Contingency Tables

Exact Tests

Demetris Athienitis

Exact inference

- The previous tests all assumed the sample size was large and that expected cell counts were big enough
- Need an approach that works for all sample sizes, *Fisher's exact test* provides us with one such option
- We will restrict attention to 2 x 2 tables, but it has been extended to bigger tables

Fisher's exact test

		Y		
		1	2	
X	1	n_{11}	n_{12}	n_{1+}
	2	n_{21}	n_{22}	n_{2+}
		n_{+1}	n_{+2}	n

H_0 : X, Y independent $\Leftrightarrow \theta = 1$ (odds ratio =1)

- If we assume that the row totals and column totals are fixed, then n_{ij} follows a hypergeometric distribution
- In other words, the exact null distribution of $\{n_{ij} | n_{1+}, n_{2+}, n_{+1}, n_{+2}\}$ is the *hypergeometric distribution*
- Once we know n_{ij} then we know all other cell counts, since we know the margin totals

Fisher's exact test

The conditional distribution of n_{11} is as follows

$$p(n_{11}) = \frac{\binom{n_{1+}}{n_{11}}\binom{n_{2+}}{n_{+1} - n_{11}}}{\binom{n}{n_{+1}}}$$

where $n_{11} \in \{\max(0, n_{+1} + n_{1+} - n), \ldots, \min(n_{+1}, n_{1+})\}$

- Since the margin totals are fixed, this expresses the probability for all 4 cells in the table
- The p-value is the sum of all hypergeometric probabilities corresponding to values of n_{11} that are as least as favorable to H_1

Fisher's exact test

Example

A lady claims to be able to tell whether milk or tea is poured first and she is told that 4 of the 8 glasses have milk poured first

		Guess		
		Milk	Tea	
Poured	Milk	3	1	4
	Tea	1	3	4
		4	4	8

- $H_1 : \theta > 1$, she is able to do better than random guessing
- In this case, n_{11} could have taken values $\{0, 1, 2, 3, 4\}$

Example (continued)

```
> cbind(0:4,dhyper(0:4,4,4,4))
     [,1]       [,2]
[1,]    0 0.01428571
[2,]    1 0.22857143
[3,]    2 0.51428571
[4,]    3 0.22857143
[5,]    4 0.01428571
```

- p-value is $p(3) + p(4) = 0.229 + 0.014 = 0.243$
- $H_0 : \theta = 1$, or that her guess is independent of the actual order and to test $H_1 : \theta \neq 1$ we would have used

$$p(0) + p(1) + p(3) + p(4) = 0.486$$

Remarks

In R you can use `fisher.test`

Remark

- The test can be conservative in many cases especially when sample size is small , total probability of 1 can only be split over a small number of scenarios and singularities will hold a lot of probability
- Confidence interval covers the true parameter more often than the chosen level

We learned

Can use Fisher's exact test that does not require aymptotic normality.

Being *exact* it can be used for small AND large sample sizes.