STAT 135 Lab 6
Duality of Hypothesis Testing and Confidence Intervals, GLRT, Pearson χ^2 Tests and Q-Q plots

Rebecca Barter

March 8, 2015

The duality between CI and hypothesis testing

The duality between CI and hypothesis testing

Recall that the acceptance region was the range of values of our test statistic for which H_0 will not be rejected at a certain level α.

For example, suppose that we have a test that rejects the null hypothesis when our sample $X_1, ..., X_n$ (where $X_i \sim F_\theta$) satisfies

$$\sum_{i=1}^{n} X_i < 1.2$$

Then our **acceptance region**, $A(\theta_0)$, is the set of values of our test statistic (which in this case is $\sum_{i=1}^{n} X_i$) that would not lead to a rejection of $H_0 : \theta = \theta_0$, i.e.

$$A(\theta_0) = \left\{ \sum_{i=1}^{n} X_i \middle| \sum_{i=1}^{n} X_i \geq 1.2 \right\}$$

The duality between CI and hypothesis testing

Assume that the true value of θ is actually θ_0 (i.e. that H_0 is true).

Then a $100 \times (1-\alpha)$ confidence interval corresponds to the values of θ for which the null hypothesis $H_0 : \theta = \theta_0$ is not rejected at significance level α.

The CI sounds an awful lot like the acceptance region, right? What's the difference?

- the **acceptance region** is the set of values of the *test statistic*, $T(X_1, ..., X_n)$, for which we would not reject H_0 at significance level α.
- the $100 \times (1-\alpha)\%$ **confidence interval** is the set of values of *the parameter* θ for which we would not reject H_0.

Exercise

Exercise: The duality between CI and hypothesis testing

Suppose that $X_1 = x_1, ..., X_n = x_1$, are such that $X_i \sim N(\mu, \sigma^2)$ with μ unknown and σ known. Show that the hypothesis test which tests

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

at significance level α corresponds to the confidence interval

$$\left[\bar{X} - z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right]$$

Generalized likelihood ratio tests

Generalized likelihood ratio tests

- Recall that the Neyman-Pearson lemma told us that when our hypotheses are simple

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta = \theta_1$$

  the likelihood ratio test has the highest power.
- Unfortunately there is no such theorem that tells us the optimal test when one of our hypotheses is composite:

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta > \theta_0 \quad , \quad H_1 : \theta < \theta_0 \quad , \quad H_1 : \theta \neq \theta_0$$

Generalized likelihood ratio tests

Suppose that observations $X_1, ..., X_n$ have a joint density $f(\mathbf{x}|\theta)$. If our hypothesis is composite e.g.

$$H_0 : \theta \leq 0 \qquad H_1 : \theta > 0$$

then we can use a (non-optimal) generalization of the likelihood ratio test where the **likelihood is evaluated at the value of θ that maximizes it**.

Generalized likelihood ratio tests

For our example ($H_0 : \theta \leq 0$), we consider a likelihood ratio test statistic of the form:

$$\Lambda = \frac{\max\limits_{\theta \leq 0}[lik(\theta)]}{\max\limits_{\theta}[lik(\theta)]}$$

where, for technical reasons, the denominator maximizes the likelihood over all possible values of θ rather than just those under H_1.

Generalized likelihood ratio tests

The generalized likelihood ratio test involves rejecting H_0 when

$$\Lambda = \frac{\max_{\theta \leq 0}[lik(\theta)]}{\max_{\theta}[lik(\theta)]} \leq \lambda_0$$

for some threshold λ_0, chosen so that

$$P(\Lambda \leq \lambda_0 | H_0) = \alpha$$

where α is the desired significance level of the test.

In order to determine λ_0, however, we need to know the distribution of our test statistic, Λ, under the null hypothesis.

Generalized likelihood ratio tests

It turns out that (under smoothness conditions on $f(\mathbf{x}|\theta)$):

Assuming that H_0 is true, then asymptotically

$$\boxed{-2\log(\Lambda) \sim \chi^2_{df}}$$

$$df = \#\{\text{ overall free parameters }\} - \#\{\text{ free parameters under } H_0\}$$

Generalized likelihood ratio tests

$$\boxed{-2\log(\Lambda) \sim \chi^2_{df}}$$

Thus, if $F_{\chi^2_k}$ is the CDF of a χ^2_k random variable,

$$\alpha = P(\Lambda \leq \lambda_0) = P(-2\log\Lambda \geq -2\log\lambda_0) = 1 - F_{\chi^2_{df}}(-2\log\lambda_0)$$

implying that we can find λ_0 by

$$\boxed{\lambda_0 = \exp\left(-\frac{F^{-1}_{\chi^2_{df}}(1-\alpha)}{2}\right)}$$

where $F^{-1}_{\chi^2_{df}}(1-\alpha)$ is the $1-\alpha$ quantile of a χ^2_{df} distribution

Generalized likelihood ratio tests: degrees of freedom

What's this degrees of freedom thing?

