Uniformly Most Powerful Tests and the Neyman-Pearson Lemma

Let’s return to the hypothesis testing problem within the Neyman-Pearson framework. Recall that we have a random variable \( X \), with PDF/PMF \( f_X(x; \theta) \), and we have a null and alternative hypothesis:

\[
H_0 : \quad \theta \in \Theta_0, \\
H_a : \quad \theta \in \Theta_0^c.
\]

We need to construct a test statistic \( T(X) \) and a critical region \( C_T \), such that we reject the null hypothesis if \( T(X) \in C_T \).

The power function of a test is defined as

\[
\beta(\theta) = \Pr_{\theta}(T(X) \in C_T)
\]

Given a prespecified significance level \( \alpha \) (for example \( \alpha = .05 \)), we require our test to satisfy, for all \( \theta \in \Theta_0 \),

\[
\beta(\theta) \leq \alpha.
\]

Subject to this restriction, we want \( \beta(\theta) \) for \( \theta \in \Theta_0^c \) to be as large as possible.

Now we define a criterion that will measure optimality of a test. It requires that the probability of a type II error is minimized for all values of the parameter consistent with the alternative hypothesis.

**Definition 1** Consider all tests of level \( \alpha \) for the null hypothesis \( \theta \in \Theta_0 \) against the alternative \( \theta \in \Theta_0^c \). A test with power function \( \beta(\theta) \) is uniformly most powerful if, for all alternative tests with level \( \alpha \) and power function \( \beta'(\theta) \), \( \beta(\theta) \geq \beta'(\theta) \) for all \( \theta \in \Theta_0^c \).

There is no guarantee that uniformly most powerful tests actually exist. We first study a simple case where such tests are easy to find. We focus on the case where both the null hypothesis and the alternative hypothesis are simple, that is, where the sets \( \Theta_0 \) and \( \Theta_0^c \) contain a single element each:

\[
H_0 : \quad \theta = \theta_0, \\
H_a : \quad \theta = \theta_1.
\]
(If a hypothesis contains more than a single point, we say that it is a composite hypothesis.)

**Result 1** *(Neyman–Pearson lemma)*

Consider testing the null hypothesis \( H_0 : \theta = \theta_0 \) against the alternative \( H_a : \theta = \theta_1 \) using a critical region of the form

\[
C_X = \{ x : f_X(x; \theta_1) \geq k \cdot f_X(x; \theta_0) \}.
\]

Let

\[
\alpha = \int_{C_X} f_X(x; \theta_0) \, dx.
\]

This test is the uniformly most powerful test of level \( \alpha \).

**Proof:** Let \( \beta(\theta) \) denote the power function of the test proposed. Consider any other test with a critical region \( C_X' \) and power function \( \beta'(\theta) \). Define

\[
\phi(x) = 1\{ x \in C_X \},
\]

and

\[
\phi'(x) = 1\{ x \in C_X' \}.
\]

Consider

\[
(\phi(x) - \phi'(x)) \cdot (f_X(x; \theta_1) - k \cdot f_X(x; \theta_0)).
\]

If this expression differs from zero, we must either have \( \phi(x) - \phi'(x) = 1 \) or \( \phi(x) - \phi'(x) = -1 \).

If \( \phi(x) - \phi'(x) = 1 \), \( (f_X(x; \theta_1) - k \cdot f_X(x; \theta_0)) \) must be nonnegative by the form of the critical region \( C_X \), so the entire expression is nonnegative.

If \( \phi(x) - \phi'(x) = -1 \), the second factor must be \( \leq 0 \), and the product again is nonnegative.

Hence,

\[
(\phi(x) - \phi'(x)) \cdot (f_X(x; \theta_1) - k \cdot f_X(x; \theta_0)) \geq 0,
\]

and therefore

\[
\int_X (\phi(x) - \phi'(x)) \cdot (f_X(x; \theta_1) - k \cdot f_X(x; \theta_0)) \, dx \\
\int_X (\phi(x) - \phi'(x)) \cdot f_X(x; \theta_1) \, dx - k \cdot \int_X (\phi(x) - \phi'(x)) \cdot f_X(x; \theta_0) \, dx \\
= \beta(\theta_1) - \beta'(\theta_1) - k \cdot (\beta(\theta_0) - \beta'(\theta_0)) \geq 0.
\]

