

# Topic 17

## Simple Hypotheses

*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*

### 17.1 Overview and Terminology

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.**

The weight of evidence that is necessary to find the defendant guilty depends on the type of trial. In a criminal trial the stated standard is that the prosecution must prove that *the defendant is guilty beyond any reasonable doubt*. In civil trials, the burden of proof may be the intermediate level of *clear and convincing evidence* or the lower level of the *preponderance of evidence*.

Given the level of evidence needed, a prosecutors task is to present the evidence in the most powerful and convincing manner possible. We shall see these notions reflected in the nature of hypothesis testing.

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**.

We now frame the issue of hypothesis testing using the classical approach. In this approach, the possible actions are:

- **Reject the hypothesis.** 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

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

- **Fail to reject the hypothesis.** 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

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

hypothesis tests			criminal trials		
	$H_0$ is true	$H_1$ is true		the defendant is	
				innocent	guilty
reject $H_0$	type I error	OK	convict		OK
fail to reject $H_0$	OK	type II error	do not convict	OK	

Thus, the *higher* level necessary to secure conviction in a criminal trial corresponds to having *lower* significance levels. This analogy should not be taken too far. The nature of the data and the decision making process is quite dissimilar. For example, the prosecutor and the defense attorney are not always out to find the most honest manner to present information. In statistical inference for hypothesis testing, the goal is something that all participants in this endeavor ought to share.

The decision for the test 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, the decision is

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

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 choice of 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

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

we have the lowest probability of a type II error, ( $\beta \leq \beta^*$ ) associated to the critical region  $C$ .

The two approaches to hypothesis testing, classical and Bayesian, begin with distinct starting points and end with different interpretations for implications of the data. Interestingly, both approaches result in a decision that is based on the values of a likelihood ratio. In the classical approach, we shall learn, based on the Neyman-Pearson lemma, that the decision is based on a level for this ratio based on setting the type I error probabilities. In the Bayesian approach, the decision on minimizing risk, a concept that we will soon define precisely.

## 17.2 The Neyman-Pearson Lemma

Many critical regions are either determined by the consequences of the **Neyman-Pearson lemma** or by using analogies of this fundamental lemma. Rather than presenting a proof of this lemma, we will provide some intuition for the choice of critical region through the following “game”.

We will conduct a single observation  $X$  that can take values from  $-11$  to  $11$  and based on that observation, decide whether or not to reject the null hypothesis. Basing a decision on a single observation, of course, is not the usual circumstance for hypothesis testing. We will first continue on this line of reasoning to articulate the logic behind the Neyman-Pearson lemma before examining more typical and reasonable data collection protocols.

To begin the game, corresponding to values for  $x$  running from  $-11$  to  $11$ , write a row of the number from 0 up to 10 and back down to 0 and add an additional 0 at each end. These numbers add to give 100. Now, scramble the numbers and write them under the first row. This can be created and displayed quickly in R using the commands:

```
> x<-c(-11:11)
> L0<-c(0,0:10,9:0,0)
> L1<-sample(L0,length(L0))
> data.frame(x,L0,L1)
```

The top row, giving the values of  $L_0$ , represents the likelihood for our one observation under the null hypothesis. The bottom row, giving the values of  $L_1$ , represents the likelihood under the alternative hypothesis. Note that the values for  $L_1$  is a rearrangement of the values for  $L_0$ . Here is the output from one simulation.

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

The goal of this game is to pick values  $x$  so that your accumulated points increase as quickly as possible from your likelihood  $L_0$  keeping your opponent's points from  $L_1$  as low as possible. The natural start is to pick values of  $x$  so that  $L_1(x) = 0$ . Then, the points you collect begin to add up without your opponent gaining anything. We find 4 such values for  $x$  and record their values along with running totals for  $L_0$  and  $L_1$ . 0

$x$	-2	3	5	7
$L_0$ total	8	15	20	23
$L_1$ total	0	0	0	0

Being ahead by a score of 23 to 0 can be translated into a best critical region in the following way. If we take as our critical region  $C$  all the values for  $x$  except -2, 3, 5, and 7. Then, because the  $L_0$ -total is 23 points out of a possible 100, then we find

$$\text{the size of the test } \alpha = 0.77 \quad \text{and} \quad \text{the power of the test } 1 - \beta = 1.00$$

because there is *no* chance of type II error with this critical region. If the results of our one observation is one of -2, 3, 5, or 7, then we are never incorrect in failing to reject  $H_0$ .

Understanding the next choice is crucial. Candidates are

$$x = 4, \text{ with } L_0(4) = 6 \text{ against } L_1(4) = 1 \quad \text{and} \quad x = 1, \text{ with } L_0(1) = 9 \text{ against } L_1(1) = 2.$$

The choice 6 against 1 is better than 9 against 2. One way to see this is to note that choosing 6 against 1 twice will put us in a better place than the single choice of 9 against 2. Indeed, after choosing 6 against 1, a choice of 3 against 1 puts us in at least as good a position than the single choice of 9 against 2. The central point is that the best choice comes to picking the remaining value for  $x$  that has the *highest ratio* of  $L_0(x)$  to  $L_1(x)$

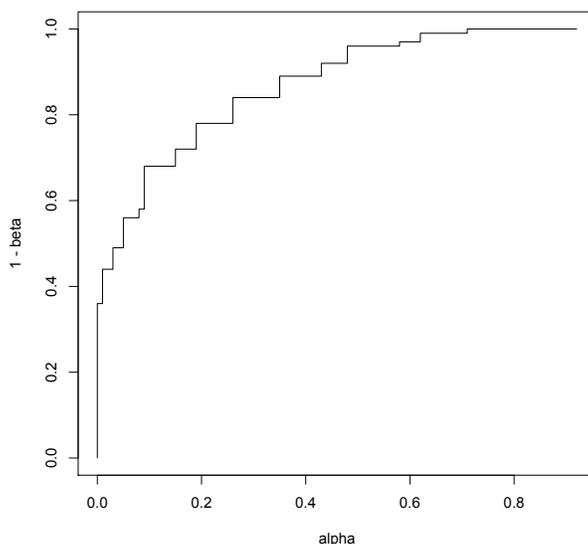
Now we can pick the next few candidates, keeping track of the size and the power of the test with the choice of critical region being the values of  $x$  not yet chosen.

$x$	-2	3	5	7	4	1	-6	0	-5
$L_0(x)/L_1(x)$	$\infty$	$\infty$	$\infty$	$\infty$	6	9/2	4	5/2	5/3
$L_0$ total	8	15	20	23	29	38	42	52	57
$L_1$ total	0	0	0	0	1	3	4	8	11
$\alpha$	0.92	0.85	0.80	<b>0.77</b>	0.71	0.62	0.58	0.48	0.43
$1 - \beta$	1.00	1.00	1.00	<b>1.00</b>	0.99	0.97	0.96	0.92	0.89

From this exercise we see how the likelihood ratio test is the choice for a most powerful test. For example, for these likelihoods, the last column states that for a  $\alpha = 0.43$  level test, the best region consists of those values of  $x$  so that

$$\frac{L_1(x)}{L_0(x)} \geq \frac{5}{3}.$$

The power is  $1 - \beta = 0.89$  and thus the type II error probability is  $\beta = 0.11$ . In genuine examples, we will typically look for level much below 0.43 and we will make many observations. We now summarize carefully the insights from this game before examining more genuine examples. A proof of this theorem is provided in Section 17.4.



