Fisher's Exact Test

Max Turgeon

STAT 3150—Statistical Computing

Lecture Objectives

- Use Fisher's Exact Tests to test for independence of two categorical variables.
- Understand the connection between Fisher's Exact Test and permutation tests.

Motivation i

- *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* by David Salsburg
- A lady in England claimed she could always tell whether milk or tea was poured in a cup first.
- Fisher designed an experiment to test this claim:
    - He randomly filled four cups with tea first, and four cups with milk first.
    - He then asked the lady to guess which four cups were filled with milk first.

Motivation ii

- Here are the results, summarised in a table:

	Guess Milk	Guess Tea
Actual Milk	3	1
Actual Tea	1	3

Motivation iii

- Fisher's null hypothesis was that the lady was guessing randomly.
- Under this null hypothesis, the value in the top left cell (i.e. correct milk guesses) follows a *hypergeometric distribution*.
    - We can therefore compute the probability she correctly guessed 3.

```r
dhyper(3, m = 4, n = 4, k = 4)
```

```
## [1] 0.2285714
```

Motivation iv

- What would be a more extreme value?
  - If she had guessed 4 correctly!
- The p-value is the sum of the probability mass function at $x = 3$ and $x = 4$.

```
dhyper(3, m = 4, n = 4, k = 4) +
  dhyper(4, m = 4, n = 4, k = 4)
```

```
## [1] 0.2428571
```

- Given that probability, Fisher concluded there was a lack of evidence that the lady could really tell which was poured first.

Fisher's Exact Test i

- Fisher's exact test is a test of independence between two factors
    - E.g. the true liquid poured first and the lady's guess. - It relies on an important assumption:
    - Both the **row totals** and the **column totals** should be *fixed* by design.
- This was the case in Fisher's experiment: he poured 4 cups of each, and told the lady, so each row and each column in the table above has to sum to 4.

Fisher's Exact Test ii

- This assumption is key:
  - It implies that the value of a single cell completely determines the value of the other three cells.
- Therefore, any of the four cells is a valid test statistic.
- E.g. if we pick the lower left cell (i.e. wrong milk guesses), we get the same p-value (but note the difference in which values are more extreme!):

```
dhyper(1, m = 4, n = 4, k = 4) +
  dhyper(0, m = 4, n = 4, k = 4)
```

```
## [1] 0.2428571
```

Permutation tests

- **What is the connection with permutation tests?**
- Under the null hypothesis, the value of any cell follows a hypergeometric distribution.
- We can quantify the likelihood that a permutation would lead to a valid 2x2 table (i.e. with the correct row and column sums).
- As we permute the data, the frequency of each configuration converges to the hypergeometric probabilities.

Example i

```r
data <- rep(c("Milk", "Tea"), each = 4)
data
```

```
## [1] "Milk" "Milk" "Milk" "Milk" "Tea" "Tea"
"Tea" "Tea"
```

Example ii

```r
results <- replicate(1000, {
  data_perm <- sample(data)
  # Treat first 4 values as if the lady
  # had guessed milk
  sum(data_perm[1:4] == "Milk")
})

mean(c(3, results) >= 3)
```

```
## [1] 0.2347652
```

- As we can see, we get a similar p-value.

Summary

- Fisher's Exact Test is very much in the same spirit as the permutation tests.
- But computational resources were limited.
  - Fisher couldn't *actually* permute the data.
  - Instead, he had to compute the probability of each permutation.
- **Note**: you can construct a valid permutation test that doesn't rely on the assumption of fixed margins.