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 reject 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
$L_1(x)$	8	0	0	1	1	7	6	9	5	3	4	4

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
$L_1(x)$	8	2	5	7	0	9	6	2	10	3	0

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	-9	5	2	-7	-2	0	-1
L_0 total	1	6	14	17	25	35	44
L_1 total	0	0	2	3	6	10	14

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	-9	5	2	-7	-2	0	-1
$L_0(x)/L_1(x)$	∞	∞	4	3	8/3	5/2	9/4
L_0 total	1	6	14	17	25	35	44
L_1 total	0	0	2	3	6	10	14
α	0.99	0.94	0.86	0.83	0.75	0.65	0.56
$1-\beta$	1.00	1.00	0.98	0.97	0.94	0.90	0.86

We see how the likelihood ratio test is the most powerful test. For example, for these likelihoods, the last column states that for a $\alpha = 0.56$ level test, the best region consists of those values of x so that

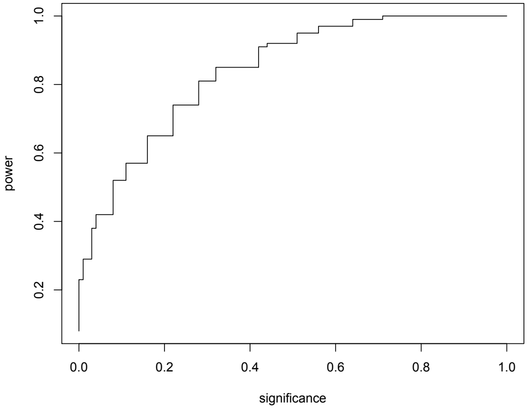
$$\frac{L_1(x)}{L_0(x)} \geq \frac{9}{4}.$$

The power is $1-\beta = 0.86$ and thus the type II error probability is $\beta = 0.14$.