**Figure 17.1: Receiver Operator Characteristic.** The graph of  $\alpha = P\{X \in C | H_0 \text{ is true}\}$  (significance) versus  $1 - \beta = P\{X \in C | H_1 \text{ is true}\}$  (power) in the example. The horizontal axis  $\alpha$  is also called the **false positive fraction (FPF)**. The vertical axis  $1 - \beta$  is also called the **true positive fraction (TPF)**.

**Theorem 17.1** (Neyman-Pearson Lemma). *Let  $L(\theta|\mathbf{x})$  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*

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

and

$$\frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} < k_\alpha \quad \text{for } \mathbf{x} \notin C, \quad (17.1)$$

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

We, thus, reject the null hypothesis if and only if the likelihood ratio exceeds a value  $k_\alpha$  with

$$\alpha = P_{\theta_0} \left\{ \frac{L(\theta_1|X)}{L(\theta_0|X)} \geq k_\alpha \right\}.$$

We shall learn that many of the standard tests use critical values for the  $t$ -statistic, the chi-square statistic, or the  $F$ -statistic. These critical values are related to the critical value  $k_\alpha$  in extensions of the ideas of likelihood ratios. In a few pages, we will take a glance at the Bayesian approach to hypothesis testing.

### 17.2.1 The Receiver Operator Characteristic

Using `R`, we can complete the table for  $L_0$  total and  $L_1$  total.

```
> 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], sumL0, sumL1, alpha, 1-beta)
```

Completing the curve, known as the **receiver operator characteristic (ROC)**, is shown in the figure above. The ROC shows the inevitable trade-offs between Type I and Type II errors. For example, by the mere fact that the graph is increasing, we can see that by setting a more rigorous test achieved by lowering  $\alpha$ , the level of significance, (decreasing the value on the horizontal axis) necessarily reduces  $1 - \beta$ , the power (decreasing the value on the vertical axis.). The unusual and slightly mystifying name is due to the fact that the ROC was first developed during World War II for detecting enemy objects in battlefields, Following the surprise military attack on Pearl Harbor in 1941, the United States saw the need to improve the prediction of the movement of aircraft from their radar signals.

**Exercise 17.2.** Consider the following (ignorant) example. Flip a coin that gives heads with probability  $\alpha$ . Ignore whatever data you have collected and reject if the coin turns up heads. This test has significance level  $\alpha$ . Describe the receiver operator characteristic curve in the case.

This shows what a minimum acceptable ROC curve looks like - any hypothesis test ought be better than a coin toss that ignores the data.

### 17.3 Examples

**Example 17.3.** Mimicry is the similarity of one species to another in a manner that enhances the survivability of one or both species - the **model** and **mimic**. This similarity can be, for example, in appearance, behavior, sound, or scent. One method for producing a mimic species is **hybridization**. This results in the transferring of adaptations from the model species to the mimic. The genetic signature of this has recently been discovered in Heliconius butterflies. Padro-Diaz et al sequenced chromosomal regions both linked and unlinked to the red colour locus and found a region that displays an almost perfect genotype by phenotype association across four species in the genus Heliconius



Figure 17.2: Heliconius butterflies

Let's consider a model butterfly species with mean wingspan  $\mu_0 = 10$  cm and a mimic species with mean wingspan  $\mu_1 = 7$  cm. Both species have standard deviation  $\sigma_0 = 3$  cm. Collect 16 specimen to decide if the mimic species has migrated into a given region. If we assume, for the null hypothesis, that the habitat under study is populated by the model species, then

- a type I error is falsely concluding that the species is the mimic when indeed the model species is resident and
- a type II error is falsely concluding that the species is the model when indeed the mimic species has invaded.

If our action is to begin an eradication program if the mimic has invaded, then a type I error would result in the eradication of the resident model species and a type II error would result in the letting the invasion by the mimic take its course.

To begin, we set a significance level. The choice of an  $\alpha = 0.05$  test means that we are accepting a 5% chance of having this error. If the goal is to design a test that has the lowest type II error probability, then the Neyman-Pearson lemma tells us that the critical region is determined by a threshold level  $k_\alpha$  for the likelihood ratio.

$$C = \left\{ \mathbf{x}; \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} \geq k_\alpha \right\}.$$

We next move to see how this critical region is determined.

**Example 17.4.** Let  $X = (X_1, \dots, X_n)$  be independent normal observations with unknown mean and known variance  $\sigma_0^2$ . The hypothesis is

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu = \mu_1.$$

For the moment consider the case in which  $\mu_1 < \mu_0$ . We look to determine the critical region.

$$\begin{aligned} \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} &= \frac{\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_0^2}\right) \cdots \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_n-\mu_1)^2}{2\sigma_0^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_1-\mu_0)^2}{2\sigma_0^2}\right) \cdots \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_n-\mu_0)^2}{2\sigma_0^2}\right)} \\ &= \frac{\exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu_1)^2\right)}{\exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu_0)^2\right)} \\ &= \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n ((x_i - \mu_1)^2 - (x_i - \mu_0)^2)\right) \\ &= \exp\left(-\frac{\mu_0 - \mu_1}{2\sigma_0^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0)\right) \end{aligned}$$

Because the exponential function is increasing, the likelihood ratio test (17.1) is equivalent to

$$\frac{\mu_1 - \mu_0}{2\sigma_0^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0), \quad (17.2)$$

exceeding some critical value. Continuing to simplify, this is equivalent to  $\bar{x}$  bounded by some critical value,

$$\bar{x} \leq \tilde{k}_\alpha,$$

where  $\tilde{k}_\alpha$  is chosen to satisfy

$$P_{\mu_0}\{\bar{X} \leq \tilde{k}_\alpha\} = \alpha.$$

(Note that division by the negative number  $\mu_1 - \mu_0$  reverses the direction of the inequality.) Pay particular attention to the fact that the probability is computed under the null hypothesis specifying the mean to be  $\mu_0$ . In this case,  $\bar{X}$  is  $N(\mu_0, \sigma_0/\sqrt{n})$  and consequently the standardized version of  $\bar{X}$ ,

$$Z = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}}, \quad (17.3)$$

is a standard normal. Set  $z_\alpha$  so that  $P\{Z \leq -z_\alpha\} = \alpha$ . (This can be determined in **R** using the `qnorm` command.) Then, by rearranging (17.3), we can determine  $\tilde{k}_\alpha$ .

$$\bar{X} \leq \mu_0 - z_\alpha \frac{\sigma_0}{\sqrt{n}} = \tilde{k}_\alpha.$$

Equivalently, we can use the standardized score  $Z$  as our test statistic and  $z_\alpha$  as the critical value. Note that the only role played by  $\mu_1$ , the value of the mean under the alternative, is that is less than  $\mu_0$ . However, it will play a role in determining the power of the test.

