Math 152 3/10/09

Coin 0 $p = .5$
Coin 1 $p = .7$

One of the coins is chosen and tossed 10 times.

X = # of heads is observed.

Which coin was tossed?

Using R:
- prob vector for Coin 0. → p0
- prob vector for Coin 1 → p1.
- likelihood ratios: p0/p1

  p1/p0

Bayesian Approach to <u>Hypothesis</u> testing:

H_0: Coin 0 was tossed

H_1: Coin 1 was tossed.

<u>Prior:</u> $P(H_0)$, $P(H_1)$: prior probabilities of each hypothesis

e.g. if there is no reason to think Coin 0 was more (or less) likely than Coin 1, $P(H_0) = P(H_1) = 1/2$.

After observing X, compute the posterior probabilities

$$P(H_0/x) = \frac{P(H_0, x)}{P(x)} = \frac{P(x/H_0)P(H_0)}{P(x)}$$

$$P(H_1/x) = \frac{P(H_1, x)}{P(x)} = \frac{P(x/H_1)P(H_1)}{P(x)}.$$

Then the ratio of posterior probabilities is

$$\frac{P(H_0/x)}{P(H_1/x)} = \frac{P(x/H_0)P(H_0)}{P(x/H_1)P(H_1)}$$

and we choose H_0 ~~iff~~ this is larger than 1.

i.e.

$$\text{iff} \quad \frac{P(x/H_0)}{P(x/H_1)} > \frac{P(H_1)}{P(H_0)}$$

i.e.

We <u>accept</u> H_0 if $X=x$ is such that

$$\frac{P(x/H_0)}{P(x/H_1)} > c = \frac{P(H_1)}{P(H_0)}$$

(and the set of such x is called an acceptance region for H_0).

if $P(H_1) = P(H_0) = 1/2$, the acceptance region is $X = 0,1,2,3,4,5,6$.

There are two possible errors that can be made.

1: reject H_0 when it is true
(= accept H_1 when it is false)

2: accept H_0 when it is false.
(= reject H_1 when it is true).

In our example, (and taking $c = 1$)

H_0 is accepted if $X = 0, 1, \ldots, 6$
and is rejected otherwise.

$$\text{Prob}(\text{reject } H_0 \,/\, H_0) \cong .117 + .044 + .0098 + .00098 \simeq .1718$$

R $\boxed{\text{sum}(p0[8:11])}$

$$\text{Prob}(\text{accept } H_0 \,/\, H_1) \simeq .3504$$

$$= \text{Prob}(X \leq 6 \,/\, H_1)$$

R $\boxed{= \text{sum}(p1[1:7])}$

e.g

If $\dfrac{P(H_0)}{P(H_1)} = 10$ then H_0 is

accepted for any x s.t.

$$\frac{P(x/H_0)}{P(x/H_1)} > \frac{1}{10}$$

i.e. for $X = 0, \ldots, 8$

and now we have

$$\text{Prob}(\text{reject } H_0 \mid H_0) \hat{=} .009765 + .0009765$$
$$\hat{=} .0107$$

$$\text{Prob}(\text{accept } H_0 \mid H_1) \hat{=} .8507.$$

The Prior probabilities control the probabilities of the two types of error. (in the Bayesian approach).

<u>The Neyman - Pearson approach to Hypothesis testing</u>.

- Set the probabilities of the ~~two~~ types I error in advance.
- To do this requires singling out one of the Hypotheses.

We usually call the ~~"special~~ singled out hypothes H_0, the Null Hypotheses.

and the "other" is called the alternative hypothesis. H_1 or H_A.

<u>Vocabulary:</u>

- Rejecting H_0 when it is true is call a type I error

- Prob. of a type I error is called the significance level of the test. (usually written α)

- accepting H_0 when it is false is called a type II error

  Prob(type II error) $= \beta$

- Prob(rejecting H_0 when it is false) $= 1-\beta$ and is called the "Power" of the test.

- In our example, the <u>Test</u> <u>Statistic</u> was the likelihood ratio and we rejected H_0 when this test statistic was $\leq C$

We could also say $X = \#$ of heads was the test statistic and reject if $X \geq X_0$ for some X_0.

- rejection region
  $= \{$ values of test statistic leading to rejection $\}$.

- acceptance region.

- The null distribution is the prob. dist of the test statistic when the null hypothesis is true.

In the Neyman-Pearson approach

? ① a Null hypothesis is singled out (How?)
a significance level α is chosen

The significance level determines a rejection region (How? there are

? ②. many possibilities)

To answer question ②. (in part).
assume for now that our Hypotheses H_0, H_1 are <u>simple</u>

i.e. they completely specify the distribution of the test statistic as in our example.