If both tests are level \( \alpha \) tests, \( \beta(\theta_0) = \beta'(\theta_0) = \alpha \), and so it must be the case that

\[
\beta(\theta_1) - \beta'(\theta_1) \geq 0,
\]

and the second test cannot be the most powerful test. □
Example 1
Let us consider some examples of applications of the Neyman-Pearson Lemma. Suppose $X$ has an exponential distribution with arrival rate $\lambda$. We wish to test the hypothesis that $\lambda = 1$ against the alternative that $\lambda = 2$:

$$H_0 : \lambda = 1;$$

$$H_a : \lambda = 2.$$ 

By the Neyman-Pearson lemma, we should use a critical region of the form

$$C_X = \{ x : f_X(x; 2) \geq k \cdot f_X(x; 1) \}$$

$$= \{ x : 2 \cdot \exp(-2x) \geq k \cdot \exp(-x) \}$$

$$= \{ x : \exp(-2x) \geq k' \exp(-x) \}$$

$$= \{ x : -2x \geq k' - x \}$$

$$= \{ x : x \leq k'' \}.$$

All that is left to determine is $k''$. Suppose we wish to test at the 0.05 level. Then we choose $k''$ to satisfy

$$0.05 = Pr(X \leq k'' | H_0)$$

$$= \int_0^{k''} \exp(-x) dx$$

$$= 1 - \exp(-k''),$$

or

$$k'' = -\ln(0.95) \approx 0.0513,$$

and the critical region is

$$C_X = [0, -\ln(0.95)].$$

□

Example 2
Suppose $X_1, \ldots, X_n$ are iid normal with mean $\mu$ and unit variance. We wish to test the null hypothesis $\mu = \mu_0$ against the alternative hypothesis that $\mu = \mu_1$, for some $\mu_1$ and $\mu_0$ with $\mu_1 > \mu_0$:

$$H_0 : \mu = \mu_0;$$

$$H_a : \mu = \mu_1.$$
By Neyman-Pearson, we want the test to reject the null if

\[ f(x_1, \ldots, x_n; \mu_1) \geq k \cdot f(x_1, \ldots, x_n; \mu_0) \]

or equivalently:

\[ \frac{f(x_1, \ldots, x_n; \mu_1)}{f(x_1, \ldots, x_n; \mu_0)} \geq k. \]

This ratio of likelihood functions is

\[
\frac{\mathcal{L}(\mu_1)}{\mathcal{L}(\mu_0)} = \frac{\exp\left(-\frac{1}{2} \sum_i (x_i - \mu_1)^2\right)}{\exp\left(-\frac{1}{2} \sum_i (x_i - \mu_0)^2\right)} \\
= \frac{\exp\left(-\frac{1}{2} \sum_i [x_i^2 - 2x_i\mu_1 + \mu_1^2]\right)}{\exp\left(-\frac{1}{2} \sum_i [x_i^2 - 2x_i\mu_0 + \mu_0^2]\right)} \\
= \exp\left((\mu_1 - \mu_0) \sum_i x_i\right) \cdot C,
\]

where \(C\) is a constant which does not depend on \(x\). Since \(\mu_1 - \mu_0 > 0\), this ratio is larger than \(k\) if and only if

\[ \sum_i x_i \geq k', \]

or equivalently,

\[ \bar{x} \equiv \frac{1}{n} \sum_i x_i \geq k''. \]

The critical region is therefore of the form

\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \geq k''\}. \]

Suppose we wish to test at the 0.05 level. Then

\[ 0.05 = Pr(\bar{x} \geq k''|\mu = \mu_0). \]

Under the null the distribution of \(\bar{x}\) is normal with mean \(\mu_0\) and variance \(1/n\):

\[ \bar{x} \sim N(\mu_0, \frac{1}{n}), \]

so

\[ \frac{\bar{x} - \mu_0}{\sqrt{1/n}} \sim N(0, 1). \]

Using a table for the standard normal distribution, we can determine that

\[ Pr\left(\frac{\bar{x} - \mu_0}{\sqrt{1/n}} \geq 1.645\right) = 0.05. \]
So
\[ Pr \left( \bar{x} - \mu_0 \geq \frac{1.645}{\sqrt{n}} \right) = 0.05, \]
and
\[ Pr \left( \bar{x} \geq \mu_0 + \frac{1.645}{\sqrt{n}} \right) = 0.05. \]
Hence the critical region should be
\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} > \mu_0 + \frac{1.645}{\sqrt{n}} \}. \]

Example 2 also illustrates an important phenomenon. There, the critical region does not depend on the value of the parameter under the alternative hypothesis, \( \mu_1 \). Whether the alternative is \( \mu_1 = \mu_0 + 1 \) or \( \mu_1 = \mu_0 + 4 \) leads to exactly the same critical region. Thus, we can use the same test if we are testing the composite alternative hypothesis \( H_a : \mu > \mu_0 \). Moreover, since the test is most powerful for each specific point in the alternative, the test is uniformly most powerful against the composite alternative.

Uniformly most powerful tests do not always exist. They exist for some special models like the normal model, when the alternative is “one-sided” (i.e. \( H_a : \mu > \mu_0 \) or \( H_a : \mu < \mu_0 \)). What if we consider the same normal model, and test
\[ H_0 : \mu = \mu_0, \]
against the two-sided alternative
\[ H_1 : \mu \neq \mu_0. \]
If the alternative is \( \mu = \mu_1 > \mu_0 \) the critical region for the most powerful test is of the form
\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \geq k \}. \]
If the alternative is \( \mu = \mu_1 < \mu_0 \) the critical region of the most powerful test is of the form
\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \leq k \}. \]
There is therefore no test that is most powerful for all values under the alternative. In other words, there is no uniformly most powerful test. One way to get around this problem, is to impose some additional restrictions on the test, and look for uniformly most powerful tests within the restricted set of tests.