**Exercise 17.5.** In the example above, give the value of  $\tilde{k}_\alpha$  explicitly in terms of  $k_\alpha$ ,  $\mu_0$ ,  $\mu_1$ ,  $\sigma_0^2$  and  $n$ .

Returning to the example of the model and mimic bird species, we now see, by the Neyman-Person lemma that the critical region can be defined as

$$C = \left\{ \mathbf{x}; \bar{x} \leq \tilde{k}_\alpha \right\} = \left\{ \mathbf{x}; \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq -z_\alpha \right\}.$$

Under the null hypothesis,  $\bar{X}$  has a normal distribution with mean  $\mu_0 = 10$  and standard deviation  $\sigma/\sqrt{n} = 3/4$ . This using the distribution function of the normal we can find either  $\tilde{k}_\alpha$

```
> qnorm(0.05, 10, 3/4)
[1] 8.76636
```

or  $-z_\alpha$ ,

```
> qnorm(0.05)
[1] -1.644854
```

Thus, the critical value is  $\tilde{k}_\alpha = 8.767$  for the test statistic  $\bar{x}$  and  $z_\alpha = -1.645$  for the test statistic  $z$ . Now let's look at data.

```
> x
[1] 8.9 2.4 12.1 10.0 9.2 3.7 13.9 9.1 8.8 6.3 12.1 11.0 12.5 4.5 8.2 10.2
> mean(x)
[1] 8.93125
```

Then

$$\bar{x} = 8.931 \quad z = \frac{8.93124 - 10}{3/\sqrt{16}} = -1.425.$$

$\tilde{k}_\alpha = 8.766 < 8.931$  or  $z_\alpha = -1.645 < -1.425$  and we fail to reject the null hypothesis.

**Exercise 17.6.** Modify the calculations in the example above to show that for the case  $\mu_0 < \mu_1$ , using the same value of  $z_\alpha$  as above, we reject the null hypothesis precisely when

$$\bar{X} \geq \mu_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}} \quad \text{or} \quad Z \geq z_\alpha$$

**Exercise 17.7.** Give an intuitive explanation why the power should

- increase as a function of  $|\mu_1 - \mu_0|$ ,
- decrease as a function of  $\sigma_0^2$ , and
- increase as a function of  $n$ .

Next we determine the type II error probability for the situation given by the previous exercise. We will be guided by the fact that

$$\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}}$$

is a *standard normal random variable* in the case that  $H_1 : \mu = \mu_1$  is true.

For  $\mu_1 > \mu_0$ , we find that the type II error probability

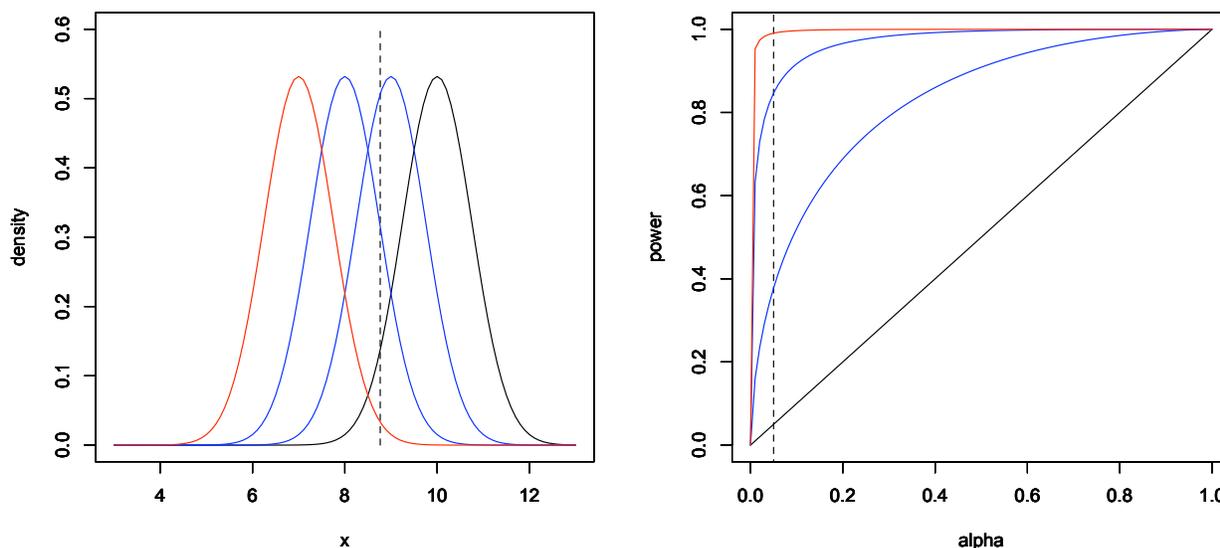
$$\begin{aligned} \beta &= P_{\mu_1}\{X \notin C\} = P_{\mu_1}\{\bar{X} < \mu_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}}\} \\ &= P_{\mu_1}\left\{\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}} < z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}}\right\} = \Phi\left(z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}}\right) \end{aligned}$$

and the power

$$1 - \beta = 1 - \Phi\left(z_\alpha + \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}}\right) \quad (17.4)$$

**Exercise 17.8.** Modify the calculations of power in (17.4) above to show that for the case  $\mu_1 < \mu_0$  to show that

$$1 - \beta = \Phi\left(-z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}}\right). \quad (17.5)$$



**Figure 17.3:** **Left:** (black) Density of  $\bar{X}$  for normal data under the null hypothesis -  $\mu_0 = 10$  and  $\sigma_0/\sqrt{n} = 3/\sqrt{16} = 3/4$ . With an  $\alpha = 0.05$  level test, the critical value  $k_\alpha = \mu_0 - z_\alpha\sigma_0/\sqrt{n} = 8.766$ . The area to the left of the dashed line and below the density function is  $\alpha$ . The alternatives shown are  $\mu_1 = 9$  and 8 (in blue) and  $\mu_1 = 7$  (in red). The areas below the curves and to the left of the dashed line is the power  $1 - \beta$ . These values are 0.3777, 0.8466, and 0.9907 for respective alternatives  $\mu_1 = 9, 8$  and 7. **Right:** The corresponding receiver operator characteristics curves of the power  $1 - \beta$  versus the significance  $\alpha$  using equation (17.5). The power for an  $\alpha = 0.05$  test are indicated by the intersection of vertical dashed line and the receiver operator characteristics curves.

