Hypothesis testing

• A hypothesis test is a claim about a parameter of a population.

• Given the data we want to make a decision about which of two hypothesis is true (or not true).

• The two hypotheses are called the “null” and “alternative” hypotheses (denoted with $H_0$ and $H_a$ respectively).

• The test has the following formulation:
  $H_0: \theta \in \Theta_0$
  $H_a: \theta \in \Theta'_0$, where $\Theta'_0$ is the complement of $\Theta_0$.

• Examples:
  1. Consider the simple regression model: $y_i = \beta_0 + \beta x_i + \epsilon_i$. We wish to test
     $H_0: \beta_1 = 0$ (the null hypothesis states that there is no association between the response $y$ and the predictor $x$).
     $H_a: \beta_1 \neq 0$ (the alternative hypothesis states that there is a linear association between $y$ and $x$).
  2. Consider an experiment in which a patient is given a treatment (some drug) and we want to test if there is a difference between before and after administrating the drug. We wish to test
     $H_0: \mu_d = 0$ (the null hypothesis states that there is no difference).
     $H_a: \mu_d \neq 0$ (the alternative hypothesis states that there is a difference).
  3. Consider an experiment where the goal is to see if on average there is a difference in the production of corn using different fertilizers. We wish to test
     $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ (the null hypothesis states that the production is the same under the different fertilizers (treatments)).
     $H_a: \text{At least two means are not equal}$ (the alternative hypothesis states that there are differences).
  4. Test for the proportion of defective items at a certain production line:
     $H_0: p = p_0$
     $H_A: p > p_0$. 
• We need to find and evaluate hypothesis tests.

• Find a procedure that will tell us for which sample values $H_0$ is accepted (and therefore for which sample values $H_0$ is rejected). These are called the acceptance region (accepts $H_0$) and the rejection region (rejects $H_0$).

• Usually the procedure of rejecting (or accepting) involves the so called test statistic $T(X)$ which a function of the data $X = (X_1, \ldots, X_n)'$.

• Type I and Type II error

<table>
<thead>
<tr>
<th>STATISTICAL DECISION</th>
<th>ACTUAL SITUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$ IS TRUE</td>
</tr>
<tr>
<td>DO NOT REJECT $H_0$</td>
<td>Correct Decision</td>
</tr>
<tr>
<td></td>
<td>$1 - \alpha$</td>
</tr>
<tr>
<td>REJECT $H_0$</td>
<td>Type I Error</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
</tr>
</tbody>
</table>

• Testing a simple hypothesis, i.e. $H_0 : \theta = \theta'$ against $H_a : \theta = \theta''$.

1. Best critical region of size $\alpha$.
   Definition: Let $R$ denote a subset of the sample space. Then $R$ is called “best critical region” of size $\alpha$ for testing the simple hypothesis $H_0 : \theta = \theta'$ against $H_a : \theta = \theta''$ if for every subset $D$ of the sample space for which $P[(X_1, \ldots, X_n) \in D|H_0] = \alpha$ it is true that
   a. $P[(X_1, \ldots, X_n) \in R|H_0] = \alpha$.
   b. $P[(X_1, \ldots, X_n) \in R|H_0] \geq P[(X_1, \ldots, X_n) \in D|H_0]$.
   Explanation:
   In general, there are many subsets $D$ for which $P[(X_1, \ldots, X_n) \in D|H_0] = \alpha$, but there is one of these subsets, denoted with $R$, such that the power of the test associated with $R$ is larger than any other subset $D$.

2. Example:
   Suppose $X \sim b(5, p)$. We want to test $H_0 : p = \frac{1}{2}$ against $H_a : p = \frac{3}{4}$ using one random value of $X$. We list all the probabilities of $b(5, \frac{1}{2})$ and $b(5, \frac{3}{4})$ in the next table:

| $x$  | $P(X = x| p = \frac{1}{2})$ | $P(X = x| p = \frac{3}{4})$ |
|------|-----------------------------|-----------------------------|
|      | $0$                         | $\frac{1}{32}$              |
|      | $1$                         | $\frac{5}{32}$              |
|      | $2$                         | $\frac{10}{32}$             |
|      | $3$                         | $\frac{10}{32}$             |
|      | $4$                         | $\frac{5}{32}$              |
|      | $5$                         | $\frac{1}{32}$              |
|      | $\frac{1}{1024}$           | $\frac{15}{1024}$           |
|      | $\frac{10}{1024}$          | $\frac{90}{1024}$           |
|      | $\frac{270}{1024}$         | $\frac{405}{1024}$          |
|      | $\frac{243}{1024}$         | $\frac{243}{1024}$          |
Suppose we decided to use $\alpha = \frac{1}{32}$. We want to find the best critical region of size $\alpha = \frac{1}{32}$. We observe that $P(X = 0|p = \frac{1}{2}) = \frac{1}{32}$ and $P(X = 5|p = \frac{1}{2}) = \frac{1}{32}$. Therefore, there are two subsets $D_1(x = 0)$ and $D_2(x = 5)$, for which $P(X \in D_1|H_0) = \frac{1}{32}$ and $P(X \in D_2|H_0) = \frac{1}{32}$. One of these subsets will be our best critical region. Which one of these two subsets has the largest power? We compute: $P(X = 0|p = \frac{3}{4}) = \frac{1}{1024}$ and $P(X = 5|p = \frac{3}{4}) = \frac{243}{1024}$, therefore the best critical region of size $\alpha = \frac{1}{32}$ is $R = \{x = 5\}$.

