**Section 6.2 – Exact inference**

Most of the inference procedures examined so far rely on a statement similar to the following:

As the sample size goes to infinity, the statistic's distribution approaches a chi-square (or normal) distribution.

Infinity is the one sample size that we can never have! Thus, we have not used the EXACT distribution for our statistics.

Fortunately, there are many situations where distributions, like a chi-square, serve as a good approximation to the actual distribution for the statistic. However, especially when the sample size is small, these distributions often serve as poor approximations. Here’s a quote R. A. Fisher’s *Statistical Methods for Research Workers* 1st edition (1926) book on this topic:

… the traditional machinery of statistical processes is wholly unsuited to the needs of practical research. Not only does it take a cannon to shoot a sparrow, but it misses the sparrow! The elaborate mechanism built on the theory of infinitely large samples is not accurate enough for simple laboratory data. Only by systematically tackling small sample problems on their merits does it seem possible to apply accurate tests to practical data.

The purpose of this section is to develop inference procedures that do not rely on large-sample approximations. These inference procedures are often referred to as EXACT in the sense that they use the *actual* distribution for a statistic.

**Section 6.2.1 – Fisher's exact test for independence**

Hypergeometric distribution

Here’s the classic set up for a random variable with a hypergeometric probability distribution:

Suppose an urn has a red balls and b blue balls with n = a + b. There are k ≤ n balls drawn at random from the urn without replacement. Let M be the number of red balls drawn out.

The random variable M has a hypergeometric distribution with probability mass function of

 for m = 0, 1,…, k

subject to m ≤ a and k – m ≤ b. Note that a, n, b, and k are FIXED values. The only random variable is M!

An important part of this set up is that each ball has an equal probability of being drawn out of the urn, regardless of color.

Example: Balls in an urn (Tea.R)

Suppose there are n = 8 balls in an urn with a = 4 of them red and b = 4 of them blue. Suppose k = 4 balls are drawn from the urn. What is the probability of observing M = 3 red balls?



The probability mass function is

|  |  |
| --- | --- |
| m | P(M = m) |
| 0 | 0.0143 |
| 1 | 0.2286 |
| 2 | 0.5143 |
| 3 | 0.2286 |
| 4 | 0.0143 |

Is it reasonable to observe M ≥ 3?

R code and output:

> M <- 0:4

> # Syntax for dhyper(m, a, b, k)

> data.frame(M, prob = round(dhyper(M, 4, 4, 4), 4))

 M prob

1 0 0.0143

2 1 0.2286

3 2 0.5143

4 3 0.2286

5 4 0.0143

Fisher’s exact test

The hypergeometric distribution can be used to find the probability of observing a particular contingency table under independence! Below are tables demonstrating how:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Response |  |
|  |  | 1 | 2 |  |
| Group | 1 | w1 | n1 – w1 | n1 |
| 2 | w2 | n2 – w2 | n2 |
|  |  | w+ | n – w+ | n |
|  |  |  |  |  |
|  |  | Urn |  |
|  |  | Drawn out | Remaining |  |
| Color | Red | m | a – m | a |
| Blue | k – m | b – k + m | b |
|  |  | k | n – k | n |

With the binomial distribution set-up for contingency tables that was given in Chapter 2, we only had n1 and n2 (row totals) as fixed prior to taking a sample. To use the hypergeometric distribution, we also need to assume W+ and n – W+ (column totals) as fixed. This can be true for some situations, but it will not be for most. Fortunately, one can condition on W+ in the binomial model to obtain the hypergeometric model. The mathematical details are discussed in an exercise of the book for this section.

Question: If n1, n2, and w+ are known, how many cell values in the 2×2 portion of the table need to be known before all four values are known?

For the first table under independence, each element of groups 1 and 2 has the same probability of obtaining the 1 response. This is the same as each ball in the second table having the same probability of being drawn out of the urn, regardless of color.

