LECTURE: POWER OF HYPOTHESIS TESTS

Goal: How "good" is a hypothesis test?

1. Power of a Test

So far: We used α and β to measure the "goodness" of our test.

Ideally we would like a function which gives us the error of the test given the true value θ of the parameter. This is called:

Definition:

The **power** of a hypothesis test is

$$\text{Power}(\theta) = P(\hat{\theta} \text{ lies in RR when true parameter value is } \theta)$$

Example 1:

Calculate Power(θ_0) where $\theta = \theta_0$ is the Null Hypothesis

$$\begin{aligned}\text{Power}(\theta_0) =& P(\hat{\theta} \text{ is in RR when } \theta = \theta_0) \\ =& P(\text{ Reject Null when Null is true }) \\ =& \alpha\end{aligned}$$

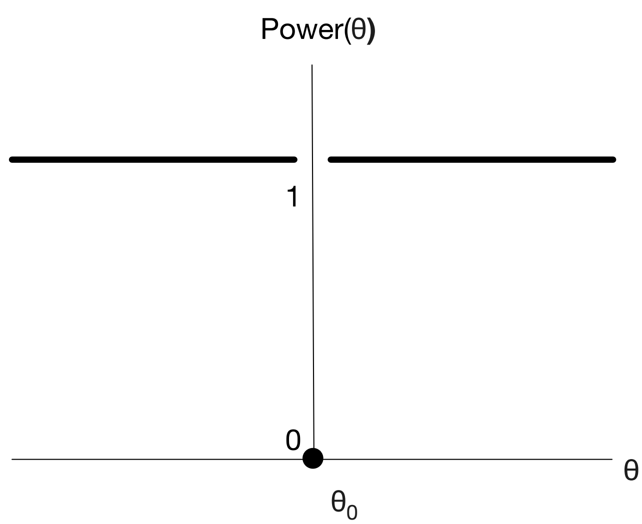
Example 2:

Power(θ_a) where $\theta = \theta_a$ is a specific value of the Alt Hypothesis

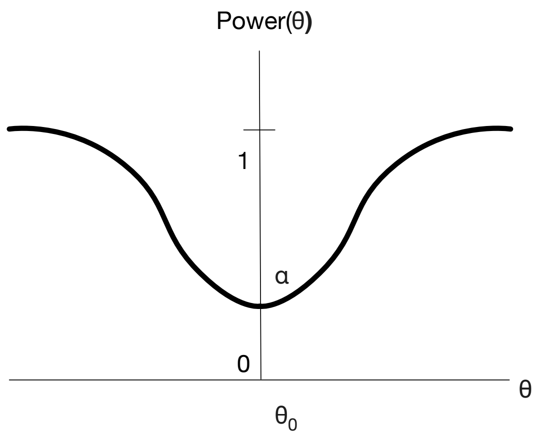
$$\begin{aligned}\text{Power}(\theta_a) =& P(\hat{\theta} \text{ is in RR when } \theta = \theta_a)\\ =& P(\text{ Reject Null when } \theta = \theta_a\ )\\ =& 1 - P(\text{ Accept Null when } \theta = \theta_a\ )\\ =& 1 - \beta(\theta_a)\end{aligned}$$

$\beta(\theta_a)$ is the probability of a type II error when the true value is $\theta = \theta_a$

For an ideal test, the power function would be 0 at θ_0 and 1 for all possible values of θ_a



No hypothesis test, however, is perfect. Realistically, the power curve for a two-tailed hypothesis test will look more like this:



We would like a test which maximizes the power function for a given α. This is called the **most powerful α-level test**. In order to do that, we need to define the concept of a simple hypothesis.

Definition:

A **simple hypothesis** is a hypothesis whose parameters completely determine the distribution of our population

Otherwise, it is called a **composite hypothesis**.

This is best understood with examples:

Example 3:

Suppose our population $\sim N(\mu, 1)$

Any hypothesis involving μ such as "$\mu = 5$" is a simple hypothesis, since its parameter μ completely determines the distribution of our population.

Example 4:

Suppose our population $\sim N(\mu, \sigma^2)$ where σ is unknown

Here we need both μ and σ^2 to determine the distribution of the population, so a hypothesis involving μ such as "$\mu = 5$" is a composite hypothesis.