Any set of outcomes which has prob $\leq \alpha$ under H_0 can be the rejection region for a test at significance level α.

If the Hypotheses are simple, then we can choose a rejection region to maximize the <u>power</u> of the test.

Neyman - Pearson Lemma:

Suppose H_0, H_1 are simple hypotheses and that the test that rejects H_0 when the likelihood ratio is less than c., has significance level α.

Then any other test for which the significance level is less than or equal to α has power less than or equal to the power of the likelihood ratio test.

remarks:

- C is chosen so that

$$\text{Prob}\left(\frac{P(x \mid H_0)}{P(x \mid H_1)} < c \,\middle|\, H_0 \right) \leq \alpha$$

- recall that the power is the probability of rejecting H_0 when it is false.

How do we single out a Null hypothesis?.
(not assuming simplicity of hypotheses)
rules of-thumb

- the null hypotheses should be "simpler" mathematically
  e.g H_0 = the data are Poisson
  H_1 = the data are not Poisson.

  H_0 only depends on one distribution

- The prob. of false rejection of H_0 is controlled by choosing α

  so, tend to choose H_0 to be the hypothesis for which false rejection carries the more severe ~~greater~~ consequences.

e.g. A: The new drug is an improvement or equal to the currently used ~~previous~~ treatment.

B: The new drug is less helpful than the currently used treatment.

If A is true ~~and falsely~~ but gets rejected no harm is caused.

If B is true but gets rejected, patients will receive a less effective treatment.

- the null hypotheses should be simpler scientifically

  e.g. ESP experiments

  A: The subject is guessing

  B: The subject has ESP.

  The null hypotheses is A.

---

e.g. $X_1, \dots X_n$ i.i.d. $N(?, \sigma^2)$

$H_0 : \mu = \mu_0$

$H_A : \mu = \mu_1$

significance level α is given.

By Neyman-Pearson the most powerful test possible with level α is based on the likelihood ratio.

$$\frac{f_0(\vec{X})}{f_1(\vec{X})} = \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu_0)^2}}{e^{-1/2\sigma^2\sum_{i=1}^{n}(X_i-\mu_1)^2}}$$

and rejects for small values of this ratio

The ratio will be small when

$$\sum_{i=1}^{n} (X_i-\mu_0)^2 - \sum_{i=1}^{n} (X_i-\mu_1)^2 \quad \text{is large.}$$

$$= \sum_{i=1}^{n} X_i^2 - 2X_i\mu_0 + \mu_0^2 - \sum (X_i^2 - 2X_i\mu_1 + \mu_1^2)$$

$$= 2\sum_{i=1}^{n} X_i(\mu_1 - \mu_0) + n(\mu_0^2 - \mu_1^2)$$

$$= 2n(\mu_1-\mu_0)\bar{X} + n(\mu_0^2-\mu_1^2)$$

If $\mu_1 < \mu_0$, the ratio is ~~large~~ small when $\bar{X}$ is small.

If $\mu_1 > \mu_0$. the ratio is small when $\bar{X}$ is large.

$\boxed{\text{Suppose } \mu_1 > \mu_0}$.

Then Neyman-Pearson says that the most powerful test rejects for large values of $\bar{X}$ i.e.

We can choose c s.t.

$$\text{Prob}\{\bar{X} > c \mid H_0\} = \alpha$$

then the rejection region $\bar{X} > c$ gives the most powerful test at level α.

The null distribution of $\bar{X}$ is

$$N(\mu_0, \sigma^2/n)$$

so

$$\text{Prob}(\bar{X} > c) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right)$$

and we choose c s.t.

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{c-\mu_0}{\sigma/\sqrt{n}}} e^{-x^2/2} dx = 1 - \alpha.$$

p-values

Given data X and a test determined by a likelihood ratio
The p-value is the smallest significance level at which the Null Hypothesis would be rejected.

Think of it as the probability under the null hypothesis of a result as or more extreme than the one actually observed.

The smaller the p-value the stronger the evidence against the Null hypothesis.

The p-value is not the same as the posterior probability that the Null hypotheses is true. (The latter depends on the prior distributions of H_0 and H_A)

e.g. In our original coin tossing example

H_0: $p = 1/2$

H_1: $p = .7$

A highest power test with significance level $\approx .05$
(.0546875)

rejects for $X = 8, 9, 10$

If 8 heads are observed the p value is .054687
9 heads are observed the p value is .0107.