STA261: Lecture 9

Likelihood Ratio Tests

Alex Stringer

July 30th, 2018

Disclaimer

The materials in these slides are intended to be a companion to the course textbook, *Mathematical Statistics and Data Analysis, Third Edition*, by John A Rice. Material in the slides may or may not be taken directly from this source. These slides were organized and typeset by Alex Stringer.

A big thanks to Jerry Brunner as well for providing inspiration for assignment questions.

License

Likelihood

"... We must return to the actual fact that one value of p, of the frequency of which we know nothing, would yield the observed result three times as frequently as would another value of p. If we need a word to characterise this relative property of different values of p, I suggest that may speak without confusion of the likelihood of one value of p being thrice the likelihood of another, bearing always in mind that likelihood is not here used loosely as a synonym of probability, but simply to express the relative frequencies with which such values of the hypothetical quantity p would in fact yield the observed sample"

- R.A. Fisher, *On the Mathematical Foundations of Theoretical Statistics*, 1921

Likelihood Ratios

If θ_0 is a candidate value for θ and $\hat{\theta}$ is the MLE, then $\hat{\theta}$ would generate the observed data Λ times more frequently than would θ_0, sort of.

If θ_0 actually is the true value of θ - the value that generated the observed data - then $-2\log\Lambda \dot{\sim} \chi^2_{p-d}$.

$-2\log\Lambda$ is a statistic, and if $\theta = \theta_0$, then we have found its *sampling distribution*.

We can use this to ask and answer the following question: *in repeated sampling, what is the probability that we would observe a likelihood ratio statistic as or more extreme as what we observed in our sample, if $\theta = \theta_0$?*

Likelihood Ratios

In the exponential example, $\Lambda = 0.915$, $-2\log\Lambda = 0.178$.

If $\theta = 5$, then

$$\begin{aligned} P(-2\log\Lambda > 0.178) &= P(\chi_1^2 > 0.178 \\ &= 0.67 \end{aligned}$$

It is pretty probable that we would observe what we observed (an average wait time of 6 minutes) in a sample of size $n = 5$ waits for the bus, if the true average wait time were actually 5 minutes.

(Aside: $n = 5$ is probably way too small for the χ^2 approximation to be any good, but that's another story).

Likelihood Ratios

What if we had taken a sample of size $n = 30$ (wait for the bus every day for a month) and observed an average wait time of $\bar{x} = 6$ minutes?

The likelihood ratio would be $\Lambda = 0.5884$, and $-2 \log \Lambda = 1.0607$.

$$P(-2 \log \Lambda > 1.0607) = 0.3031$$

This would be less probable. Observing a 6 minute average wait time all month gives stronger *evidence* against the notion that the average wait time is actually 5 minutes than does observing an average wait time of 6 minutes for just one week.

Hypothesis Tests

Putting formal terminology on this procedure gives us our first **Hypothesis Test**.

We call $H_0 : \theta = 5$ the **null hypothesis**.

$H_1 : \theta \neq 5$ is the **alternative hypothesis**.

$-2 \log \Lambda$ is called the **test statistic**, and χ^2_1 is its **null distribution**. Note we only need to know its distribution when H_0 is *true* in order for this procedure to work.

$P(-2 \log \Lambda(\mathbf{X}) > -2 \log \Lambda(\mathbf{x})$ is called the **p-value**, and quantifies the evidence the observed data gives against H_0.

Example

Suppose we throw a coin $n = 10$ times and observe 7 heads. Test the hypothesis that the coin is fair.

What about 100 throws and 70 heads?

Size of the Test

We can always compute a test statistic and a p-value if we have the mathematical tools to do so.

If a hard decision is required ("is the coin fair or not?") then we need to set a cutoff, i.e. a maximum p-value for which we would reject H_0.

We call this value α the **significance level** or the **size** of the test.

The decision rule is: reject H_0 if $p_0 < \alpha$.

Types of Errors

If we're making a decision, we could be wrong. How we could be wrong depends on the underlying unknown true value of θ.

A **Type I** error is to reject H_0 if in fact it is actually true.

A **Type II** error is to fail to reject H_0 if it is false.

Proposition: if H_0 is true, $P(\text{Type I Error}) = \alpha$.

Proof: ...

Example

Suppose $X_i \overset{IID}{\sim} N(\mu, \sigma^2)$. We take a sample of size $n = 100$ and get $(\bar{x}, s^2) = (2.1, 1.1)$. Test the hypothesis that $\mu = 0$ at the 5% level.

"At the 5% level" means $\alpha = 0.05$.

...

Categorical Data Analysis

Let's apply this framework to a more complicated example. Tim Hortons claims that each of their cup sizes has an equal chance of winning in Roll Up The Rim. Suppose I go to Tims $N = 67$ times during the competition, I buy $(S, M, L) = (9, 41, 17)$ coffees of each size, and I win $(1, 14, 2)$ times.

Can we develop a hypothesis test to test whether the probability of winning is the same for each cup size?

...