1 The Neyman Pearson Lemma

The probability of a type I error is controlled by construction; it is at most α.

A good test should also have a small probability of type II error. In other words it should also be a powerful test.

Definition 1.1 (Most Powerful)**.** *Let $\mathscr{T}$ be a statistical test for testing*

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

with significance level α and maximum power $\gamma = 1 - \beta$ among all the other statistical tests with the same significance level. Such a test $\mathscr{T}$ with the maximum power is said to be the most powerful (MP) test.

The following theorem which is known as the Neyman–Pearson Lemma solve the problem of the existence and construction of MP tests for testing a simple null hypothesis against a simple alternative hypothesis.

Lemma 1.1. *Suppose we have a sample $x_1, x_2, \cdots, x_n$ and have two simple hypotheses H_0 and H_1. Suppose the likelihood, is $\mathscr{L}(H_0)$ under H_0 and $\mathscr{L}(H_1)$ under H_1. The most powerful test of H_0 against H_1 has a critical region of the form*

$$\frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)} \leq \text{ a constant.} \tag{1}$$

*P*roof.

We give the proof for continuous random variables. For discrete random variables just replace integrals with sums. For the definition of the critical region of significance level α,

$$\alpha = P(\mathbf{X} \in C | f_0) = \int \dots \int_C \mathscr{L}(H_0) d\mathbf{x}.$$

Suppose that D is an other critical region of significance level α. In other words,

$$\alpha = \int \dots \int_D \mathscr{L}(H_0) d\mathbf{x}.$$

We will prove that

$$\text{power of} \quad C = \int \dots \int_C \mathscr{L}(H_1) d\mathbf{x} \geq \int \dots \int_D \mathscr{L}(H_1) d\mathbf{x} = \text{power of} \quad D.$$

The following equalities are obvious,

$$C = (C \cap D) \cup (C \cap D') \quad \text{and} \quad D = (D \cap C) \cup (D \cap C').$$

So,
$$\alpha = \int \dots \int_C \mathscr{L}(H_0)\,d\mathbf{x} = \int \dots \int_{C\cap D} \mathscr{L}(H_0)\,d\mathbf{x} + \int \dots \int_{C\cap D'} \mathscr{L}(H_0)\,d\mathbf{x}$$
and
$$\alpha = \int \dots \int_D \mathscr{L}(H_0)\,d\mathbf{x} = \int \dots \int_{D\cap C} \mathscr{L}(H_0)\,d\mathbf{x} + \int \dots \int_{D\cap C'} \mathscr{L}(H_0)\,d\mathbf{x}$$
Therefore
$$\int \dots \int_{C\cap D'} \mathscr{L}(H_0)\,d\mathbf{x} = \int \dots \int_{D\cap C'} \mathscr{L}(H_0)\,d\mathbf{x}. \tag{2}$$
Inside C we have the Neyman-Pearson region so $\frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)} \le k$ (a constant) or $\mathscr{L}(H_1) \ge \mathscr{L}(H_0)/k$, while outside C we have the **non** Neyman-Pearson region $\mathscr{L}(H_1) \le \mathscr{L}(H_0)/k$. Accordingly from Eq. (2),
$$\int \dots \int_{C\cap D'} \mathscr{L}(H_1)\,d\mathbf{x} \ge \int \dots \int_{C\cap D'} \mathscr{L}(H_0)/k\,d\mathbf{x} = \int \dots \int_{D\cap C'} \mathscr{L}(H_0)/k\,d\mathbf{x} \ge \int \dots \int_{D\cap C'} \mathscr{L}(H_1)\,d\mathbf{x} \tag{3}$$
Adding $\int \dots \int_{C\cap D} \mathscr{L}(H_1)\,d\mathbf{x}$ in Eq. (3),
$$\int \dots \int_{C\cap D} \mathscr{L}(H_1)\,d\mathbf{x} + \int \dots \int_{C\cap D'} \mathscr{L}(H_1)\,d\mathbf{x} \ge \int \dots \int_{D\cap C} \mathscr{L}(H_1)\,d\mathbf{x} + \int \dots \int_{D\cap C'} \mathscr{L}(H_1)\,d\mathbf{x} \Longleftrightarrow$$
$$\text{power of} \quad C = \int \dots \int_C \mathscr{L}(H_1)\,d\mathbf{x} \ge \int \dots \int_D \mathscr{L}(H_1)\,d\mathbf{x} = \text{power of} \quad D.$$

■

To reiterate, by most powerful we mean that any other test will have power which is less that or equal to the power of the test based on the Neyman-Pearson critical region. This enables us to construct the critical region of the MP test in the sense that any other test will have *inferior power.*

In the most common case we have a random sample from a distribution $f(x)$. We can then formulate the lemma as:

Proposition 1.1. *Suppose we have a null hypothesis which completely specifies the distribution $f(x)$, say*
$$H_0 : f(x) = f_0(x)$$
where $f_0(x)$ is a known function. We further assume that the alternative is
$$H_1 : f(x) = f_1(x)$$
where again $f_1(x)$ is a known function.