Example: Lady tasting tea (Tea.R)

Ronald Fisher exploited the use of the hypergeometric distribution one day in Cambridge, England, in the late 1920s. During an afternoon tea with a group of colleagues, a “lady” in this group (Muriel Bristol) claimed that she could determine whether the milk or the tea was poured first into a cup. Fisher devised the follow experiment to test this lady’s claim:

1. Eight cups were set aside for the experiment.
2. Tea was poured first into 4 cups and milk was poured first into the remaining 4 cups.
3. The lady was blinded to which had tea or milk poured first, but she knew there were four of each.
4. The lady tasted the cups of tea and provided a tea or milk poured first response.

The television show Nova presents a re-enactment in their episode “Prediction by the Numbers” (Season 45, Episode 6); see <https://youtu.be/9OIel5NUG7Q?t=1346>. Salsburg (2001) gives a more thorough discussion in his book “The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century”.

Below is a **hypothetical** outcome of the experiment:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Lady’s response |  |
|  |  | Milk | Tea |  |
| Actual | Milk | 3 | 1 | 4 |
| Tea | 1 | 3 | 4 |
|  |  | 4 | 4 | 8 |

If the lady could not actually differentiate which was added to a cup first, then her selection of milk responses would in reality be four random cups, and the odds of a milk response would be the same regardless of whether milk or tea was truly added first (odds for row 1 equal to odds for row 2). Thus, we have independence.

Under independence, the probabilities for each possible contingency table are the same as we found earlier with the hypergeometric:

|  |  |  |
| --- | --- | --- |
| m | P(M = m) |  |
| 0 | 0.0143 | 0 |
| 1 | 0.2286 | 1/9 |
| 2 | 0.5143 | 1 |
| 3 | 0.2286 | 9 |
| 4 | 0.0143 | >9 |

This table could have been written in terms of w1 as well. I have added odds ratios to the table to help you see the amount of observed dependence.

P-values are probabilities an event is at least as extreme as what was observed. Under independence, the above table then provides information needed to compute a p-value for a test of independence! Thus, the probability of randomly guessing three or more of the milk cups correctly is P(M ≥ 3) = 0.2286 + 0.0143 = 0.2429.

Questions:

* Would you be surprised if the lady obtained 3 or more milk cups correct if she did not truly know the difference?
* Would you be surprised if the lady obtained 4 milk cups correct if she did not truly know the difference?
* What does OR > 1 mean here?
* What does OR < 1 mean here?
* What are H0 and Ha here for a test of independence?

The lady did actually get all 4 correct! Here’s how Fisher could have used R to perform the test:

> c.table <- array(data = c(4, 0, 0, 4), dim = c(2,2),

 dimnames = list(Actual = c("Milk", "Tea"), Response =

 c("Milk", "Tea")))

> c.table

 Response

Actual Milk Tea

 Milk 4 0

 Tea 0 4

> fisher.test(x = c.table, alternative = "greater")

 Fisher's Exact Test for Count Data

data: c.table

p-value = 0.01429

alternative hypothesis: true odds ratio is greater than 1

95 percent confidence interval:

 2.003768 Inf

sample estimates:

odds ratio

 Inf

> fisher.test(x = c.table)

 Fisher's Exact Test for Count Data

data: c.table

p-value = 0.02857

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 1.339059 Inf

sample estimates:

odds ratio

 Inf

The default for fisher.test() is to perform a two-sided test (alternative = "two.sided"). The p-value for a two-sided test is simply found by adding all probabilities that are less than or equal to the probability corresponding to what was observed. For this case, the probability was simply 0.0143 + 0.0143 = 0.0286.

Note: All exact inference procedures tend to be conservative for hypothesis testing.

Larger than 2×2 tables

Fisher’s exact test can be extended to tables larger than 2×2 by using the multiple hypergeometric distribution. This distribution is shown in my book, and the application of it is very similar to the 2×2 case.