Power of Tests

April 20, 2020

So far in this Chapter you have been given a Hypothesis test, and then used this test to analyze the data. This basically meant deciding to reject or fail to reject the Null Hypothesis. This is somewhat similar to in Chapter 8, where you were given a estimator and used it to estimate a parameter. Then in chapter 9, we did some analysis of these estimators to see which ones are good, and then saw how to derive a estimator. Now we will begin to see how to analyze Hypothesis tests and to derive them.

Section 10.10 has several new ideas, so I'll split them over several lectures.

1 The power of a test

In our tests there are two types of errors, type I and type II. The level of an acceptable type I error is given in the problem, because of the philosophy to "fail to reject" the null hypothesis rather than accept it, this makes sense. Because we are going to accept the alternative hypothesis when the data suggests it, we need to start by defining an acceptable amount of error that we can handle, this level of error is exactly the probability of a type I error.

Nevertheless, we would like the probability of a type II error to be as small as possible. This comes from choosing a good test. A way we can quantify this, is by what is called the power of the a test:

DEFINITION 1.1. *The* ***power of the test*** *is the probability that the test will lead to a rejection of* H_0 *when the actual parameter value is* θ.

In other words, if W *is the test statistic,* θ *is the unknown parameter, and the rejection rejection is* RR*, then*

$$\textit{power}(\theta) = \mathbb{P}(W \in RR \textit{ when the parameter value is } \theta).$$

So the power of a test is a function of θ, the unknown parameter. Since we don't know the actual value of θ is, we would like have a test that has a good power for all values of θ. We think of the power of the test differently for θ in the null hypothesis and for θ in the alternate hypothesis.

When the null hypothesis is $H_0 : \theta = \theta_0$, then

$$\text{power}(\theta_0) = \mathbb{P}(W \in RR \text{ when the parameter value is } \theta_0) = \alpha.$$

The first equality is the definition of the power and the second is the definition of α. So we see when $\theta = \theta_0$, a small power is desirable.

On the other hand when θ is a value in H_a, then we want to reject the null hypothesis with high probability and therefore a large power is desirable. For θ_a in H_a we have

$$\begin{aligned}\text{power}(\theta_a) &= \mathbb{P}(W \in RR \text{ when the parameter value is } \theta_a) \\ &= 1 - \mathbb{P}(W \notin RR \text{ when the parameter value is } \theta_a) \\ &= 1 - \beta(\theta_a).\end{aligned}$$

Once again the first line is the definition of power, the second is just switching from $W \in RR$ to the complement event $W \notin RR$, the final line is the definition β for the given θ_a.

So we would like β to be large in H_a and small in H_0. The ideal curve is in the book Figure 10.14, where power $(\theta_0) = 0$ for θ in H_0 and for θ in H_a we want power$(\theta) = 1$. But if we could design such a test, we would need to know θ already, and not need to do statistics. So we instead we will get a function that looks like Figure 10.13 in the book.

The power function will generally be a continuous function of θ, with power$(\theta_0) = \alpha$. Our goal is, for a given α, to find a test whose RR maximizes the power for all θ in H_a.

Let's do an example, where we're given a test and an α and we'll compute the power function. The first example will not be a very good statistical test, but everything will be quite explicit to compute. Then we'll do a more standard example.

EXAMPLE 1.2. *We have a coin that we suspect is biased. In order to test this null hypothesis that the coin is fair against the alternative hypothesis that coin is not fair, we flip it 4 times. If we don't get exactly 2 heads we reject the null hypothesis in favor of the coin not being fair.*

Compute the power of this test.

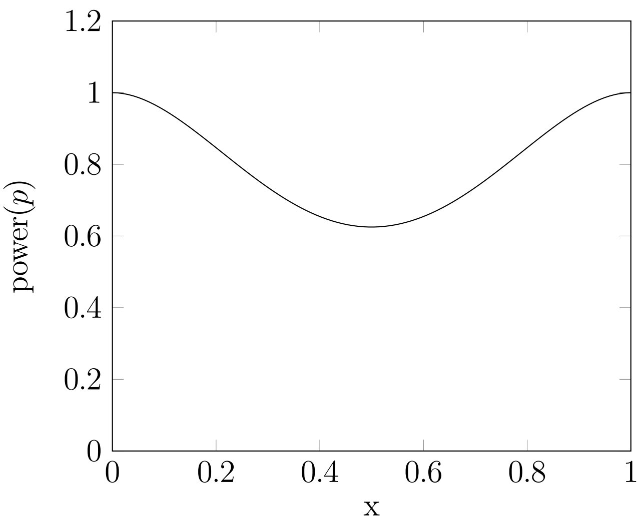
Solution: Let p be the probability the coin is heads when flipped. Let W be the number of heads. The RR is $W \neq 2$.