A test is unbiased if the power function \( \beta(\theta_1) \geq \beta(\theta_0) \) for all \( \theta_1 \in \Theta_0^c \) and all \( \theta_0 \in \Theta_0 \). That is, the probability of rejecting the null hypothesis, or of an observation in the critical region, is at least
as large for values of the parameters consistent with the alternative \((\theta \in \Theta^c_0)\) as for values of the parameters consistent with the null hypothesis \((\theta \in \Theta_0)\).

Let us consider this approach in detail for the case with a normal distribution with unknown mean and known variance. Let \(X_1, \ldots, X_n\) be independent and normally distributed with unknown mean \(\mu\) and known variance \(\sigma^2\). We are interested in testing the null hypothesis

\[ H_0 : \mu = \mu_0, \]

against the alternative

\[ H_1 : \mu \neq \mu_0. \]

Let us consider the ratio of density functions to determine the critical region:

\[
\frac{f(x_1, \ldots, x_n | \mu_1)}{f(x_1, \ldots, x_n | \mu_0)} = \frac{(2\pi\sigma^2)^{n/2} \exp\left(\frac{-1}{2\sigma^2} (\sum_{i=1}^n x_i^2 - 2\mu_1 \sum_{i=1}^n x_i + n\mu_1^2) \right)}{(2\pi\sigma^2)^{n/2} \exp\left(\frac{-1}{2\sigma^2} (\sum_{i=1}^n x_i^2 - 2\mu_0 \sum_{i=1}^n x_i + n\mu_0^2) \right)}
\]

\[= \exp\left(\frac{1}{\sigma^2} \cdot (\mu_1 - \mu_0) \cdot \sum_{i=1}^n x_i \right) \cdot \exp\left(\frac{- (\mu_1^2 - \mu_0^2)}{n/2}\right).\]

Hence if we are looking for a uniformly most powerful test against the alternative hypothesis \(H_1 : \mu > \mu_0\), the critical region ought to be of the form

\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \geq k\}. \]

If we were to test against the alternative hypothesis \(H_1 : \mu < \mu_0\), the critical region ought to be of the form

\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \leq k\}. \]

It therefore appears sensible to base a test on the value of \(\bar{x}\), the sample average, which is a sufficient statistic for \(\mu\). It seems fairly clear that the critical region should be of the form

\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \leq a \text{ or } \bar{x} \geq b\}. \]

Unbiasedness of the test implies that

\[ 1 - \beta(\mu) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left(\frac{-1}{2\sigma^2/n} (\bar{x} - \mu)^2 \right) d\bar{x}, \]

is maximized at \(\mu_0\). The function is maximized at \(\mu = (a + b)/2\), so that for unbiasedness we must have \(b - \mu_0 = \mu_0 - a\). Hence the critical region is

\[ C_X = \{(x_1, \ldots, x_n) : \bar{x} \leq \mu_0 - c \text{ or } \bar{x} \geq \mu_0 + c\}, \]
with the value of $c$ determined by the size of the test. Under the null hypothesis the distribution of $\bar{x}$ is normal with mean $\mu_0$ and variance $\sigma^2/n$. Hence, if we wish to test at the 10% level, recalling that for a standard normal random variable $Z$

$$Pr(-1.645 < Z < 1.645) = 0.90,$$

the critical region is

$$C_X = \{(x_1, \ldots, x_n) : \bar{x} \leq \mu_0 - 1.645 \cdot \sigma / \sqrt{n}, \bar{x} \geq \mu_0 + 1.645 \cdot \sigma / \sqrt{n}\}.$$  

This is the uniformly most powerful unbiased test.

If we wish to test at the 5% level, the critical region is

$$C_X = \{(x_1, \ldots, x_n) : \bar{x} \leq \mu_0 - 1.96 \cdot \sigma / \sqrt{n}, \bar{x} \geq \mu_0 + 1.96 \cdot \sigma / \sqrt{n}\}. $$

Equivalently we can use the critical region

$$C_X = \{(x_1, \ldots, x_n) : n \cdot (\bar{x} - \mu_0)^2 / \sigma^2 \geq 3.84 \},$$

which uses the Chi–squared distribution for the square of a standard normal random variable. In fact a common way of doing the test is to calculate the test statistic, here $n \cdot (\bar{x} - \mu_0)^2 / \sigma^2$ which under the null hypothesis has a known distribution, in this case a $\chi^2(1)$ distribution. We reject the null hypothesis if the test statistic exceeds the critical value, in this case 3.84 at the 5% level or 2.706 at the 10% level.