Topic 17
Simple Hypotheses
Terminology and the Neyman-Pearson Lemma
Outline

Overview

Terminology

The Neyman-Pearson Lemma
Overview

Statistical hypothesis testing is designed to address the question:

*Do the data provide sufficient evidence to conclude that we must depart from our original assumption concerning the state of nature?*

The logic of hypothesis testing is similar to the one a juror faces in a criminal trial:

*Is the evidence provided by the prosecutor sufficient for the jury to depart from its original assumption that the defendant is not guilty of the charges brought before the court?*

Two of the jury’s possible actions are

- Find the defendant guilty.
- Find the defendant not guilty.

Given the level of evidence needed, a prosecutor’s task is to present the evidence in the most powerful and convincing manner possible.
The simplest set-up for understanding the issues of statistical hypothesis, is the case of two values $\theta_0$ and $\theta_1$ in the parameter space. We write the test, known as a simple hypothesis as

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

$H_0$ is called the null hypothesis. $H_1$ is called the alternative hypothesis.

The possible actions are

- Reject the hypothesis.
- Fail to reject the hypothesis.
### Terminology

#### Criminal Trials

<table>
<thead>
<tr>
<th></th>
<th>Innocent</th>
<th>Guilty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convict</td>
<td></td>
<td>OK</td>
</tr>
<tr>
<td>Do not convict</td>
<td>OK</td>
<td></td>
</tr>
</tbody>
</table>

#### Hypothesis Tests

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is True</th>
<th>$H_1$ is True</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td>Type I Error</td>
<td>OK</td>
</tr>
<tr>
<td>Fail to reject $H_0$</td>
<td>OK</td>
<td>Type II Error</td>
</tr>
</tbody>
</table>
Terminology

- Rejecting the hypothesis when it is true is called a type I error or a false positive. Its probability $\alpha$ is called the size of the test or the significance level. In symbols, we write

$$\alpha = P_{\theta_0}\{\text{reject } H_0\}.$$  

- Failing to reject the hypothesis when it is false is called a type II error or a false negative, has probability $\beta$. The power of the test, $1 - \beta$, the probability of rejecting the test when it is indeed false, is also called the true positive fraction. In symbols, we write

$$\beta = P_{\theta_1}\{\text{fail to reject } H_0\}.$$
Terminology

The action is often based on first determining a critical region \( C \). Data \( x \) in this region is determined to be too unlikely to have occurred when the null hypothesis is true. Thus,

\[
\text{reject } H_0 \quad \text{if and only if} \quad x \in C.
\]

Given a choice \( \alpha \) for the size of the test, the choice of a critical region \( C \) is called best or most powerful if for any other critical region \( C^* \) for a size \( \alpha \) test, i.e., both critical region lead to the same type I error probability,

\[
\alpha = P_{\theta_0} \{ X \in C \} = P_{\theta_0} \{ X \in C^* \},
\]

but perhaps different type II error probabilities

\[
\beta = P_{\theta_1} \{ X \notin C \}, \quad \beta^* = P_{\theta_1} \{ X \notin C^* \},
\]

the lowest probability of a type II error, \( (\beta \leq \beta^*) \) is associated to the critical region \( C \).
The Neyman-Pearson Lemma

Consider two likelihoods for $x$ running from $-11$ to $11$,

<table>
<thead>
<tr>
<th>$x$</th>
<th>-11</th>
<th>-10</th>
<th>-9</th>
<th>-8</th>
<th>-7</th>
<th>-6</th>
<th>-5</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_0(x)$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>$L_1(x)$</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_0(x)$</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$L_1(x)$</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

The goal of this game is to pick values $x$ to accumulate points as quickly as possible from your likelihood $L_0$ keeping your opponent’s points from $L_1$ as low as possible.

<table>
<thead>
<tr>
<th>$x$</th>
<th>-2</th>
<th>3</th>
<th>-6</th>
<th>-9</th>
<th>0</th>
<th>1</th>
<th>-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_0$ total</td>
<td>8</td>
<td>15</td>
<td>19</td>
<td>20</td>
<td>30</td>
<td>39</td>
<td>46</td>
</tr>
<tr>
<td>$L_1$ total</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>
The Neyman-Pearson Lemma

Keeping track of the size $\alpha$ and the power $1 - \beta$ of the test with the choice of critical region being the values of $x$ not yet chosen, we have the following table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>-2</th>
<th>3</th>
<th>-6</th>
<th>-9</th>
<th>0</th>
<th>1</th>
<th>-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_0(x)/L_1(x)$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>10</td>
<td>9</td>
<td>7/4</td>
</tr>
<tr>
<td>$L_0$ total</td>
<td>8</td>
<td>15</td>
<td>19</td>
<td>20</td>
<td>30</td>
<td>39</td>
<td>46</td>
</tr>
<tr>
<td>$L_1$ total</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.92</td>
<td>0.85</td>
<td>0.81</td>
<td>0.80</td>
<td>0.70</td>
<td>0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>$1 - \beta$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.94</td>
</tr>
</tbody>
</table>

We see how the likelihood ratio test is the most powerful test. For example, for these likelihoods, the last column states that for a $\alpha = 0.54$ level test, the best region consists of those values of $x$ so that

$$\frac{L_1(x)}{L_0(x)} \geq \frac{7}{4}.$$ 

The power is $1 - \beta = 0.94$ and thus the type II error probability is $\beta = 0.06$. 

The Neyman-Pearson Lemma

Exercise. Repeat exercise above. The R code follows.

```r
> x<-c(-11:11)
> L0<-c(0,0:10,9:0,0)
> L1<-sample(L0,length(L0))
> data.frame(x,L0,L1)
> o<-order(L1/L0)
> sumL0<-cumsum(L0[o])
> sumL1<-cumsum(L1[o])
> alpha<-1-sumL0/100
> beta<-sumL1/100
> data.frame(x[o],L1[o]/L1[o],L0[o],L1[o],alpha,beta)
> plot(alpha,1-beta,type="s")
```

The graph $\alpha$ versus $1 - \beta$ is called the receiver operator characteristic (ROC) curve.
The Neyman-Pearson Lemma

Theorem. (Neyman-Pearson Lemma) Let $L(\theta|x)$ denote the likelihood function for the random variable $X$ corresponding to the probability $P_\theta$. If there exists a critical region $C$ of size $\alpha$ and a nonnegative constant $k_\alpha$ such that

$$\frac{L(\theta_1|x)}{L(\theta_0|x)} \geq k_\alpha \quad \text{for} \ x \in C$$

and

$$\frac{L(\theta_1|x)}{L(\theta_0|x)} < k_\alpha \quad \text{for} \ x \notin C,$$

then $C$ is the most powerful critical region of size $\alpha$. 