Example 5:

Suppose our population $\sim$ Exp (λ)

Any hypothesis involving λ is simple since λ completely determines the distribution of the population.

Since $\mu = 1/\lambda$ any hypothesis involving μ is also simple

2. Neyman-Pearson Lemma

Consider a hypothesis test where we are testing a **simple** null hypothesis $H_0 : \theta = \theta_0$ against a **simple** alternative hypothesis $H_a : \theta = \theta_a$.

We want to find a rejection region such that:

(1) Power$(\theta_0) = \alpha$

(2) Power(θ_a) is as large as possible

The following theorem tells us how to derive the most powerful α-level test in this case

Neyman-Pearson Lemma:

Let $Y_1, \ldots, Y_n$ be iid samples from a population

Let $L(Y_1, \ldots, Y_n|\theta)$ be the likelihood function for the sample

Then the most powerful $\alpha-$level test has a RR given by

$$\frac{L(Y_1, \ldots, Y_n|\theta_0)}{L(Y_1, \ldots, Y_n|\theta_a)} < k$$

Here k is chosen so that the test has the desired α

Note: The fraction above is called a **likelihood ratio**

Example 6:

Suppose Y is a single observation from a population with pdf

$$f_\theta(y) = \begin{cases} \theta y^{\theta-1} & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the most powerful hypothesis test with $\alpha = 0.05$ to test $H_0 : \theta = 2$ against $H_a : \theta = 1$

STEP 1: Both hypotheses are simple, so we can use Neyman-Pearson

Since there is only one sample $n = 1$ the likelihood ratio is

$$\frac{L(Y|\theta_0)}{L(Y|\theta_a)} = \frac{f_{\theta_0}(Y)}{f_{\theta_a}(Y)} = \frac{2Y^{2-1}}{1Y^{1-1}} = 2Y$$

We used $\theta_0 = 2$ by Null and $\theta_a = 1$ by the alternative hypothesis.

STEP 2: By Neyman-Pearson, the RR is $\{2Y < k\} = \{Y < k/2\}$ where we will determine k based on $\alpha = 0.05$

$$\begin{aligned} 0.05 = \alpha &= P(Y \text{ lies in RR when Null is true}) \\ &= P(Y < k/2 \text{ when } \theta = 2) \\ &= \int_0^{k/2} 2y dy \\ &= \left(\frac{k}{2}\right)^2 \end{aligned}$$

This gives $\frac{k}{2} = \sqrt{0.05} = 0.2236$

STEP 3: Hence the RR for the 0.05-most powerful test is:

$$\left\{Y < \frac{k}{2}\right\} = \{Y < 0.2236\}$$

In other words, among all hypothesis tests for $H_0 : \theta = 2$ versus $H_a : \theta = 1$ based on $n = 1$ and $\alpha = 0.05$, a test with this rejection region has the largest possible value for $\text{Power}(\theta_a) = \text{Power}(1)$

(b) In this case, find $\text{Power}(\theta_a) = \text{Power}(1)$

$$\begin{aligned} \text{Power}(1) &= P(Y \text{ lies in RR when } \theta = 1) \\ &= P(Y < 0.2236 \text{ when } \theta = 1) \\ &= \int_0^{0.2236} 1 dy \\ &= 0.2236 \end{aligned}$$

The value 0.2236 is the maximum value of the power of the test among all tests with $\alpha = 0.05$. But for this test, $\beta = 1 - 0.2236 = 0.7764$, which is very large. So this test is not very good. However, no other test with these same parameters is any better

What about our hypothesis tests from the previous sections? It turns out that they are the best hypothesis tests for the given situations:

Fact:

Let $Y_1, \dots, Y_n$ be iid $N(\mu, \sigma^2)$ where μ is unknown but σ^2 is known.

Suppose we're testing $H_0 : \mu = \mu_0$ against $H_a : \mu > \mu_0$

Then the α-most powerful test is given by $\left\{\bar{Y} > k\right\}$ where

$$k = \mu_0 + z_\alpha \left(\frac{\sigma}{\sqrt{n}}\right)$$

So in fact the tests we've seen are the best tests. The proof of this result is an application of the Neyman-Pearson lemma