A type II error is falsely failing to conclude that the mimic species have inhabited the study area when indeed they have. To compute the probability of a type II error, note that for  $\alpha = 0.05$ , we substitute into (17.5),

$$-z_\alpha + \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} = -1.645 + \frac{3}{3/\sqrt{16}} = 2.355$$

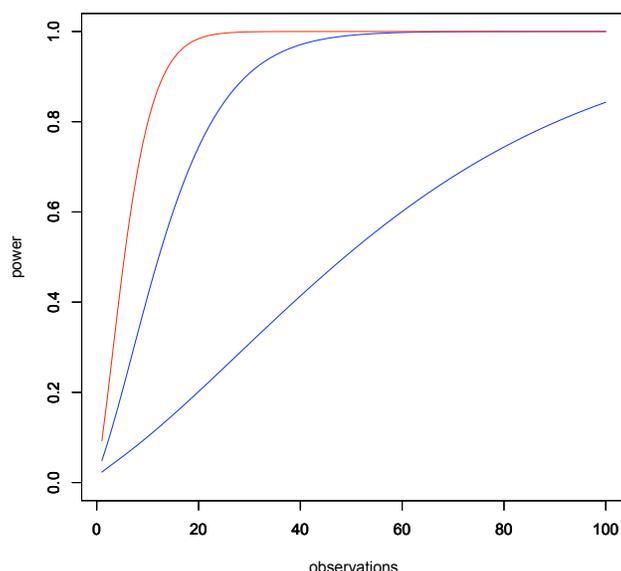
```
> pnorm(2.355)
[1] 0.9907386
```

and the type II error probability is  $\beta = 1 - 0.9907 = 0.0093$ , a bit under 1%.

Let's expand the examination of equation (17.5). As we move the alternative value  $\mu_1$  downward, the density of  $\bar{X}$  moves leftward. The values for  $\mu_1 = 9, 8$ , and 7 are displayed on the left in Figure 17.2. This shift in the values is a way of saying that the alternative is becoming more and more distinct as  $\mu_1$  decreases. The mimic species becomes easier and easier to detect. We express this by showing that the test is more and more powerful with decreasing values of  $\mu_1$ . This is displayed by the increasing area under the density curve to the left of the dashed line from 0.377 for the alternative  $\mu_1 = 9$  to 0.9907 for  $\mu_1 = 7$ . We can also see this relationship in the receiver operator characteristic graphed, the graph of the power  $1 - \beta$  versus the significance  $\alpha$ . This is displayed for the significance level  $\alpha = 0.05$  by the dashed line.

**Exercise 17.9.** Determine the power of the test for  $\mu_0 = 10$  cm and  $\mu_1 = 9, 8$ , and 7 cm with the significance level  $\alpha = 0.01$ . Does the power increase or decrease from its value when  $\alpha = 0.01$ ? Explain your answer. How would the graphs in Figure 17.2 be altered to show this case?

Often, we wish to know in advance the number of observations  $n$  needed to obtain a given power. In this case, we use (17.4) with a fixed value of  $\alpha$ , the size of the test, and determine the power of the test as a function of  $n$ . We display this in Figure 17.3 with the value of  $\alpha = 0.01$ . Notice how the number of observations needed to achieve a desired power is high when the wingspan of the mimic species is close to that of the model species.



**Figure 17.4:** Power as a function of the number of observations for an  $\alpha = 0.01$  level test. The null hypothesis -  $\mu_0 = 10$ . The alternatives shown are  $\mu_1 = 9$  and  $8$  (in blue) and  $\mu_1 = 7$  (in red). Here  $\sigma_0 = 3$ . The low level for  $\alpha$  is chosen to reflect the desire to have a stringent criterion for rejecting the null hypothesis that the resident species is the model species.

The example above is called the  $z$ -test. If  $n$  is sufficiently large, then even if the data are not normally distributed,  $\bar{X}$  is well approximated by a normal distribution and, as long as the variance  $\sigma_0^2$  is known, the  $z$ -test is used in this case. In addition, the  $z$ -test can be used when  $g(\bar{X}_1, \dots, \bar{X}_n)$  can be approximated by a normal distribution using the delta method.

**Example 17.10** (Bernoulli trials). Here  $X = (X_1, \dots, X_n)$  is a sequence of Bernoulli trials with unknown success probability  $p$ , the likelihood

$$\begin{aligned} L(p|\mathbf{x}) &= p^{x_1}(1-p)^{1-x_1} \dots p^{x_n}(1-p)^{1-x_n} = p^{x_1+\dots+x_n}(1-p)^{n-(x_1+\dots+x_n)} \\ &= (1-p)^n \left( \frac{p}{1-p} \right)^{x_1+\dots+x_n} \end{aligned}$$

For the test

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p = p_1$$

the likelihood ratio

$$\frac{L(p_1|\mathbf{x})}{L(p_0|\mathbf{x})} = \left( \frac{1-p_1}{1-p_0} \right)^n \left( \left( \frac{p_1}{1-p_1} \right) / \left( \frac{p_0}{1-p_0} \right) \right)^{x_1+\dots+x_n}. \quad (17.6)$$

**Exercise 17.11.** Show that the likelihood ratio (17.6) results in a test to reject  $H_0$  whenever

$$\sum_{i=1}^n x_i \geq \tilde{k}_\alpha \quad \text{when } p_0 < p_1 \quad \text{or} \quad \sum_{i=1}^n x_i \leq \tilde{k}_\alpha \quad \text{when } p_0 > p_1. \quad (17.7)$$

In words, if the alternative is a higher proportion than the null hypothesis, we reject  $H_0$  when the data have too many successes. If the alternative is lower than the null, we reject  $H_0$  when the data do not have enough successes.

In either situation, the number of successes  $N = \sum_{i=1}^n X_i$  has a  $\text{Bin}(n, p_0)$  distribution under the null hypothesis. Thus, in the case  $p_0 < p_1$ , we choose  $\tilde{k}_\alpha$  so that

$$P_{p_0} \left\{ \sum_{i=1}^n X_i \geq \tilde{k}_\alpha \right\} \leq \alpha. \quad (17.8)$$

In general, we cannot choose  $k_\alpha$  to obtain exactly the value  $\alpha$ . Thus, we take the minimum value of  $k_\alpha$  to achieve the inequality in (17.8).

To give a concrete example take  $p_0 = 0.6$  and  $n = 20$  and look at a part of the cumulative distribution function.

$x$	$\cdots$	13	14	15	16	17	18	19	20
$F_N(x) = P\{N \leq x\}$	$\cdots$	0.7500	0.8744	0.9491	0.9840	0.9964	0.9994	0.99996	1

If we take  $\alpha = 0.05$ , then

$$P\{N \geq 16\} = 1 - P\{N \leq 15\} = 1 - 0.9491 = 0.0509 > 0.05$$

$$P\{N \geq 17\} = 1 - P\{N \leq 16\} = 1 - 0.9840 = 0.0160 < 0.05$$

Consequently, we need to have at least 17 successes in order to reject  $H_0$ .

**Exercise 17.12.** Find the critical region in the example above for  $\alpha = 0.10$  and  $\alpha = 0.01$ . For what values of  $\alpha$  is  $C = \{16, 17, 18, 19, 20\}$  a critical region for the likelihood ratio test.

