Topic 17

Simple Hypotheses

Terminology and the Neyman-Pearson Lemma

Outline

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 prosecutors task is to present the evidence in the most powerful and convincing manner possible.

Terminology

The simplest set-up for understanding the issues of statistical hypothesis, is the case of two values θ_0 and θ_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 eject the hypothesis.

Terminology

criminal trials		
	innocent	guilty
convict		OK
do not convict	OK	

hypothesis tests		
	H_0 is true	H_1 is true
reject H_0	type I error	OK
fail to reject H_0	OK	type II error

Terminology

- Rejecting the hypothesis when it is true is called a type I error or a false positive. Its probability α 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 β. 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 $\mathbf{x}$ in this region is determined to be too unlikely to have occurred when the null hypothesis is true. Thus,

$$\text{reject } H_0 \quad \text{if and only if} \quad \mathbf{x} \in C.$$

Given a choice α 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 α 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 probabiities

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

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,

x	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
$L_0(x)$	0	0	1	2	3	4	5	6	7	8	9	10
$L_1(x)$	4	2	0	3	7	0	6	8	4	0	9	1

x	1	2	3	4	5	6	7	8	9	10	11
$L_0(x)$	9	8	7	6	5	4	3	2	1	0	0
$L_1(x)$	1	9	0	8	5	5	2	10	6	3	7

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.

x	-2	3	-6	-9	0	1	-3
L_0 total	8	15	19	20	30	39	46
L_1 total	0	0	0	0	1	2	6

The Neyman-Pearson Lemma

Keeping track of the size α 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.

x	-2	3	-6	-9	0	1	-3
$L_0(x)/L_1(x)$	∞	∞	∞	∞	10	9	7/4
L_0 total	8	15	19	20	30	39	46
L_1 total	0	0	0	0	1	2	6
α	0.92	0.85	0.81	0.80	0.70	0.61	0.54
$1-\beta$	1.00	1.00	1.00	1.00	0.99	0.98	0.94

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

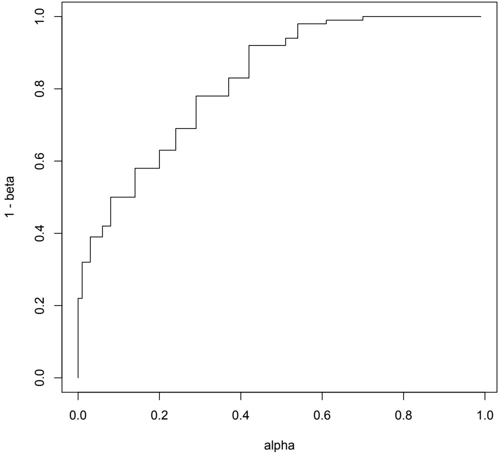
$$\frac{L_1(x)}{L_0(x)} \geq \frac{7}{4}.$$

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.

```
> 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 α versus $1-\beta$ is called the receiver operator characteristic (ROC) curve.

The Neyman-Pearson Lemma

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

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

and

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

then C is the most powerful critical region of size α.