Then the most powerful test of $H_0 : f(x) = f_0(x)$ against $H_1 : f(x) = f_1(x)$ has a critical region of the form
$$\frac{\prod_{i=1}^n f_0(x_i)}{\prod_{i=1}^n f_1(x_i)} \le \textit{ constant} \tag{4}$$

Example 1.1. *Suppose we have one observation x from an exponential distribution $f(x) = \theta\exp(-\theta x)$ and we wish to test*

$$H_0 : \theta = \theta_0 \textit{ against } H_1 : \theta = \theta_1 > \theta_0$$

Then the critical region takes the form

$$\begin{aligned}
C &= \left\{x : \frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)} \le k_1\right\} \\
&= \left\{x : \frac{\theta_0\exp(-\theta_0 x)}{\theta_1\exp(-\theta_1 x)} \le k_1\right\} \\
&= \left\{x : \left(\frac{\theta_0}{\theta_1}\right)\exp\{(\theta_1-\theta_0)x\} \le k_1\right\} \\
&= \left\{x : \exp\{(\theta_1-\theta_0)x\} \le k_2\right\} \\
&= \left\{x : (\theta_1-\theta_0)x \le k_3\right\} \\
&= \left\{x : x \le k_4\right\} \textit{ since } \theta_1 > \theta_0
\end{aligned}$$

For the computation of k_4 we have that

$$\alpha = P(x \le k_4|\theta=\theta_0) = \int_0^{k_4}\theta_0\exp(-\theta_0 x) = 1-\exp(-\theta_0 k_4) \Longleftrightarrow$$

$$k_4 = -\log(1-\alpha)/\theta_0.$$

Therefore the test with statistic x and critical region $\{x : x \le -\log(1-\alpha)/\theta_0\}$ is the MP test of

$$H_0 : \theta = \theta_0 \textit{ against } H_1 : \theta = \theta_1 > \theta_0.$$

Example 1.2. *Suppose we have a sample $x_1, x_2, \cdots, x_n$ from a normal distribution*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

and we wish to test

$$H_0 : \mu = \mu_0 \textit{ against } H_1 : \mu = \mu_1 > \mu_0$$

The two values of the likelihood functions $\mathscr{L}(H_0)$ and $\mathscr{L}(H_1)$ are

$$\mathscr{L}(H_0) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_0)^2\right\}$$

$$\mathscr{L}(H_1) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_1)^2\right\}$$

Then the critical region takes the form

$$
\begin{aligned}
C &= \left\{x : \frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)} \le k_1\right\} \\
&= \left\{x : \exp\left\{\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\mu_1)^2 - \sum_{i=1}^{n}(x_i-\mu_0)^2\right]\right\} \le k_1\right\} \\
&= \left\{x : \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_1)^2 - \sum_{i=1}^{n}(x_i-\mu_0)^2\right) \le k_2\right\} \\
&= \left\{x : n(\mu_1^2-\mu_0^2) + 2(\mu_0-\mu_1)\sum_{i=1}^{n} x_i \le k_3\right\} \\
&= \left\{x : \bar{x} \ge k_4\right\} \textit{ since } \mu_1 > \mu_0
\end{aligned}
$$

For the computation of k_4 *we have that*

$$\alpha = P(\bar{x} \ge k_4 | \mu = \mu_0).$$

We know that $\bar{x} \sim N(\mu, \sigma^2/n)$ *and under the null hypothesis* H_0*,*

$$P\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \ge \frac{k_4-\mu_0}{\sigma/\sqrt{n}}\right) \textit{ or } P(z \ge z_\alpha) \Longleftrightarrow$$

$$k_4 = \mu_0 + (\sigma/\sqrt{n})z_\alpha.$$

Therefore the test with statistic $\bar{x}$ *and critical region* $\{x : \bar{x} \ge \mu_0 + (\sigma/\sqrt{n})z_\alpha\}$ *is the MP test of*