**Example 17.13.** If  $np_0$  and  $n(1 - p_0)$  are sufficiently large, then, by the central limit theorem,  $\sum_{i=1}^n X_i$  has approximately a normal distribution. If we write the sample proportion

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i,$$

then, under the null hypothesis, we can apply the central limit theorem to see that

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

is approximately a standard normal random variable and we perform the  $z$ -test as in the previous exercise.

For example, if we take  $p_0 = 1/2$  and  $p_1 = 3/5$  and  $\alpha = 0.05$ , then with 110 heads in 200 coin tosses

$$Z = \frac{0.55 - 0.50}{0.05/\sqrt{2}} = \sqrt{2}.$$

```
> qnorm(0.95)
```

```
[1] 1.644854
```

Thus,  $\sqrt{2} < 1.645 = z_{0.05}$  and we fail to reject the null hypothesis.

**Example 17.14.** Honey bees store honey for the winter. This honey serves both as nourishment and insulation from the cold. Typically for a given region, the probability of survival of a feral bee hive over the winter is  $p_0 = 0.7$ . We are checking to see if, for a particularly mild winter, this probability moved up to  $p_1 = 0.8$ . This leads us to consider the hypotheses

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p = p_1.$$

for a test of the probability that a feral bee hive survives a winter. If we use the central limit theorem, then, under the null hypothesis,

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

has a distribution approximately that of a standard normal random variable. For an  $\alpha$  level test, the critical value is  $z_\alpha$  where  $\alpha$  is the probability that a standard normal is at least  $z_\alpha$ . If the significance level is  $\alpha = 0.05$ , then we will reject  $H_0$  for any value of  $z > z_\alpha = 1.645$

For this study, 112 colonies have been chosen and 88 survive. Thus  $\hat{p} = 0.7875$  and

$$z = \frac{0.7875 - 0.7}{\sqrt{0.7(1 - 0.7)/112}} = 1.979.$$

Consequently, reject  $H_0$ .

For both of these previous examples, the usual method is to compute the  $z$ -score with the continuity correction. We shall soon see this with the use of `prop.test` in R.

### 17.4 Summary

For a simple hypothesis

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

we have two possible action, **reject**  $H_0$  and **fail to reject**  $H_0$ , this leads to two possible types of errors

error	probability	alternative names
type I	$\alpha = P_{\theta_0} \{\text{reject } H_0\}$	level significance false positive
type II	$\beta = P_{\theta_1} \{\text{fail to reject } H_0\}$	false negative

The probability  $1 - \beta = P_{\theta_1} \{\text{reject } H_0\}$  is called the true positive probability or **power**. The procedure is to set a **significance level**  $\alpha$  and find a critical region  $C$  so that the type II error probability is as small as possible. The Neyman-Pearson lemma lets us know that the critical region is determined by setting a level  $k_\alpha$  for the likelihood ratio.

$$C = \left\{ \mathbf{x}; \frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} \geq k_\alpha \right\}$$

We continue, showing the procedure in the examples above.

	normal observations $\mu_1 \geq \mu_0$	Bernoulli trials $p_1 > p_0$
Simplify likelihood ratio to obtain a test statistic $T(\mathbf{x})$	$\bar{x}$ $z = \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}$	$\sum_{i=1}^n x_i$
Use the distribution of $T(\mathbf{x})$ under $H_0$ to set a critical value $\tilde{k}_\alpha$ so that $P_{\theta_0} \{T(X) \geq \tilde{k}_\alpha\} = \alpha$	$\bar{X} \sim N(\mu_0, \sigma_0/\sqrt{n})$ $Z \sim N(0, 1)$	$\sum_{i=1}^n X_i \sim \text{Bin}(n, p_0)$
Determine type II error probability $\beta = P_{\theta_1} \{T(X) \geq \tilde{k}_\alpha\}$	$P_{\mu_1} \{\bar{X} \geq \tilde{k}_\alpha\}$	$P_{p_1} \{\sum_{i=1}^n X_i \geq \tilde{k}_\alpha\}$

### 17.5 Proof of the Neyman-Pearson Lemma

For completeness in exposition, we include a proof of the Neyman-Pearson lemma.

Let  $C$  be the  $\alpha$  critical region determined by the likelihood ratio test. In addition, let  $C^*$  be a critical region for a second test of size greater than or equal to  $\alpha$ . In symbols,

$$P_{\theta_0} \{X \in C^*\} \geq P_{\theta_0} \{X \in C\} = \alpha \tag{17.9}$$

As before, we use the symbols  $\beta$  and  $\beta^*$  denote, respectively, the probability of type II error for the critical regions  $C$  and  $C^*$  respectively. The Neyman-Pearson lemma is the statement that  $\beta^* \geq \beta$ .

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  $E = C \setminus C^*$  and  $E^* = C^* \setminus C$  that are exclusive to one region. In symbols, we write this as

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

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

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

(See Figure 17.5)

First, we will describe the proof in words.

- The contribution to type I errors from data in  $S$  and for type II errors from data outside  $E \cup E^*$  are the same for both types of test. Consequently, 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 tests have level  $\alpha$ , 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, the null hypothesis is not rejected in  $E$ .
- $E^*$  is outside likelihood ratio critical region. So, under the alternative hypothesis, the probability that the data land in  $E$  is at most  $k_\alpha$  times as large as it is under the null hypothesis. This contributes to the type II error for the likelihood ratio based test.
- $E$  is in the likelihood ratio critical region. So, under the alternative hypothesis, the probability that the data land in  $E$  is at least  $k_\alpha$  times as large as it is under the null hypothesis. 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.

To carry out the proof, first consider the parameter value  $\theta_0$  and subtract from both sides in (17.9) the probability  $P_{\theta_0}\{X \in S\}$  that the data land in the shared critical regions and thus would be rejected by both tests to obtain

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

or

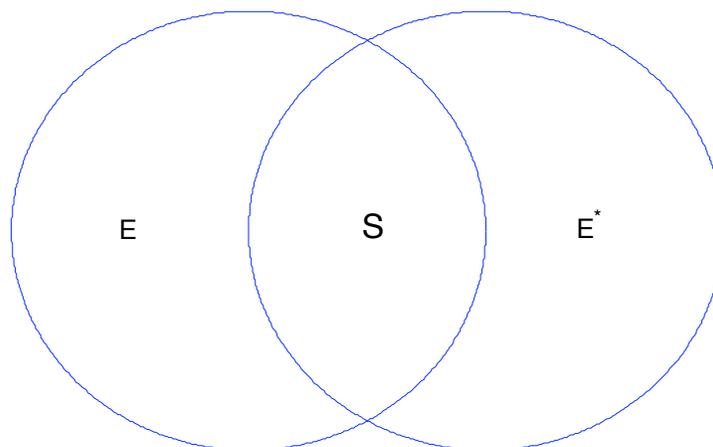
$$P_{\theta_0}\{X \in E^*\} - P_{\theta_0}\{X \in E\} \geq 0. \quad (17.10)$$

Moving to the parameter value  $\theta_1$ , 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 of the integrals the quantity  $P_{\theta_1}\{X \in S\}$ , the probability that the hypothesis would be falsely rejected by both tests to obtain

$$\beta^* - \beta = P_{\theta_1}\{X \in E\} - P_{\theta_1}\{X \in E^*\} \quad (17.11)$$



**Figure 17.5:** Critical region  $C$  as determined by the Neyman-Pearson lemma is indicated by the circle on the left. The circle on the right  $C^*$  is the critical region for a second  $\alpha$  level test. Thus,  $C = S \cup E$  and  $C^* = S \cup E^*$ .

