Seminar 10 - 2025

The Neyman-Pearson Lemma

Decision	Actual Situation	
	H_0 true	H_1 true
Reject H_0	Type I error (prob. α)	Correct decision
Accept H_0	Correct decision	Type II error (prob. β)

Recall that for a target parameter θ, we are testing

$$H_0 : \theta = \theta_0, \text{ versus one of}$$
$$H_1 : \begin{cases} \theta < \theta_0 \\ \theta > \theta_0 \\ \theta \neq \theta_0 \end{cases}$$

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

$$\alpha = \mathbb{P}(\text{ type I error }) = \mathbb{P}(\text{ reject } H_0 \mid H_0)$$
$$\beta = \mathbb{P}(\text{ type II error }) = \mathbb{P}(\text{ accept } H_0 \mid H_1)$$

The smaller both of them are, the more reliable the test is. In general, α is preset, at most 0.05 and the test is designed so that β is also small enough to be acceptable.

So far, type II errors were not discussed much, as the computation of β 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 θ is less than, greater than or not equal to a value is not enough to help us compute the probability. However, if the alternative H_1 is also a simple hypothesis $H_1 : \theta = \theta_1$, then β can be computed. Thus, β, unlike α, depends on the value specified in the alternative hypothesis, that is $\beta = \beta(\theta_1)$.

Example:

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?

When we looked at this example the test for the mean was $H_0 : \mu = 20, H_1 : \mu < 20$. Now, consider

$$H_0 : \quad \mu = \mu_0 = 20$$
$$H_1 : \quad \mu = \mu_1 = 18 < 20$$

For this test find $\beta(\mu_1)$.

Solution:

We tested a left-tailed alternative for the mean

$$H_0 : \quad \mu = 20$$
$$H_1 : \quad \mu < 20.$$

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

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

then

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1), \text{ if } \mu = \mu_0 \, (H_0 \text{ holds })$$

and

$$Z_1 = \frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1), \text{ if } \mu = \mu_1 \, (H_1 \text{ holds })$$

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

$$\text{reject } H_0 \Longleftrightarrow Z_0 \in RR \Longleftrightarrow \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z_{0.05} \Longleftrightarrow \bar{X} \leq 18.9$$

We compute

$$\beta(\mu_1) = \mathbb{P}(\text{accept } H_0 \mid H_1) = \mathbb{P}\left(\bar{X} > 18.9 \mid \mu = \mu_1\right)$$

Hence,

$$\begin{aligned}
\beta(\mu_1) &= \mathbb{P}\left(\bar{X} > 18.9 \mid \mu = \mu_1\right) \\
&= \mathbb{P}\left(\frac{\bar{X} - 18}{\frac{4}{6}} > \frac{18.9 - 18}{\frac{4}{6}} \middle| \mu = 18\right) \\
&= \mathbb{P}(Z_1 > 1.35 \mid \mu = 18) \\
&= 1 - \mathbb{P}(Z_1 \leq 1.35 \mid \mu = 18) \\
&= 1 - \text{norm.cdf}(1.35) = 1 - 0.9115 = 0.0885 \,(\text{ since } Z_1 \sim N(0,1))
\end{aligned}$$

Remark: Let us take a closer look at the computation of α and β 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 μ is $\mu_0 = 20$, then $Z_0 \sim N(0,1)$ and when the value is $\mu_1 = 18$, then $Z_1 \sim N(0,1)$. However, in the end, we expressed the error probabilities α and β, by looking at the distribution of $\bar{X}$ itself, not its reduced version. In other words, we used the fact that, when the true value of μ is $\mu_0 = 20$, then

$$\bar{X} \sim N\left(\mu_0, \sigma^2/n\right) \text{ and } \alpha = \mathbb{P}(\text{ reject } H_0 \mid H_0) = \mathbb{P}\left(\bar{X} \leq 18.9 \mid H_0\right)$$

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

$$\bar{X} \sim N\left(\mu_1, \sigma^2/n\right) \text{ and } \beta = \beta(\mu_1) = \mathbb{P}(\text{accept} H_0 \mid H_1) = \mathbb{P}\left(\bar{X} > 18.9 \mid H_1\right).$$

In order to have a better control over β, we introduce the following notion.