We also observe that the best critical region of size $\alpha = \frac{1}{32}$ corresponds to the point in $D$ for which $\frac{P(X = x|p = \frac{1}{2})}{P(X = x|p = \frac{3}{4})}$ is the minimum. We see this in the next table where we compute the ratios $\frac{P(X = x|p = \frac{1}{2})}{P(X = x|p = \frac{3}{4})}$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x</td>
<td>p = \frac{1}{2})$</td>
<td>$\frac{1}{32}$</td>
<td>$\frac{5}{32}$</td>
<td>$\frac{10}{32}$</td>
<td>$\frac{10}{32}$</td>
<td>$\frac{5}{32}$</td>
</tr>
<tr>
<td>$P(X = x</td>
<td>p = \frac{3}{4})$</td>
<td>$\frac{1}{1024}$</td>
<td>$\frac{15}{1024}$</td>
<td>$\frac{90}{1024}$</td>
<td>$\frac{270}{1024}$</td>
<td>$\frac{405}{1024}$</td>
</tr>
<tr>
<td>$\frac{P(X = x</td>
<td>p = \frac{1}{2})}{P(X = x</td>
<td>p = \frac{3}{4})}$</td>
<td>32</td>
<td>$\frac{32}{3}$</td>
<td>$\frac{32}{9}$</td>
<td>$\frac{32}{27}$</td>
</tr>
</tbody>
</table>

Another example: Suppose $\alpha = \frac{6}{32}$. Find the best critical region of size $\alpha = \frac{6}{32}$. 

3
3. Neyman-Pearson theorem:
Suppose $X$ is a random variable and we need to decide whether the probability distribution is either $f_0(x)$ or $f_1(x)$. For example, we may want to test that $f_0(x)$ is $N(18, 1)$ against the alternative that $f_1(x)$ is $N(28, 1)$.

Let $k$ be some positive number, and define the following two sets:

$$A = \left\{ x \mid \frac{f_0(x)}{f_1(x)} > k \right\}$$

and

$$R = \left\{ x \mid \frac{f_0(x)}{f_1(x)} < k \right\}$$

The Neyman-Pearson decision rule is the following:
If data $x$ is in set $A$, then accept $H_0$.
If data $x$ is in set $R$, then accept $H_a$.

Let $\alpha$ be the probability of Type I error based on $A$ and $R$ above.

Therefore for the Neyman-Pearson Lemma we have:

$$\alpha = \int_R f_0(x)dx \quad \text{and} \quad 1 - \alpha = \int_A f_0(x)dx,$$

and

$$\beta = \int_A f_1(x)dx$$

Suppose that there is a competitor test with acceptance region $A^*$ and rejection region $R^*$, such that $\alpha^* \leq \alpha$.

Therefore for this competitor test we have:

$$\alpha^* = \int_{R^*} f_0(x)dx \quad \text{and} \quad 1 - \alpha^* = \int_{A^*} f_0(x)dx,$$

and

$$\beta^* = \int_{A^*} f_1(x)dx$$

The Neyman-Pearson Lemma claims that this test is the best, in the sense that any other competitor test with Type I error $\alpha^*$ such that $\alpha^* \leq \alpha$ will have higher probability of Type II error. Therefore, $\beta^* - \beta \geq 0$.

Proof:

This is the entire data space:
The value of $X$ must fall here.

This is the data space partitioned by the Neyman-Pearson Lemma:

This is data set $A$.

This is data set $R$.

This is the data space partitioned by the competitor test:
This is set $A^*$.

This is set $R^*$.

This is the final picture showing the partitioned of the data space based on the two tests:

\[
\begin{array}{c|c}
A \cap A^* & A \cap R^* \\
\hline
R \cap A^* & R \cap R^* \\
\end{array}
\]

We need to calculate $\beta^* - \beta$:

\[
\beta^* - \beta = \int_{A^*} f_1(x) dx - \int_A f_1(x) dx \\
= \int_{(A^* \cap A) \cup (A^* \cap R)} f_1(x) dx - \int_{(A \cap A^*) \cup (A \cap R^*)} f_1(x) dx \\
= \int_{A^* \cap A} f_1(x) dx + \int_{A^* \cap R} f_1(x) dx - \left\{ \int_{A \cap A^*} f_1(x) dx + \int_{A \cap R^*} f_1(x) dx \right\} \\
= \int_{A^* \cap R} f_1(x) dx - \int_{A \cap R^*} f_1(x) dx.
\]
For the first integral \( \int_{A^* \cap R} f_1(x) \, dx \), it should be true that \( \frac{f_0(x)}{f_1(x)} < k \) because it is done over the subset \( R \). Therefore the integral will be smaller if we replace \( f_1(x) \) with \( \frac{f_0(x)}{k} \) since \( f_1(x) > \frac{f_0(x)}{k} \).

