6. Neyman-Pearson Lemma

Definition 10. A hypothesis is a statement about a population parameter.

The definition of a hypothesis is quite general, but the important point is that a hypothesis makes a statement about the population. The goal of a hypothesis test is to decide (based on a sample from the population) which of two complementary hypotheses is true.

Definition 11. The two complementary hypotheses in a hypothesis testing problem are called the *null hypothesis* and the *alternative hypothesis*. They are denoted by H_0 and H_1, respectively.

If θ denotes a population parameter, the general format of the null and alternative hypotheses is as follows:

$$H_0 : \theta = \theta_0 \text{ and } H_1 : \theta = \theta_1.$$

In a hypothesis testing problem, after observing the sample, the experimenter must decide either to accept H_0 as true, or to reject H_0 as false (and decide H_1 is true).

Definition 12. A hypothesis testing procedure (or, hypothesis test) is a rule that specifies:
i. For which sample values the decision is made to accept H_0 as true,
ii. For which sample values H_0 is rejected (and H_1 is accepted as true).
The subset of the sample space for which H_0 will be rejected is called the *rejection region* (or, critical region). The complement of the rejection region is called the *acceptance region*.

Recall that in deciding to accept (or, reject) the null hypothesis H_0, an experimenter might be making a mistake. Usually, hypothesis tests are evaluated and compared through their probabilities of making mistakes. In this module, we will discuss how these error probabilities can be controlled. In some cases, it can even be determined which tests have the smallest possible error probabilities.

A hypothesis test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ might make one of two types of errors. These two types of errors traditionally have been given the non-mnemonic names, Type-I error and Type-II error. If $\theta = \theta_0$ but the hypothesis test incorrectly decides to reject H_0, then the test has made a Type-I error. If, on the other hand $\theta = \theta_1$, but the test decides to accept H_0, a Type-II error has been made. These two different situations are depicted in the table below:

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct decision	Type-I Error
	H_1	Type-II Error	Correct decision

Suppose R denotes the rejection region for a test. Then, for $\theta = \theta_0$, the test will make a mistake if $X \in R$, so the probability of a Type-I error is $\mathbf{P}_\theta(X \in R)$. For $\theta = \theta_1$, the probability of a Type-II error is $\mathbf{P}_\theta (X \in R^c)$. This switching from R to R^c is a bit confusing, but if we realize that $\mathbf{P}_\theta (X \in R^c) = 1 - \mathbf{P}_\theta(X \in R)$, then $\mathbf{P}_\theta(X \in R)$ (as a function of θ) contains all the information about the test with rejection region R. We have

$$\mathbf{P}_\theta(X \in R) = \begin{cases} \text{probability of a Type-I error} & \text{if } \theta = \theta_0, \\ \text{one minus the probability of a Type-II error} & \text{if } \theta = \theta_1. \end{cases}$$

The power function of a hypothesis test with rejection region R is defined by $\beta(\theta) = \mathbf{P}_\theta(X \in R)$, which is a function of θ.

Let C be the collection of all level α $(0 < \alpha < 1)$ tests. The test described in the definition is then called a *uniformly most powerful* (UMP) level α test. The following famous theorem clearly describes which tests are UMP level α tests in the situation, where the null and alternative hypotheses both consist of only one probability distribution for the sample (i.e., when both H_0 and H_1 are simple hypotheses).

Theorem 13. *(Neyman-Pearson Lemma) Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf (or, pmf) corresponding to θ_i is f_{θ_i} for $i = 0, 1$, using a test with rejection region R that satisfies*

i. $x \in R$ if $f_{\theta_1}(x) > k f_{\theta_0}(x)$ and $x \in R^c$ if $f_{\theta_1}(x) < k f_{\theta_0}(x)$ for some $k \geq 0$, and
ii. $\alpha = P_{\theta_0}(X \in R)$.

Then, any test that satisfies these two conditions is a UMP level α test.

Let us now look into two examples.

Example 14. (UMP Binomial test) Let $X \sim \text{Bin}(2, \theta)$. We want to test

$$H_0 : \theta = \frac{1}{2} \text{ versus } H_1 : \theta = \frac{3}{4}.$$

Calculating the ratios of the pmfs gives:

$$\frac{f\left(0 \mid \theta = \frac{3}{4}\right)}{f\left(0 \mid \theta = \frac{1}{2}\right)} = \frac{1}{4}, \quad \frac{f\left(1 \mid \theta = \frac{3}{4}\right)}{f\left(1 \mid \theta = \frac{1}{2}\right)} = \frac{3}{4} \quad \text{and} \quad \frac{f\left(2 \mid \theta = \frac{3}{4}\right)}{f\left(2 \mid \theta = \frac{1}{2}\right)} = \frac{9}{4}.$$

(i) If we choose $\frac{3}{4} < k < \frac{9}{4}$, the NP Lemma says that the test which rejects H_0 if $X = 2$ is the UMP level $\alpha = P_{H_0}\left(X = 2 \mid \theta = \frac{1}{2}\right) = \frac{1}{4}$ test,

(ii) If we choose $\frac{1}{4} < k < \frac{3}{4}$, the NP Lemma says that the test which rejects H_0 if $X = 1$ or 2 is the UMP level $\alpha = P_{H_0}\left(X = 1 \text{ or } 2 \mid \theta = \frac{1}{2}\right) = \frac{3}{4}$ test,

(iii) Choosing $k < \frac{1}{4}$ (or, $k > \frac{9}{4}$) yields the UMP level $\alpha = 1$ (or, $\alpha = 0$) test.

If $k = \frac{3}{4}$, then we must reject H_0 for the sample point $X = 2$ and accept H_0 for $X = 0$, but leaves our action for $X = 1$ undetermined. Now, if we accept H_0 for $X = 1$, we get the UMP level $\alpha = \frac{1}{4}$ (test (i) as above). If we reject H_0 for $X = 1$, we get the UMP level $\alpha = \frac{3}{4}$ (test (ii) as above).

This example also shows that for a discrete distribution, the level α at which a test can be done is a function of the particular pmf with which we are dealing. (No such problem arises in the continuous case. Any α level can be attained as we see below.)

Example 15. (UMP normal test) Let $X_1, \ldots, X_n$ be a random sample from a $N\left(\mu, \sigma^2\right)$ population, when σ^2 is known. Consider testing

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu = \mu_1,$$

where $\mu_1 > \mu_0$. The inequality $f_{\mu_1}(x) > k f_{\mu_0}(x)$ is equivalent to

$$\bar{X}_n > \frac{\left(2\sigma^2 \log k\right)/n - \mu_0^2 + \mu_1^2}{2\left(\mu_1 - \mu_0\right)}.$$

The fact that $\mu_1 - \mu_0 > 0$ was used to obtain this inequality. The right hand side increases from $-\infty$ to ∞ as k increases from 0 to ∞. Thus, the test with rejection region $\bar{X}_n > c$ is the UMP level α test, where $\alpha = P_{\mu_0}(\bar{X}_n > c)$. If a particular α is specified, then the UMP test rejects H_0 if $\bar{X}_n > c = \mu_0 + \sigma z_\alpha/\sqrt{n}$. This choice of c ensures that the test is level α.

This is indeed the one sample t-test that we had discussed earlier in Note 14!