test is called a **uniformly most powerful test**.

**Example 6.7.** Let  $X_1, \dots, X_n$  be a random sample drawn from a Normal  $N(\mu, \sigma)$  distribution, with  $\mu \in \mathbb{R}$  unknown and  $\sigma > 0$  known. At the significance level  $\alpha \in (0, 1)$ , find the most powerful right-tailed test for testing

$$\begin{aligned} H_0 : \quad & \mu = \mu_0 \\ H_1 : \quad & \mu > \mu_0. \end{aligned}$$

**Solution.** First we use the NPL to find a most powerful test for a *simple* alternative, i.e.

$$\begin{aligned} H_0 : \quad & \mu = \mu_0 \\ H_1 : \quad & \mu = \mu_1 > \mu_0. \end{aligned}$$

We have the Normal pdf

$$f(x; \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}.$$

The likelihood function is

$$\begin{aligned} L(\mu) &= \prod_{i=1}^n f(X_i; \mu) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right). \end{aligned}$$

Then, by the NPL, we find

$$\frac{L(\mu_1)}{L(\mu_0)} = \exp\left(\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (X_i - \mu_0)^2 - \sum_{i=1}^n (X_i - \mu_1)^2 \right]\right) \geq k_\alpha,$$

or, taking the logarithm  $\ln$  (which is an increasing function) on both sides,

$$\begin{aligned} \frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (X_i - \mu_0)^2 - \sum_{i=1}^n (X_i - \mu_1)^2 \right] &\geq \ln k_\alpha, \\ \sum_{i=1}^n X_i^2 - 2\mu_0 \sum_{i=1}^n X_i + n\mu_0^2 - \left( \sum_{i=1}^n X_i^2 - 2\mu_1 \sum_{i=1}^n X_i + n\mu_1^2 \right) &\geq 2\sigma^2 \ln k_\alpha. \end{aligned}$$

After cancellations and using  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , we have

$$2n\bar{X}(\mu_1 - \mu_0) \geq 2\sigma^2 \ln k_\alpha + n(\mu_1^2 - \mu_0^2).$$

Since  $\mu_1 > \mu_0$ , we get

$$\bar{X} \geq \frac{\sigma^2 \ln k_\alpha}{n(\mu_1 - \mu_0)} + \frac{\mu_1 + \mu_0}{2} = K_\alpha.$$

Then we use the test statistic  $TS = \bar{X}$ , for which we found the rejection region

$$RR = \{\bar{X} \geq K_\alpha\}.$$

But

$$\begin{aligned} \alpha &= P(\bar{X} \geq K_\alpha \mid \mu = \mu_0) \\ &= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{K_\alpha - \mu_0}{\sigma/\sqrt{n}} \mid \mu = \mu_0\right) \\ &= P\left(Z_0 \geq \frac{K_\alpha - \mu_0}{\sigma/\sqrt{n}} \mid Z_0 \in N(0, 1)\right) \\ &= P(Z_0 \geq z_{1-\alpha}), \end{aligned}$$

since  $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \in N(0, 1)$ . Then we must have

$$\frac{K_\alpha - \mu_0}{\sigma/\sqrt{n}} = z_{1-\alpha}, \quad K_\alpha = \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}},$$



so  $K_\alpha$  is *independent* of  $\mu_1$ . Then the test with  $RR = \{\bar{X} \geq K_\alpha\}$  is a *uniformly* most powerful test for testing

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0,$$

at the significance level  $\alpha$ .

■

**Remark 6.8.** In a similar manner, we can find a uniformly most powerful test for the left-tailed case

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0.$$