For the second integral \( \int_{A \cap R^*} f_1(x) \, dx \), it should be true that \( \frac{f_0(x)}{f_1(x)} > k \) because it is done over the subset \( A \). Therefore the integral will be larger if we replace \( f_1(x) \) with \( \frac{f_0(x)}{k} \) since \( f_1(x) < \frac{f_0(x)}{k} \).

These two changes above will give us:

\[
\beta^* - \beta \geq \int_{A^* \cap R} \frac{1}{k} f_0(x) \, dx - \int_{A \cap R^*} \frac{1}{k} f_0(x) \, dx
\]

\[
= \frac{1}{k} \int_{A^* \cap R} f_0(x) \, dx - \frac{1}{k} \int_{A \cap R^*} f_0(x) \, dx.
\]

Now, add an subtract \( \frac{1}{k} \int_{A \cap A^*} f_0(x) \, dx \) to get:

\[
\beta^* - \beta \geq \frac{1}{k} \int_{A^* \cap R} f_0(x) \, dx + \frac{1}{k} \int_{A \cap A^*} f_0(x) \, dx
\]

\[- \frac{1}{k} \int_{A \cap R^*} f_0(x) \, dx - \frac{1}{k} \int_{A \cap A^*} f_0(x) \, dx.
\]

Because, \( A^* = (A^* \cap A) \cup (A^* \cap R) \) and \( A = (A \cap A^*) \cup (A \cap R^*) \) we finally get:

\[
\beta^* - \beta \geq \frac{1}{k} \int_{A^*} f_0(x) \, dx - \frac{1}{k} \int_{A} f_0(x) \, dx
\]

\[
\geq \frac{1}{k}(1 - \alpha^*) - \frac{1}{k}(1 - \alpha)
\]

\[
\geq \frac{1}{k}(\alpha - \alpha^*) \geq 0.
\]

Therefore, the competitor test with equal or better Type I error probability must have larger Type II error probability.

4. Neyman-Pearson theorem. Summary and examples:

Suppose we wish to test the simple hypothesis

\( H_0 : \theta = \theta_0 \)

against the alternative simple hypothesis

\( H_a : \theta = \theta_a \).

As always, a sample of \( X_1, X_2, \ldots, X_n \) is selected from a probability distribution with unknown parameter \( \theta \). Let \( L(\theta_0) \) denote the likelihood function when \( \theta = \theta_0 \) and \( L(\theta_a) \) denote the likelihood function when \( \theta = \theta_a \). Then for a given significance level \( \alpha \), the test that maximizes the power has a rejection region determined by \( \frac{L(\theta_0)}{L(\theta_a)} < k \), where \( k \) is some constant. This test will be the most powerful test for testing \( H_0 \) against \( H_a \).
The previous result applies to simple hypotheses. Usually one of the two hypotheses is composite. For example: $H_0: \theta = \theta_0$
against the alternative composite hypothesis
$H_a: \theta > \theta_0$.
We say that a test that is most powerful for every simple alternative in $H_a$ is uniformly most powerful.

**Example 1:**
Let $X$ be a single observation from the probability density function $f(x) = \theta x^{\theta-1}, 0 < x < 1$. Find the most powerful test using significance level $\alpha = 0.05$ for testing
$H_0: \theta = 1$
$H_a: \theta = 2$.

**Example 2:**
Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma)$, with known $\sigma^2$. Find the uniformly most powerful test using significance level $\alpha$ for testing
$H_0: \mu = \mu_0$
$H_a: \mu > \mu_0$. 

Example 3:
Let \( X \sim exp \left( \frac{1}{\lambda} \right) \). Therefore, \( f(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda} x}, \lambda > 0, x > 0 \). Let \( X_1, X_2, \ldots, X_n \) be a random sample from this distribution.

a. Show that the best critical region for testing
\[
H_0 : \lambda = 3 \\
H_a : \lambda = 5
\]
is based on \( \sum_{i=1}^{n} x_i \).

b. If \( n = 12 \) and using \( \frac{2}{3} \sum_{i=1}^{n} x_i \sim \chi^2_{24} \) find the best critical region when the significance level \( \alpha = 0.05 \).
Example 4:
Suppose $X$ has the possible values 0,1,2,3,4. Suppose that the null hypothesis says that $X$ is uniform on these integers, while the alternative hypothesis says that $X \sim b(4, \frac{1}{2})$. Let’s see what happens if we let $k$ of the Neyman-Pearson lemma be equal to 0.6. The following table will be help us find the best critical region when $k = 0.6$ and the power of the test.

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x</td>
<td>H_0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(X = x</td>
<td>H_a)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{P(X=x</td>
<td>H_0)}{P(X=x</td>
<td>H_a)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>