# 9 Hypothesis Testing

Statistical Hypotheses testing is a formal means of choosing between two distributions on the basis of a particular statistic or random variable generated from one of them.

## 9.1 Neyman-Pearson Paradigm

- Null hypothesis $\mathcal{H}_0$
- Alternate hypothesis $\mathcal{H}_A$ or $\mathcal{H}_1$

There are two types of hypotheses, simple ones where the hypothesis completely specifies the distribution. Here is an example when they are both compositie:
$X_i \sim$ **Poisson** with unknown parameter
$X_i$ is not **Poisson**
Simple hypotheses test one value of the parameter against another, the form of the distribution remaining fixed.

## 9.2 Two Types of Error-the burnt toast and the smoke detector.

| | | Reality | |
|---|---|---|---|
| | | $\mathcal{H}_0$ true | $\mathcal{H}_0$ false |
| Test says | accept $\mathcal{H}_0$ | Good | Type II error |
| | reject $\mathcal{H}_0$ | Type I error | Good |

## Setting up tests

1. Define the null hypothesis $\mathcal{H}_0$(devil's advocate).
2. Define the alternative $\mathcal{H}_A$(one sided /two sided).
3. Find the test statistic.
4. Decide on the type I error: $\alpha$ that you are willing to take.
5. Compute the Probability of observing the data given the null hypothesis: P-value.
6. Compare the P-value to $\alpha$, if its smaller, reject $\mathcal{H}_0$.

### 9.2.1 Example

ESP experiment : guess the color of 52 cards with replacement.

$$\begin{array}{lll} \mathcal{H}_0 & T \sim \mathcal{B}(1/2, n = 10) & \\ \mathcal{H}_1 & T \sim \mathcal{B}(p, n = 10), & p > 1/2 \end{array}$$

| p | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| .7 | .0000 | .0001 | .0014 | .0090 | .0368 | .1029 | .2001 | .2668 | .2335 | .1211 | .0282 |
| .6 | .0001 | .0016 | .0106 | .0425 | .1115 | .2007 | .2508 | .2150 | .1209 | .0403 | .0060 |
| .5 | .0010 | .0098 | .0439 | .1172 | .2051 | .2461 | .2051 | .1172 | .0439 | .0098 | .0010 |

Rejection region $R = \{8, 9, 10\}$.

$$\alpha = P(X > 7) = P(R) = 0.0439 + 0.0098 + 0.0010 = 0.0547$$

Rejection region= $R = \{7, 8, 9, 10\}$.

$$\alpha = P(X > 6) = P(R) = 0.1172 + 0.0439 + 0.0098 + 0.0010 = 0.1719$$

In the setup, we have to choose $\alpha$ first. Then we can compute what the power will be under various values of $p$:

$$\begin{array}{ll} p = 0.6 & P(X > 7|p = 0.6) = 0.1673 \\ p = 0.7 & P(X > 7|p = 0.7) = 0.3828 \end{array}$$

As p approaches 1, the power approaches one. As p approaches 0.5 the power becomes equal to $\alpha$.

## 9.3 Bayesian Version

We can give a probability of $\mathcal{H}_0$ being true given the data, this can actually be more meaningful than the P-value.

Consider testing:
$\mathcal{H}_0 : \theta \leq \theta_0$ against $\mathcal{H}_1 : \theta > \theta_0$,
we can calculate under the Bayesian model $P(\theta \leq \theta_0)$, the integral of the pdf over $(-\infty, \theta)$.
Of course this poses a problem if the pdf is continuous and one of the hypotheses is simple.
Example: Ten sets of twins, randomized, one is given A, one is given B.

$$p = P(\text{trt A better than trt B})$$

$\mathcal{H}_0 : p = \frac{1}{2}$
$\mathcal{H}_A : p \neq \frac{1}{2}$

We observe $Y$, number of cases where A is better, $Y \sim \mathcal{B}(10, p)$.

Assign a positive prior point probability to the null hypothesis: $P(p = \frac{1}{2}) = \gamma$. Suppose we take a symmetric prior pdf for $p \neq \frac{1}{2}$,

$$g(p|p \neq \frac{1}{2}) = 6p(1-p), \qquad 0 < p < 1$$

If the experiment results in 9 wins for A and one for B, the likelihood, given this data is: $L(p) = P(Y = 9|p) = 10p^9(1-p)$ We need the unconditional probability of $Y = 9$.

$$\begin{aligned} P(Y = 9) &= \gamma \times (Y = 9|p = .5) + (1-\gamma) \times \int_0^1 P(Y = 9|p \neq .5) g(p) dp \\ &= \gamma \times 10(.5)^{10} + (1-\gamma) \times 10 \int_0^1 6p^{10}(1-p)^2 dp \\ &= 10\{\frac{\gamma}{1024} + \frac{1-\gamma}{143}\} \end{aligned}$$

$$P(\mathcal{H}_0|\text{data}) = \frac{143\gamma}{143\gamma + 1024(1-\gamma)}$$

When $\gamma = 0.5$, P(no trt effect)=0.123 after nine success; if the prior is $\gamma = .1$, the posterior probability of $\mathcal{H}_0$ beccomes .015.