We can use the likelihood ratio criterion on each of the two integrals above.

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

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

- For  $\mathbf{x} \in E^*$ , then  $\mathbf{x}$  is not in the critical region and consequently  $L(\theta_1|\mathbf{x}) \leq k_\alpha L(\theta_0|\mathbf{x})$  and

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

Apply these two inequalities to (17.11)

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

This difference is at least 0 by (17.10) 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.

## 17.6 An Brief Introduction to the Bayesian Approach

As with other aspects of the Bayesian approach to statistics, hypothesis testing is closely aligned with Bayes theorem. For a simple hypothesis, we begin with a **prior probability** for each of the competing hypotheses.

$$\pi\{\theta_0\} = P\{H_0 \text{ is true}\} \quad \text{and} \quad \pi\{\theta_1\} = P\{H_1 \text{ is true}\}.$$

Naturally,  $\pi\{\theta_0\} + \pi\{\theta_1\} = 1$ . Although this is easy to state, the choice of a prior ought to be grounded in solid scientific reasoning.

As before, we collect data and with it compute the **posterior probabilities** of the two parameter values  $\theta_0$  and  $\theta_1$ . This gives us the posterior probabilities that  $H_0$  is true and  $H_1$  is true.

We can see, in its formulation, the wide difference in perspective between the Bayesian and classical approaches.

- In the Bayesian approach, we begin with a prior probability that  $H_0$  is true. In the classical approach, the assumption is that  $H_0$  is true.
- In the Bayesian approach, we use the data and Bayes formula to compute the posterior probability that  $H_1$  is true. In the classical approach, we use the data and a significance level to make a decision to reject  $H_0$ . The question: *What is the probability that  $H_1$  is true?* has no meaning in the classical setting.
- The decision to reject  $H_0$  in the Bayesian setting is based on minimizing risk using presumed losses for type I and type II errors. In classical statistics, the choice of type I error probability is used to construct a critical region. This choice is made with a view to making the type II error probability as small as possible. We reject  $H_0$  whenever the data fall in the critical region.

Both approaches use as a basic concept, the likelihood function  $L(\theta|\mathbf{x})$  for the data  $\mathbf{x}$ . Let  $\tilde{\Theta}$  be a random variable taking on one of the two values  $\theta_0, \theta_1$  and having a distribution equal to the prior probability  $\pi$ . Thus,

$$\pi\{\theta_i\} = P\{\tilde{\Theta} = \theta_i\}, i = 0, 1.$$

Recall Bayes formula for events  $A$  and  $C$ ,

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)},$$

we set  $C$  to be the event that the alternative hypothesis is true and  $A$  to be the event that the data take on the value  $\mathbf{x}$ . In symbols,

$$C = \{\tilde{\Theta} = \theta_1\} = \{H_1 \text{ is true}\} \quad \text{and} \quad A = \{X = \mathbf{x}\}.$$

Focus for the moment on the case in which the data are discrete, we have the conditional probabilities for the alternative hypothesis.

$$P(A|C) = P_{\theta_1}\{X = \mathbf{x}\} = f_X(\mathbf{x}|\theta_1) = L(\theta_1|\mathbf{x}).$$

Similarly, for the null hypothesis,

$$P(A|C^c) = P_{\theta_0}\{X = \mathbf{x}\} = f_X(\mathbf{x}|\theta_0) = L(\theta_0|\mathbf{x}).$$

The posterior probability that  $H_1$  is true can be written symbolically in several ways.

$$f_{\tilde{\Theta}|X}(\theta_1|\mathbf{x}) = P\{H_1 \text{ is true}|X = \mathbf{x}\} = P\{\tilde{\Theta} = \theta_1|X = \mathbf{x}\}$$

Returning to Bayes formula, we make the substitutions,

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)},$$

$$f_{\tilde{\Theta}|X}(\theta_1|\mathbf{x}) = \frac{L(\theta_1|\mathbf{x})\pi\{\theta_1\}}{L(\theta_0|\mathbf{x})\pi\{\theta_0\} + L(\theta_1|\mathbf{x})\pi\{\theta_1\}}.$$

By making a similar argument involving limits, we can reach the same identity for the density of continuous random variables. The formula for the posterior probability can be more easily understood if we rewrite the expression above in terms of odds, i. e., as the ratio of probabilities.

$$\frac{f_{\tilde{\Theta}|X}(\theta_1|\mathbf{x})}{f_{\tilde{\Theta}|X}(\theta_0|\mathbf{x})} = \frac{P\{H_1 \text{ is true}|X = \mathbf{x}\}}{P\{H_0 \text{ is true}|X = \mathbf{x}\}} = \frac{P\{\tilde{\Theta} = \theta_1|X = \mathbf{x}\}}{P\{\tilde{\Theta} = \theta_0|X = \mathbf{x}\}} = \frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} \cdot \frac{\pi\{\theta_1\}}{\pi\{\theta_0\}}. \quad (17.12)$$

With this expression we see that the posterior odds are equal to the likelihood ratio times the prior odds.

The decision whether or not to reject  $H_0$  depends on the values assigned for the loss obtained in making an incorrect conclusion. We begin by setting values for the loss. This can be a serious exercise in which a group of experts weighs the evidence for either adverse outcome. We will take a loss of 0 for making a correct decision, a loss of  $\ell_I$  for a type I error and  $\ell_{II}$  for a type II error. We summarize this in a table.

loss function table		
decision	$H_0$ is true	$H_1$ is true
$H_0$	0	$\ell_{II}$
$H_1$	$\ell_I$	0

The Bayes procedure is to make the decision that has the smaller posterior expected loss, also known as the **risk**. If the decision is  $H_0$ , the loss  $\mathcal{L}_0(\mathbf{x})$  takes on two values

$$\mathcal{L}_0(\mathbf{x}) = \begin{cases} 0 & \text{with probability } P\{H_0 \text{ is true}|X = \mathbf{x}\}, \\ \ell_{II} & \text{with probability } P\{H_1 \text{ is true}|X = \mathbf{x}\}. \end{cases}$$

The expected loss

$$E\mathcal{L}_0(\mathbf{x}) = \ell_{II}P\{H_1 \text{ is true}|X = \mathbf{x}\} = \ell_{II}(1 - P\{H_0 \text{ is true}|X = \mathbf{x}\}) \quad (17.13)$$

is simply the product of the loss and the probability of incorrectly choosing  $H_1$ .