$$H_0 : \mu = \mu_0 \textit{ against } H_1 : \mu = \mu_1 > \mu_0.$$

1.1 Discrete Distributions

You will have noticed that we have avoided discrete distributions. Suppose we have a Binomial problem, so $f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \cdots, n$

$$H_0 : p = p_0 \text{ against } H_1 : p = p_1 < p_0$$

For a given α Neyman Pearson gives the critical region

$$\begin{aligned}
C &= \left\{x : \frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)} \leq k_1\right\} \\
&= \left\{x : \frac{\binom{n}{x} p_0^x (1-p_0)^{n-x}}{\binom{n}{x} p_1^x (1-p_1)^{n-x}} \leq k_1\right\} \\
&= \left\{x : x \log(p_0/p_1) - x \log\left(\frac{1-p_0}{1-p_1}\right) \leq k_2\right\} \\
&= \left\{x : x \log\left(\frac{p_0 - p_1 p_0}{p_1 - p_0 p_1}\right) \leq k_3\right\} \\
&= \left\{x : x \leq k_4\right\} \text{ since } \left(\frac{p_0 - p_1 p_0}{p_1 - p_0 p_1}\right) > 1 \Longleftrightarrow p_1 < p_0
\end{aligned}$$

Now suppose we try to go a bit further, with $p_0 = 0.5$ and $n = 12$. If we choose $\alpha = 0.05$ then

$$\alpha = 0.05 = P[x \leq k_4]$$

If we check the Binomial tables we see that such a k_4 is not possible. We can choose k_4 to give a variety of α values but we cannot have 0.05, as you can see from below

k	0	1	2	3	4	5
$P[X \leq k]$	0.000	0.003	0.019	0.073	0.194	0.387

This is a theoretical problem and some would argue that this invalidates the whole scheme so Neyman-Pearson cannot be applied in this case.

It is possible by using randomization methods to come up with a modified scheme but I have never ever seen it used. In practice one chooses the best α value from those available and we shall refer to these tests as being *non-randomized* when we want to be quite precise.

1.1.1 Randomization procedure

As an aside we demonstrate the randomization process. In the example above we would like $\alpha = 0.5$ but we can only bracket the value with 0.019 and 0.073. Take a *biased* coin where the probability of a head is π. Suppose then we toss the coin and

- If a head choose the critical region based on 0.019.
- If a tail choose the critical region based on 0.073.

Then on average our α is $0.019\pi + 0.073(1-\pi)$. We can make this equal 0.05 simply by choosing $0.019\pi + 0.073(1-\pi) = 0.05$ or $0.023 = 0.054\pi$. So choosing $\pi = 0.426$ solves our problem!

2 Uniformly Most Powerful Tests (UMP)

Definition 2.1 (Uniformly Most Powerful)**.** *Let $\mathscr{T}$ be a statistical test for testing a simple null hypothesis H_0 against a composite H_1 with significance level α. Let also for each $\theta \in H_1$ its power $\gamma(\theta) = 1 - \beta(\theta)$ is more or equal to the power of any other statistical test with the same significance level. Such a test $\mathscr{T}$ with the maximum power is said to be the uniformly most powerful (UMP) test.*

The next theorem introduce a general approach of finding an UMP test.

Theorem 2.1. *Let H_0 be a simple null hypothesis and H_1 a composite alternative hypothesis. Let also the MP test of significance level α for testing the H_0 against the alternative $\theta = \theta_1$ be the same for every $\theta \in H_1$. With other words the test statistic is independent of the special value of the alternative hypothesis. Then this test is the UMP test of significance level α for testing H_0 against the composite H_1.*

Example 2.1. *Suppose we have a sample $x_1, x_2, \cdots, x_n$ from a normal distribution*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

and we wish to test

$$H_0 : \mu = \mu_0 \textit{ against } H_1 : \mu > \mu_0$$

Let $\mu_1 > \mu_0$. We have seen in Example 1.2 that the MP test of

$$H_0 : \mu = \mu_0 \textit{ against } H_1 : \mu = \mu_1 > \mu_0.$$

is the test with statistic $\bar{x}$ and critical region $\{x : \bar{x} \geq \mu_0 + (\sigma/\sqrt{n})z_\alpha\}$. This test is the same for each $\mu_1 > \mu_0$, since its critical region does not depend to μ_1.

There is not always an UMP test. This is easily seen using the same set up as above but with

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu = \mu_1 \neq \mu_0$$

Now we have the two possibilities $\mu_1 < \mu_0$ and $\mu_1 > \mu_0$. Each of these alternatives has an UMP test but they are based on different critical regions so we cannot have a combined UMP test!