## 9.4 Optimal Tests for simple hypotheses

- Null hypothesis $\mathcal{H}_0 : f = f_0$
- Alternate hypothesis $\mathcal{H}_A : f = f_1$

We want to find a rejection region $R$ such that the error of both types are as small as possible.

$$\int_R f_0(x)dx = \alpha \qquad \text{and} \qquad 1 - \beta = \int_R f_1 dx$$

## Neyman-Pearson Lemma

**Theorem 9.1** *For testing $f_0(x)$ against $f_1(x)$ a critical region of the form*

$$\Lambda(x) = \frac{f_0(x)}{f_1(x)} < K$$

*where K is constant has the greatest power (smallest $\beta$) in the class of tests with the same $\alpha$*

$$\left(\int_R - \int_S\right) f = \int_{R\cap S^c} f - \int_{S\cap R^c} f$$

$$\beta_R - \beta_S = \left(\int_R - \int_S\right) f_1 = \int_{R\cap S^c} f_1 - \int_{S\cap R^c} f_1$$

Since in $R$ , $f_1 \geq f_0/K$ and in $R^c$, $-f_1 \geq -f_0/K$ we have:

$$\begin{aligned}\beta_S - \beta_R &\geq \frac{1}{K}\left(\int_{R\cap S^c} f_0 - \int_{S\cap R^c} f_0\right) \\ &= \frac{1}{K}\left(\int_R f_0 - \int_S f_0\right) = \frac{1}{K}(\alpha_R - \alpha_S)\end{aligned}$$

Example: Two simple hypotheses about Bernouilli parameter:

$$\mathcal{H}_0 : p = 0.5 \qquad \mathcal{H}_1 : p = 0.25$$

| $y$ | $f_0(y)$ | $f_1(y)$ | $\Lambda(y)$ |
|---|---|---|---|
| 0 | .0312 | .2373 | .13 |
| 1 | .1562 | .3955 | .40 |
| 2 | .3125 | .2637 | 1.2 |
| 3 | .3125 | .0879 | 3.6 |
| 4 | .1562 | .0146 | 10.6 |
| 5 | .0312 | .0010 | 31.2 |

| R | $\alpha$ | $\beta$ | Power |
|---|---|---|---|
| { } | 0 | 1 | 0 |
| {0 } | 0.031 | .763 | 0.237 |
| {0,1 } | 0.187 | .367 | 0.633 |
| {0,1,2 } | 0.500 | .103 | 0.897 |

### 9.4.1 Neyman Pearson Test for $\mathcal{N}(0, \theta)$)

Given a random sample from $X$ the likelihood is

$$L(\theta) = \theta^{-n/2} \exp -\frac{1}{2\theta} \sum X_i^2$$

For the NP likelihood ration test the rejection region for testing $\theta_1$ against $\theta_2$is of the form:

$$\Lambda = (\frac{\theta_1}{\theta_2})^{-n/2} \exp -(\frac{1}{2\theta_1} - \frac{1}{2\theta_2}) \sum X_i^2 < K'$$

If $\theta_1 < \theta_2$ this inequality holds for a given $K'$ if and only if, for some $K$, $\sum X_i^2 > K$

The distribution of $\sum X_i^2/\theta$ is $\chi_n^2$ and we can thus find the error sizes in this case:

$$\alpha = P(\sum X_i^2 > K|\theta_1) = 1 - F_{\chi_n^2}(\frac{K}{\theta_1})$$

$$\beta = P(\sum X_i^2 < K|\theta_2) = F_{\chi_n^2}(\frac{K}{\theta_2})$$

## 9.5 Composite Hypotheses

Typically one or both hypotheses are composite, suppose $\mathcal{H}_A$ is composite, a test that is most powerful for all simple hypotheses conatined in $\mathcal{H}_A$ is called uniformly most powerful.

Example:

Suppose we test $\mathcal{H}_0 : p = 0.5$ based on $Y$, the number of 1's in a random sample of size $n = 5$. We saw that rejecting $\mathcal{H}_0$ for $Y < K$. We saw that rejecting $\mathcal{H}_0$ for $Y < K$ is most powerful agains $p = 0.25$ in the class of tests with $\alpha = 0.187$. But the same is true if we were to use an arbitrary $p < 0.5$. The likelihood ratio is:

$$\Lambda = \frac{.5^Y .5^{5-Y}}{p^Y(1-p)^{5-Y}} \propto (\frac{1}{p} - 1)^Y$$

An increasing function of $Y$ when $p < \frac{1}{2}$. So, $\Lambda$ is less than a constant when $Y$ is less than a constant. Thus $Y < K$ is most powerful against every $p < .5$; it is UMP against $\mathcal{H}_A : p < .5$.

Note:

The rejection region $\{0, 1\}$ for $Y$ is UMP for $p = 0.5$ against $p < 0.5$, in the class of tests with $\alpha = 0.187$. The power function is

$$\pi(p) = P(Y = 0 \text{ or } 1|p) = (1-p)^5 + 5p(1-p)^4 = (1-p)^4(1+4p)$$

This is a decreasing function of p on (0,1), so its largest value on $p \geq .5$ is at $p = 0.5$. So $\alpha^* = .187 = \pi(.5) = max \alpha's$.

Since $\{Y \leq 1\}$ has the greatest power at any alternative $p < .5$ among tests with $\alpha \leq .187$ it has greatest power at any alternative among tests with $\alpha^* = .187$, it is UMP for $p \geq .5$ against $p < .5$.