If the decision is  $H_1$ , the loss  $\mathcal{L}_1(\mathbf{x})$  also takes on two values

$$\mathcal{L}_1(\mathbf{x}) = \begin{cases} \ell_I & \text{with probability } P\{H_0 \text{ is true}|X = \mathbf{x}\}, \\ 0 & \text{with probability } P\{H_1 \text{ is true}|X = \mathbf{x}\}. \end{cases}$$

In this case, the expected loss

$$E\mathcal{L}_1(\mathbf{x}) = \ell_I P\{H_0 \text{ is true} | X = \mathbf{x}\} \quad (17.14)$$

is a product of the loss and the probability of incorrectly choosing  $H_0$ .

We can now express the Bayesian procedure in symbols using the criterion of smaller posterior expected loss:

$$\text{decide on } H_1 \text{ if and only if } E\mathcal{L}_1(\mathbf{x}) \leq E\mathcal{L}_0(\mathbf{x}).$$

Now substituting for  $E\mathcal{L}_0(\mathbf{x})$  and  $E\mathcal{L}_1(\mathbf{x})$  in (17.13) and (17.14), we find that we make the decision on  $H_1$  and reject  $H_0$  if and only if

$$\begin{aligned} \ell_I P\{H_0 \text{ is true} | X = \mathbf{x}\} &\leq \ell_{II} (1 - P\{H_0 \text{ is true} | X = \mathbf{x}\}) \\ (\ell_I + \ell_{II}) P\{H_0 \text{ is true} | X = \mathbf{x}\} &\leq \ell_{II} \\ P\{H_0 \text{ is true} | X = \mathbf{x}\} &\leq \frac{\ell_{II}}{\ell_I + \ell_{II}} \end{aligned}$$

or stated in terms of odds

$$\frac{P\{H_1 \text{ is true} | X = \mathbf{x}\}}{P\{H_0 \text{ is true} | X = \mathbf{x}\}} \geq \frac{\ell_{II}}{\ell_I}, \quad (17.15)$$

we reject  $H_0$  whenever the posterior odds exceed the ratio of the losses for each type of error.

**Exercise 17.15.** Show that (17.15) follows from the inequality above it.

As we saw in (17.12), this ratio of posterior odds is dependent on the ratio of prior odds. Taking this into account, we see that the criterion for rejecting  $H_0$  is a level test for the likelihood ratio:

Reject  $H_0$  if and only if

$$\frac{L(\theta_1 | \mathbf{x})}{L(\theta_0 | \mathbf{x})} \geq \frac{\ell_{II} / \pi\{\theta_1\}}{\ell_I / \pi\{\theta_0\}}.$$

Thus, the criterion for rejecting  $H_0$  is a level test on the likelihood ratio. This is exactly the same type of criterion as that used in classical statistics. However, the rationale, thus the value for the ratio necessary to reject, is quite different. For example, the lower the value of the prior odds, the higher the likelihood ratio needed to reject  $H_0$  under the Bayesian framework.

**Example 17.16.** For normal observations with means  $\mu_0$  for the null hypothesis and  $\mu_1$  for the alternative hypothesis. If the variance has a known value,  $\sigma_0$ , we have from Example 17.4, the likelihood ratio

$$\frac{L(\mu_1 | \mathbf{x})}{L(\mu_0 | \mathbf{x})} = \exp -\frac{\mu_0 - \mu_1}{2\sigma_0^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0) = \exp \left( -\frac{\mu_0 - \mu_1}{2\sigma_0^2} n(2\bar{x} - \mu_1 - \mu_0) \right).$$

For Example 17.3 on the model and mime butterfly species,  $\mu_0 = 10$ ,  $\mu_1 = 7$ ,  $\sigma_0 = 3$ , and sample mean  $\bar{x} = 8.931$  based on  $n = 16$  observations, we find the likelihood ratio 9.96. Thus,

$$\frac{P\{H_1 \text{ is true} | X = \mathbf{x}\}}{P\{H_0 \text{ is true} | X = \mathbf{x}\}} = \frac{P\{\tilde{M} = \mu_1 | X = \mathbf{x}\}}{P\{\tilde{M} = \mu_0 | X = \mathbf{x}\}} = 9.96 \frac{\pi\{\mu_1\}}{\pi\{\mu_0\}}.$$

where  $\tilde{M}$  is a random variable having a distribution equal to the prior probability  $\pi$  for the model and mimic butterfly wingspan. Consequently, the posterior odds for the mimic vs. mime species is approximately ten times the prior odds.

Finally, the decision will depend on the ratio of  $\ell_{II}/\ell_I$ , i. e., the ratio of the loss due to letting the invasion by the mimic take its course versus eradication of the resident model species.

**Example 17.17.** Returning to a previous example, give the likelihood ratios for  $n = 20$  Bernoulli trials with  $p_0 = 0.6$  and  $p_1 = 0.7$  for values  $x = 0, \dots, 20$  for the number of successes. Give the values for the number of successes in which the number of successes change the prior odds by a factor of 5 or more as given by the posterior odds.

## 17.7 Answers to Selected Exercises

17.2 Flip a biased coin in which the probability of heads is  $\alpha$  under both the null and alternative hypotheses and reject whenever heads turns up. Then

$$\alpha = P_{\theta_0}\{\text{heads}\} = P_{\theta_1}\{\text{heads}\} = 1 - \beta.$$

Thus, the receiver operator characteristic curve is the line through the origin having slope 1.

17.4. The likelihood ratio

$$\frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} = \exp -\frac{\mu_0 - \mu_1}{2\sigma_0^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0) \geq k_\alpha.$$

Thus,

$$\begin{aligned} \frac{\mu_1 - \mu_0}{2\sigma_0^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0) &\geq \ln k_\alpha \\ \sum_{i=1}^n (2x_i - \mu_1 - \mu_0) &\geq \frac{2\sigma_0^2}{\mu_1 - \mu_0} \ln k_\alpha \\ 2\bar{x} - \mu_1 - \mu_0 &\leq \frac{2\sigma_0^2}{n(\mu_1 - \mu_0)} \ln k_\alpha \\ \bar{x} &\leq \frac{1}{2} \left( \frac{2\sigma_0^2}{n(\mu_1 - \mu_0)} \ln k_\alpha + \mu_1 + \mu_0 \right) = \tilde{k}_\alpha \end{aligned}$$

Notice that since  $\mu_1 < \mu_0$ , division by  $\mu_1 - \mu_0$  changes the direction of the inequality.

17.6. If  $c_\alpha$  is the critical value in expression in (17.2) then

$$\frac{\mu_1 - \mu_0}{2\sigma_0^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0) \geq c_\alpha$$

Since  $\mu_1 > \mu_0$ , division by  $\mu_1 - \mu_0$  does not change the direction of the inequality. The rest of the argument proceeds as before. we obtain that  $\bar{x} \geq \tilde{k}_\alpha$ .