Definition:

The *power of a test* on a parameter θ, unknown, is the probability of rejecting the null hypothesis

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

when the true value of the parameter is $\theta = \theta^\circ$. If the null hypothesis is true, i.e. $\theta = \theta_0$, then

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

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

$$\begin{aligned}\pi(\theta_1) &= \mathbb{P}(\text{ reject } H_0 \mid \theta = \theta_1) = \mathbb{P}(\text{ reject } H_0 \mid H_1) \\ &= 1 - \mathbb{P}(\text{accept } H_0 \mid H_1) = 1 - \beta(\theta_1)\end{aligned}$$

Remark: By "accept H_0" we actually mean "not reject H_0", everywhere.

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

$$H_0 : \theta = \theta_0, H_1 : \theta = \theta_1,$$

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

$$\pi(\theta_1) = 1 - \beta(\theta_1) \text{ is the largest possible.}$$

Such a test is called a *most powerful test.*

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.

Recall:

The *likelihood function* is defined as:

$$L(x_1, \ldots, x_n; \theta) = \begin{cases} \mathbb{P}(X = x_1) \cdot \ldots \cdot \mathbb{P}(X = x_n), \text{if } X \text{ is discrete} \\ f(x_1) \cdot \ldots \cdot f(x_n), \text{if } X \text{ is continuous with pdf } f \end{cases}$$

i.e. this tells us how likely it is to observe the data set $x_1, x_2, \ldots, x_n$ under specific values of the parameter θ.

Lemma (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 θ the simple hypotheses

$$H_0 : \theta = \theta_0, H_1 : \theta = \theta_1,$$

based on a random sample $X_1, \ldots, X_n$. Let $L(X_1, \ldots, X_n; \theta)$ 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

$$RR = \left\{(x_1, \ldots, x_n) : \frac{L(x_1, \ldots, x_n; \theta_1)}{L(x_1, \ldots, x_n; \theta_0)} \geq k_\alpha\right\}$$

where the constant $k_\alpha > 0$ depends only on α and the sample data.

Example:

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with known variance $\sigma^2 = 1$ but unknown mean μ. The hypotheses are:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1, \quad \text{with } \mu_0 < \mu_1.$$

We aim to design the most powerful test for these simple hypotheses at a fixed significance level $\alpha \in (0, 1)$.

Likelihood Ratio Test

The likelihood function for a sample $X_1, \ldots, X_n$ is:

$$L(X_1, \ldots, X_n; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \mu)^2}.$$

Simplifying, the log-likelihood is:

$$\log L(X_1, \ldots, X_n; \mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Under H_0 and H_1, the log-likelihoods become:

$$\log L(X_1, \ldots, X_n; \mu_0) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_0)^2,$$

$$\log L(X_1, \ldots, X_n; \mu_1) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_1)^2.$$

The likelihood ratio is:

$$\Lambda = \frac{L(X_1, \ldots, X_n; \mu_1)}{L(X_1, \ldots, X_n; \mu_0)}.$$

Taking the log of Λ, we get:

$$\log \Lambda = -\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_1)^2 + \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_0)^2.$$

Expanding the squared terms:

$$\log \Lambda = -\frac{1}{2}\left[\sum_{i=1}^{n} X_i^2 - 2\mu_1\sum_{i=1}^{n} X_i + n\mu_1^2\right] + \frac{1}{2}\left[\sum_{i=1}^{n} X_i^2 - 2\mu_0\sum_{i=1}^{n} X_i + n\mu_0^2\right].$$

Simplifying:

$$\log \Lambda = n(\mu_1 - \mu_0)\overline{X} - \frac{n}{2}(\mu_1^2 - \mu_0^2),$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean.

