Assuming H_0 is true, there is a fundamental result by Samuel S. Wilks: As the sample size n approaches ∞, the test statistic $-2\log(\lambda)$ asymptotically will be chi-squared distributed (χ^2) with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0.[13] This implies that for a great variety of hypotheses, we can calculate the likelihood ratio λ for the data and then compare $-2\log(\lambda)$ to the χ^2 value corresponding to a desired statistical significance as an *approximate* statistical test. Other extensions exist.[which?]

STA 138 Discussion 6 – solutions

Fall 2020

Likelihood ratio tests for multinomial models

For our discussion this week, we will implement likelihood ratio tests. To do so, we'll pick up where we left off last week.

Suppose that Amelia has sampled 76 newts out of a tank. There are four species in the tank (A, B, C, and D); the observed counts are given in the table below.

A	B	C	D
14	22	25	15

Beatrice, the lab assistant, feeds the newts in the tank regularly. She takes particular notice when she does so of the brightly colored species A and B. From her experience, she claims that 20% of the newts in the tank are from species A, and 30% from species B.

Assuming that she is correct, use numerical methods to obtain the MLE of the proportions of the species in the tank.

In the problems below, we will interpret the test results individually; there's no need to control for multiple testing. We'll carry out each test at $\alpha = 0.01$.

1. Let's test whether we can rule out the possibility that the four species in the tank exist in equal proportions. Carry out a likelihood ratio test of

$$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.25$$

against

$$H_1 : \pi_i \neq \pi_j \text{ for some } i \neq j$$

Remember that the likelihood ratio test asks us to compare sets of possible parameters Θ_0 and Θ, where the former represents the null hypothesis. In this case, let's write

$$\Theta_0 = \{(\pi_1, \pi_2, \pi_3, \pi_4) : \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.25\}$$

and

$$\Theta = \left\{ (\pi_1, \pi_2, \pi_3, \pi_4) : 0 \leq \pi_1, \pi_2, \pi_3, \pi_4 \leq 1, \sum_{j=1}^{4} \pi_j = 1 \right\}$$

We can find without difficulty that the maximum likelihood under H_0 is

$$\arg\max_{\pi \in \Theta_0} L(\pi, y) = L(\pi_0, y) \approx 0.00017 \ .$$

Furthermore, we found last week that

$$\arg\max_{\pi \in \Theta} L(\pi, y) \approx 0.00160 \ .$$

unconstrained MLE: Part 1 in last discussion

4 parameters - 1 constraint = 3

Is the difference between these significant at $\alpha = 0.01$? From the above, we compute the LRT statistic to be 4.52953. Because Θ has $\dim(\Theta) = 3$ freely varying parameters, and Θ_0 has none, we use this in combination with a χ^2 distribution with $3 - 0 = 3$ degrees of freedom to obtain $p-$value 0.21.

So, at $\alpha = 0.01$, we fail to reject H_0.

2. Let's put Beatrice's claim to the test: using a likelihood ratio test, test

$$H_0 : \pi_1 = 0.2, \pi_2 = 0.3$$

against

$$H_1 : \pi_1 \neq 0.2 \text{ or } \pi_2 \neq 0.3$$

Here Θ_1 is as it was in the previous problem, while

$$\Theta_0 = \{(\pi_1, \pi_2, \pi_3, \pi_4) : \pi_1 = 0.2, \pi_2 = 0.3, 0 \leq \pi_3, \pi_4 \leq 1, \pi_3 + \pi_4 = 0.5\}$$

Since Θ_0 now has one freely varying parameter, we have $\dim(\Theta_0) = 1$.

Last week, we found that

$$\arg\max_{\pi \in \Theta_0} L(\pi, y) \approx 0.00143 \ .$$ constrained MLE: Part 1 in last discussion

This gives us LRT statistic 0.22466, and, from the χ^2_{3-1} null distribution, we get $p-$value 0.894. So, we fail to reject this this claim as well.

3. *Assuming that Beatrice's claim is true*, let's test whether there is evidence to rule out the possibility that species C and D occur in the tank in equal proportions.

$$H_0 : \pi_1 = 0.2, \pi_2 = 0.3, \pi_3 = \pi_4 = 0.25$$

against

$$H_1 : \pi_1 = 0.2 \text{ and } \pi_2 = 0.3 \text{ but either } \pi_3 \neq 0.25 \text{ or } \pi_4 \neq 0.25$$

Here, we have

$$\Theta = \{(\pi_1, \pi_2, \pi_3, \pi_4) : \pi_1 = 0.2, \pi_2 = 0.3, 0 \leq \pi_3, \pi_4 \leq 1, \pi_3 + \pi_4 = 0.5\}$$

and

$$\Theta_0 = \{(\pi_1, \pi_2, \pi_3, \pi_4) : \pi_1 = 0.2, 0.3, \pi_3 = \pi_4 = 0.25\}$$

This gives us LRT statistic 2.53074, and, from the χ^2_{1-0} null distribution, we get $p-$value 0.112. So, we fail to reject this this claim too!

Appendix: R Script

```r
## problem 1
y <-  c(14, 22, 25, 15)
lik0 <- dmultinom(y,prob=rep(0.25,4))
lik1 <- 0.00160
lrtstat <- -2*log(lik0/lik1)
pval <- 1-pchisq(lrtstat,3)
## problem 2
lik0 <- 0.00143
lrtstat <- -2*log(lik0/lik1)
pval <- 1-pchisq(lrtstat,2)
## problem 3
lik1 <- lik0
lik0 <- dmultinom(y,prob=c(0.2,0.3,0.25,0.25))
lrtstat <- -2*log(lik0/lik1)
pval <- 1-pchisq(lrtstat,1)
```