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)
> 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],L0[o]/L1[o],L0[o],
  L1[o],alpha,beta)
> plot(alpha,1-beta,type="s")
```



The graph α versus $1-\beta$ is called the receiving 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.

Proof of the Neyman-Pearson Lemma

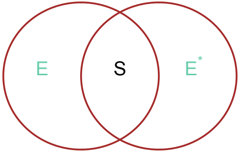
Let C be the α critical region determined by the likelihood ratio test. In addition, let C^* be a critical region for a second test of size α. In symbols,

$$P_{\theta_0}\{X \in C^*\} = P_{\theta_0}\{X \in C\} = \alpha.$$

As before, we use the symbols β and β^* denote, respectively, the probability of type II error for the critical regions C and C^* respectively.

The Neyman-Pearson lemma is the statement that $\beta^* \geq \beta$.

Proof of the The Neyman-Pearson Lemma



Divide both critical regions C and C^* into two disjoint subsets, the subset that the critical regions share $S = C \cap C^*$ and the subsets $E = C \backslash C^*$ and $E^* = C^* \backslash C$ that are exclusive to one region.

In symbols, we write this as the disjoint unions

$$C = S \cup E, \quad \text{and} \quad C^* = S \cup E^*.$$

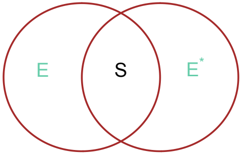
Thus under either parameter value $\theta_i, i = 1, 2$,

$$P_{\theta_i}\{X \in C\} = P_{\theta_i}\{X \in S\} + P_{\theta_i}\{X \in E\}$$

and

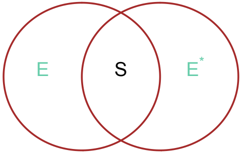
$$P_{\theta_i}\{X \in C^*\} = P_{\theta_i}\{X \in S\} + P_{\theta_i}\{X \in E^*\}.$$

Proof of the The Neyman-Pearson Lemma



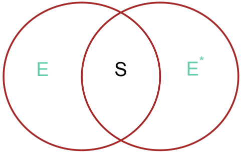
- The contribution to type I errors from data in S and type II errors from data outside $E \cup E^*$ are the same for both tests. Thus, we can focus on differences in types of error by examining the case in which the data land in either E and E^*.
- Because both test have level α, the probability that the data land in E or in E^* are the same under the null hypothesis.
- Under the likelihood ratio critical region, the null hypothesis is not rejected in E^*.
- Under the second test critical region, the null hypothesis is not rejected in E.

Proof of the The Neyman-Pearson Lemma



- E^* is outside likelihood ratio critical region. Under the alternative hypothesis, the probability that the data land in E^* is *at most* k_α times as large as it is under the null. This contributes to the type II error for the likelihood ratio based test.
- E is inside the likelihood ratio critical region. So, under the alternative hypothesis, the probability that the data land in E is *at least* k_α times as large as it is under the null. This contributes a larger amount to the type II error for the second test than is added from E^* to the likelihood ratio based test.
- Thus, the type II error for the likelihood ratio based test is *smaller* than the type II error for the second test.

Proof of the The Neyman-Pearson Lemma



Consider the parameter value θ_0. Start with

$$\alpha = P_{\theta_0}\{X \in C\} = P_{\theta_0}\{X \in C^*\}$$

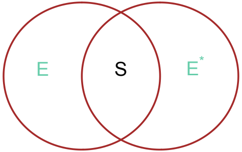
and subtract the probability that the data land in the shared critical regions $P_{\theta_0}\{X \in S\}$, i.e., the region rejected by both tests to obtain

$$P_{\theta_0}\{X \in E^*\} = P_{\theta_0}\{X \in E\}$$

or

$$P_{\theta_0}\{X \in E^*\} - P_{\theta_0}\{X \in E\} = 0$$

Proof of the The Neyman-Pearson Lemma



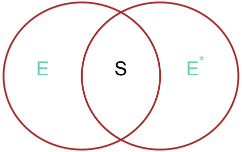
For θ_1, the difference in the corresponding type II error probabilities is

$$\begin{aligned}\beta^* - \beta &= P_{\theta_1}\{X \notin C^*\} - P_{\theta_1}\{X \notin C\} \\ &= (1 - P_{\theta_1}\{X \in C^*\}) - (1 - P_{\theta_1}\{X \in C\}) = P_{\theta_1}\{X \in C\} - P_{\theta_1}\{X \in C^*\}.\end{aligned}$$

Now subtract from both probabilities the quantity $P_{\theta_1}\{X \in S\}$, the probability that the hypothesis would be *falsely* rejected by both tests to obtain

$$\beta^* - \beta = P_{\theta_1}\{X \in E\} - P_{\theta_1}\{X \in E^*\}.$$

Proof of the The Neyman-Pearson Lemma



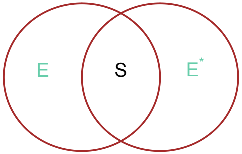
- For $\mathbf{x} \in E$, then $\mathbf{x}$ is *inside* the critical region and consequently $L(\theta_1|\mathbf{x}) \geq k_\alpha L(\theta_0|\mathbf{x})$ and

$$P_{\theta_1}\{X \in E\} = \int_E L(\theta_1|\mathbf{x})\ \nu(d\mathbf{x}) \geq k_\alpha \int_E L(\theta_0|\mathbf{x})\ \nu(d\mathbf{x}) = k_\alpha P_{\theta_0}\{X \in E\}.$$

- For $\mathbf{x} \in E^*$, then $\mathbf{x}$ is *outside* the critical region and consequently $L(\theta_1|\mathbf{x}) \leq k_\alpha L(\theta_0|\mathbf{x})$ and

$$P_{\theta_1}\{X \in E^*\} = \int_{E*} L(\theta_1|\mathbf{x})\ \nu(d\mathbf{x}) \leq k_\alpha \int_{E*} L(\theta_0|\mathbf{x})\ \nu(d\mathbf{x}) = k_\alpha P_{\theta_0}\{X \in E^*\}.$$

Proof of the The Neyman-Pearson Lemma



Thus,

$$P_{\theta_1}\{X \in E\} \geq k_\alpha P_{\theta_0}\{X \in E\} \quad \text{and} \quad -P_{\theta_1}\{X \in E^*\} \geq -k_\alpha P_{\theta_0}\{X \in E^*\}$$

Add these two inequalities to obtain

$$\beta^* - \beta = P_{\theta_1}\{X \in E\} - P_{\theta_1}\{X \in E^*\} \geq k_\alpha(P_{\theta_0}\{X \in E\} - P_{\theta_0}\{X \in E^*\}).$$

This difference is at least 0 and consequently $\beta^* \geq \beta$, i. e., the critical region C^* has at least as large type II error probability as that given by the likelihood ratio test. QED

The Neyman-Pearson Lemma

I can point to the particular moment when I understood how to formulate the undogmatic problem of the most powerful test of a simple statistical hypothesis against a fixed simple alternative. At the present time, the problem appears entirely trivial and within reach of a beginning undergraduate. But, with a degree of embarrassment, I must confess that it took something like half a decade of combined effort of E.S.P. and myself to put things straight.

- Jerzy Neymann in the Festschrift in honor of Herman Wold, 1970,
E.S.P is Egon Sharpe Pearson