Then W is binomial(4,p) and then for any value of p, the probability $W \in RR$ is

$$\begin{aligned}\mathbb{P}(W \in RR) = \mathbb{P}(W \neq 2) &= \binom{4}{0}(1-p)^4 + \binom{4}{1}p^1(1-p)^3 + \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4 \\ &= (1-p)^4 + 4(1-p)^3 p + 4(1-p)p^3 + p^4\end{aligned}$$

so the answer is

$$\text{power}(p) = (1-p)^4 + 4(1-p)^3 p + 4(1-p)p^3 + p^4$$



In this problem the RR was just given, somewhat arbitrarily, instead of being determined by α. In this case we see α is the minimum value of the power curve and equals

$$\alpha = \mathbb{P}(W=2 \text{ given that } p=1/2) = (1-1/2)^4 + 4(1-1/2)^3 1/2 + 4(1-1/2)1/2^3 + 1/2^4 = \frac{10}{2^4}.$$

Let's now do an example that is a bit more similar to what we usually look at.

EXAMPLE 1.3. *Let $Y_1,\ldots,Y_{100}$ be a random sample of normal random variables with mean θ and variance 1. Use the text statistic $W = \frac{1}{100}\sum_{i=1}^{100} Y_i$ to test the hypothesis $H_0 : \theta_0 = 10$ against $H_a : \theta_0 \neq 10$. Determine the rejection region for an $\alpha = .05$ test and the power for this test.*

Solution:

To determine the rejection region, we assume the null hypothesis is true, so the test statistic $W = \frac{1}{100}\sum_{i=1}^{100} Y_i$ is a normal random variable with mean 10 and variance $1/100$. It is normal because the Y_i's are normal, and the sum of normal random variables is also normal. The mean is 10 because the mean of the Y_i's is 10, and the variance is $\text{Var}(W) = \text{Var}(\frac{1}{100}\sum_{i=1}^{100} Y_i) = 1/100^2 * 100 * \text{Var}(Y_1) = 1/100$.

We choose the rejection region so that when H_0 is true $\mathbb{P}(W \notin RR) = .05$. Choosing a symmetric region around $\theta_0 = 10$ we want to find a such that

$$\mathbb{P}(10 - a \leq W \leq 10 + a) = .05.$$

We then standardize W to get that this is equivalent to:

$$\mathbb{P}(-\frac{a}{1/\sqrt{100}} \leq \frac{W-10}{1/\sqrt{100}} \leq \frac{a}{1/\sqrt{100}}) = .05.$$

so we see $10a = z_{.025}$. (since we're doing a two sided test, we want half of the probability of a type I error to be in the upper trail and half in the lower tail). Another way of writing the above equation, substituting $10a = z_{.025}$, is

$$\mathbb{P}(-z_{.025} \leq 10(W-10) \leq z_{.025}) = .05.$$

This means the rejection region is:

$$RR \text{ is } W \text{ such that } |W-10| > z_{.05/2}/\sqrt{100} = 1.96/10 = .196$$

Now for any other value of θ we want to compute 1 minus the probability that a normal random variable with mean θ and variance $1/100$ lands in the set $(10-.196, 10+.196)$, mathematically:

$$\text{power}(\theta) = 1 - \mathbb{P}(W \in (10-.196, 10+.196) \text{ with } W \text{ a normal random variable with mean } \theta \text{ and variance } 1/100$$

Standardizing the normal random variable gives:

$$\mathbb{P}(W \in (10-.196, 10+.196)) = \mathbb{P}(10-.196 \le W \le 10+.196) = \mathbb{P}\left(\frac{10-\theta-.196}{10} \le \frac{W-\theta}{10} \le \frac{10-\theta+.196}{10}\right)$$

$$= \mathbb{P}\left(Z \in \left(\frac{10-\theta-.196}{10}, \frac{10-\theta+.196}{10}\right)\right) \text{ for } Z \text{ a standard normal rv}$$

The number on the right is $\beta(\theta)$ so the power is

$$\text{power}(\theta) = 1 - \mathbb{P}\left(Z \in \left(\frac{10-\theta-.196}{10}, \frac{10-\theta+.196}{10}\right)\right) \text{ for } Z \text{ a standard normal rv}$$

It is not so explicit, because we don't have a way of explicitly compute the integral of the pdf of a normal random variable. But for any θ, this number can be computed from the tables or with a computer. As θ gets farther from 10, this gets smaller, so the power gets closer and closer to 1.