17.7. If power means easier to distinguish using the data, then this is true when the means are farther apart, the measurements are less variable or the number of measurements increases. This can be seen explicitly in the power equation (17.4).

17.8. For  $\mu_0 > \mu_1$ ,

$$\begin{aligned} \beta &= P_{\mu_1}\{X \notin C\} = P_{\mu_1}\{\bar{X} > \mu_0 - \frac{\sigma_0}{\sqrt{n}} z_\alpha\} \\ &= P_{\mu_1}\left\{\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}} > -z_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right\} = 1 - \Phi\left(-z_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right) \end{aligned}$$

and the power

$$1 - \beta = \Phi\left(-z_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right).$$

17.9 Interpreting equation (17.4) in R, we find that

```
> mu0<-10
> sigma0<-3
> n<-16
```

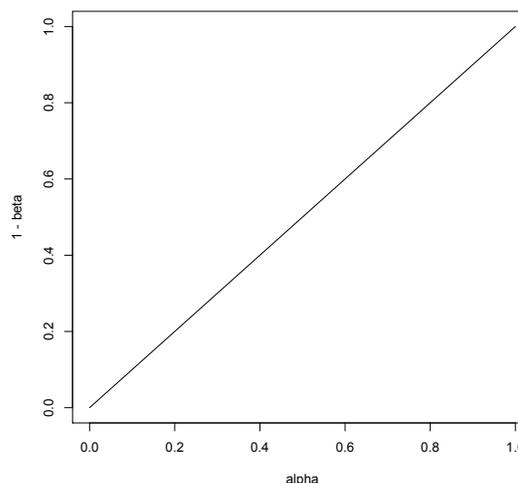


Figure 17.6: Receiver Operator Characteristic based on a biased coin toss. Thus, any viable ROC should be above the line the the graph.

```

> zalpha<-qnorm(0.995)
> mu1<-c(9, 8, 7)
> power<-1-pnorm(zalpha-abs(mu1-mu0)/(sigma0/sqrt(n)))
> data.frame(mu1,power)
  mu1    power
1    9 0.1070268
2    8 0.5361891
3    7 0.9228015

```

Notice that the power has decreased from the case  $\alpha = 0.05$ . This could be anticipated. In reducing the significance level from  $\alpha = 0.05$  to  $\alpha = 0.01$ , we make the criterion for rejecting more stringent by reducing the critical region  $C$ . The effect can be seen in Figure 17.2. On the left, the vertical dashed line is moved left to reduce the area under the black curve to the left of the dashed line. This, in turn, reduces the area under the other curves to the left of the dashed line. On the right figure, the vertical dashed line is the line  $\alpha = 0.01$  and, because the ROC curve is increasing, the values for the power is decreased

17.11. For the likelihood ratio (17.6), take the logarithm to obtain

$$\ln \left( \frac{L(p_1|\mathbf{x})}{L(p_0|\mathbf{x})} \right) = n \ln \left( \frac{1-p_1}{1-p_0} \right) + (x_1 + \cdots + x_n) \ln \left( \left( \frac{p_1}{1-p_1} \right) / \left( \frac{p_0}{1-p_0} \right) \right) \geq \ln k_\alpha.$$

If  $p_0 < p_1$  then the ratio in the expression for the logarithm in the second term is greater than 1 and consequently, the logarithm is positive. Thus, we isolate the sum  $\sum_{i=1}^n x_i$  to give the test (17.7). For  $p_0 > p_1$ , the logarithm is negative and the direction of the inequality in (17.7) is reversed.

17.12. If we take  $\alpha = 0.10$ , then

$$\begin{aligned} P\{N \geq 15\} &= 1 - P\{N \leq 14\} = 1 - 0.8744 = 0.1256 > 0.10 \\ P\{N \geq 16\} &= 1 - P\{N \leq 15\} = 1 - 0.9491 = 0.0509 < 0.10 \end{aligned}$$

Consequently, we need to have at least 16 successes in order to reject  $H_0$ . If we take  $\alpha = 0.01$ , then

$$\begin{aligned} P\{N \geq 17\} &= 1 - P\{N \leq 16\} = 1 - 0.9840 = 0.0160 > 0.01 \\ P\{N \geq 18\} &= 1 - P\{N \leq 17\} = 1 - 0.9964 = 0.0036 < 0.01 \end{aligned}$$

Consequently, we need to have at least 18 successes in order to reject  $H_0$ . For  $C = \{16, 17, 18, 19, 20\}$ ,

$$P\{N \in C\} = 1 - P\{N \leq 15\} = 1 - 0.9491 = 0.0509.$$

Thus,  $\alpha$  must be less than 0.0509 for  $C$  to be a critical region. In addition,  $P\{N \geq 17\} = 0.0160$ . Consequently, if we take any value for  $\alpha < 0.0160$ , then the critical region will be smaller than  $C$ .

17.15. Begin with

$$P\{H_0 \text{ is true} | X = \mathbf{x}\} \leq \frac{\ell_{\text{I}}}{\ell_{\text{I}} + \ell_{\text{II}}}.$$

or, taking reciprocals,

$$\frac{1}{P\{H_0 \text{ is true} | X = \mathbf{x}\}} \geq \frac{\ell_{\text{I}} + \ell_{\text{II}}}{\ell_{\text{I}}}.$$

Because  $P\{H_0 \text{ is true} | X = \mathbf{x}\} + P\{H_1 \text{ is true} | X = \mathbf{x}\} = 1$ , we have that

$$P\{H_1 \text{ is true} | X = \mathbf{x}\} \geq \frac{\ell_{\text{II}}}{\ell_{\text{I}} + \ell_{\text{II}}}.$$

Now multiply the two inequalities above. Because all the quantities are positive, the product of the left sides is greater than the product of the right sides and so the inequality (17.15) holds.

17.17. Using the likelihood ratio formula in Example 17.9.

```
> x<-c(0:20)
> n<-20
> p0<-0.6
> p1<-0.7
> ratio<-((1-p1)/(1-p0))^n*((p1/(1-p1))/(p0/(1-p0)))^x
> data.frame(x[1:7],ratio[1:7],x[8:14],ratio[8:14],x[15:21],ratio[15:21])
```

	x.1.7.	ratio.1.7.	x.8.14.	ratio.8.14.	x.15.21.	ratio.15.21.
1	0	0.003171212	7	0.06989143	14	1.540361
2	1	0.004932996	8	0.10872001	15	2.396118
3	2	0.007673550	9	0.16912001	16	3.727294
4	3	0.011936633	10	0.26307558	17	5.798013
5	4	0.018568096	11	0.40922867	18	9.019132
6	5	0.028883705	12	0.63657794	19	14.029761
7	6	0.044930208	13	0.99023235	20	21.824072

Thus, values  $x \geq 17$  increase the posterior odds in favor of  $H_1$  by a factor greater than 5. values  $x \leq 9$  increase the posterior odds in favor of  $H_0$  by a factor greater than 5.