We begin with the test criterion from the Neyman-Pearson Lemma. The null hypothesis H_0 is rejected if:

$$\frac{L(X_1, \ldots, X_n; \mu_1)}{L(X_1, \ldots, X_n; \mu_0)} \geq k_\alpha,$$

where $k_\alpha > 0$ is a threshold determined by the significance level α. Taking the natural logarithm (which preserves the inequality) leads to:

$$\log \Lambda = n(\mu_1 - \mu_0)\overline{X} - \frac{n}{2}(\mu_1^2 - \mu_0^2) \geq \log k_\alpha.$$

Rearranging the terms, we isolate the sample mean $\overline{X}$:

$$n(\mu_1 - \mu_0)\overline{X} \geq \log k_\alpha + \frac{n}{2}(\mu_1^2 - \mu_0^2).$$

Dividing through by $n(\mu_1 - \mu_0)$ (noting that $\mu_1 > \mu_0$, so $n(\mu_1 - \mu_0) > 0$):

$$\overline{X} \geq \frac{\log k_\alpha + \frac{n}{2}(\mu_1^2 - \mu_0^2)}{n(\mu_1 - \mu_0)}.$$

Letting the right-hand side be c, we simplify the test criterion to:

$$\overline{X} \geq c,$$

where:

$$c = \frac{\log k_\alpha + \frac{n}{2}(\mu_1^2 - \mu_0^2)}{n(\mu_1 - \mu_0)}.$$

Distribution of $\overline{X}$ Under H_0

Under the null hypothesis H_0, the sample mean $\overline{X}$ follows a normal distribution:

$$\overline{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right).$$

Thus, the probability of rejecting H_0 is determined by the upper tail of this distribution. For a fixed significance level α, the threshold c is chosen so that:

$$\mathbb{P}(\overline{X} \geq c \mid H_0) = \alpha.$$

Standardizing $\overline{X}$ under H_0:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

The critical value z_α is the point in the standard normal distribution such that:

$$\mathbb{P}(Z \geq z_\alpha) = \alpha.$$

Rewriting the rejection criterion in terms of Z:

$$\overline{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Rejection Region

The rejection region for the test is therefore:

$$RR = \left\{ \overline{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\}.$$

Here:

- z_α is the critical value from the standard normal distribution corresponding to the significance level α,
- μ_0 is the mean under the null hypothesis,
- $\frac{\sigma}{\sqrt{n}}$ is the standard error of the sample mean.

In summary, the rejection region depends on the sample mean $\overline{X}$ and is determined by the critical threshold $\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

This example illustrates the likelihood ratio test applied to a normal distribution with known variance and unknown mean. The rejection region is based on the sample mean, and the test is most powerful for the given hypotheses.

Intuition Behind the Neyman-Pearson Lemma

The Neyman-Pearson Lemma (NPL) provides a fundamental result in hypothesis testing, offering a clear and optimal strategy for distinguishing between two simple hypotheses H_0 and H_1. Here's the intuition in words:

1. **Likelihood Ratio as Evidence**:
   - The likelihood function $L(X_1, \ldots, X_n; \theta)$ measures how plausible the observed data is under a given parameter value θ.
   - The likelihood ratio $\frac{L(X_1,\ldots,X_n;\theta_1)}{L(X_1,\ldots,X_n;\theta_0)}$ compares how much more likely the observed data is under θ_1 (corresponding to H_1) than under θ_0 (corresponding to H_0).
   - A large likelihood ratio suggests that the data is more consistent with θ_1 than θ_0, making it reasonable to reject H_0 in favor of H_1.
2. **Rejection Region**:
   - The test is based on defining a rejection region RR, where we reject H_0 if the likelihood ratio exceeds a threshold k_α.
   - This threshold k_α is chosen to ensure that the probability of rejecting H_0 when H_0 is true (Type I error) is equal to the fixed significance level α.
3. **Optimality**:

- The Neyman-Pearson Lemma guarantees that this test is the most powerful for the given significance level α. This means it maximizes the probability of correctly rejecting H_0 (when H_1 is true) among all possible tests with the same Type I error rate α.
- In other words, it provides the best possible trade-off between Type I and Type II errors for these specific hypotheses.

4. **Interpretation**:

   - The test boils down to a direct comparison: is the evidence in favor of H_1 strong enough relative to H_0 to justify rejecting H_0?
   - The likelihood ratio serves as a natural and intuitive measure of this "relative evidence," making the test both mathematically elegant and practically meaningful.

In summary, the Neyman-Pearson Lemma formalizes the idea that the likelihood ratio provides the most efficient way to decide between two competing simple hypotheses, ensuring that the decision rule is as effective as possible given the constraints on error rates.