$df = \#\{$ overall free parameters $\} - \#\{$ free parameters under $H_0\}$

Suppose that we have a sample $X_1, ..., X_n$ such that $X_i \sim N(\mu, \sigma^2 = 8)$, and we want to test the hypothesis

$$H_0 : \mu < 0 \quad , \quad H_1 : \mu \geq 0$$

Then:

- Overall we have two parameters: μ and σ
- We have only **one free parameter overall** (we have specified $\sigma^2 = 8$, but made no assumptions on μ).
- We have **no free parameters under $\mathbf{H}_0$** (we have specified both $\mu < 0$ and $\sigma^2 = 8$).
- Thus $df = 1 - 0 = 1$.

Pearson's χ^2-test

Pearson's χ^2-test

Pearson's χ^2 test, often called the goodness-of-fit test, can be used to test the adequacy of a model to our data. For example, we can use this test to test the null hypothesis that the observed frequency distribution in a sample is consistent with a particular theoretical distribution.

The chi-squared test statistic, X^2, is given by

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{n} \frac{O_i^2}{E_i} - N$$

- O_i is an observed frequency
- E_i is the expected (theoretical frequency) under the null hypothesis
- n is the number of groups in the table
- N is the sum of the observed frequencies

Pearson's χ^2-test

Under the null hypothesis,

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-1-dim(\theta)}$$

where θ is our parameter of interest.

Thus our p-value is given by

$$p\text{-value} = P(\chi^2_{n-1-dim(\theta)} > X^2)$$

Exercise

Exercise: Pearson's χ^2-test (Rice, Chaper 9 exercise 42)

1. A student reported getting 9207 heads and 8743 tails in 17,950 coin tosses. Is this a significant discrepancy from the null hypothesis $H_0 : p = \frac{1}{2}$?
2. To save time, the student had tossed groups of five coins at a time and had recorded the results:

Number of Heads	Frequency
0	100
1	524
2	1080
3	1126
4	655
5	105

   Are the data consistent with the hypothesis that all the coins were fair ($p = \frac{1}{2}$)?

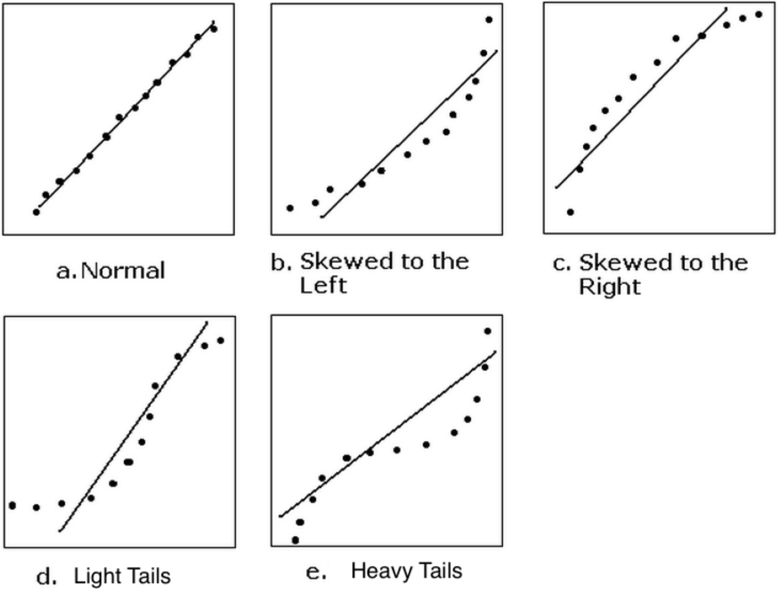
Q-Q plots

Normal Q-Q plots

Normal Quantile-Quantile (Q-Q) plots:

- Graphically compare the quantiles of observed data (reflected by the ordered observations) to the theoretical quantiles from the normal distribution (can also do this for any other distribution).
- If the resultant plot looks like a straight line, then this implies that the observed data comes from a normal distribution.
- If the resultant plot does not look like a straight line, you could use it to figure out if the data comes from a distribution with heavier or lighter tails, or is skewed etc

Normal Q-Q plots



Exercise

Exercise: normal Q-Q plots

Simulate 500 observations from the following distributions, and plot Q-Q plots

- t_2
- $Exponential(5)$
- $Normal(2, 3^2)$

Plot Q-Q plots and discuss the properties of these distributions.