Fisher's exact test

Observed data

	N	Y	
A	18	2	20
B	11	9	20
	29	11	40

- Assume the null hypothesis (independence) is true.
- Constrain the marginal counts to be as observed.
- What's the chance of getting this exact table?
- What's the chance of getting a table at least as "extreme"?

Hypergeometric distribution

- Imagine an urn with K white balls and N – K black balls.
- Draw n balls without replacement.
- Let x be the number of white balls in the sample.
- x follows a hypergeometric distribution (w/ parameters K, N, n).

	In urn		
	white	black	
sampled	x		n
not sampled			N – n
	K	N – K	N

Hypergeometric probabilities

Suppose X $\sim$ Hypergeometric (N, K, n).

No. of white balls in a sample of size n, drawn without replacement from an urn with K white and N – K black.

$$\Pr(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

Example:

	In urn		
	0	1	
sampled	18		20
not			20
	29	11	40

N = 40, K = 29, n = 20

$$\Pr(X = 18) = \frac{\binom{29}{18}\binom{40-29}{20-18}}{\binom{40}{20}} \approx 1.4\%$$

The hypergeometric in R

```
dhyper(x, m, n, k)

phyper(q, m, n, k)

qhyper(p, m, n, k)

rhyper(nn, m, n, k)
```

In R, things are set up so that

`m` = no. white balls in urn

`n` = no. black balls in urn

`k` = no. balls sampled (without replacement)

`x` = no. white balls in sample

`nn` = no. of observations

Back to Fisher's exact test

Observed data

	N	Y	
A	18	2	20
B	11	9	20
	29	11	40

- Assume the null hypothesis (independence) is true.
- Constrain the marginal counts to be as observed.
- Pr(observed table | H_0) = Pr(X=18)

  X $\sim$ Hypergeometric (N=40, K=29, n=20)

Fisher's exact test

1. For all possible tables (with the observed marginal counts), calculate the relevant hypergeometric probability.

2. Use that probability as a statistic.

3. P-value (for Fisher's exact test of independence):

   $\longrightarrow$ The sum of the probabilities for all tables having a probability equal to or smaller than that observed.

An illustration

The observed data

	N	Y	
A	18	2	20
B	11	9	20
	29	11	40

All possible tables (with these marginals):

20 0 9 11	$\rightarrow$ 0.00007	14 6 15 5	$\rightarrow$ 0.25994
19 1 10 10	$\rightarrow$ 0.00160	13 7 16 4	$\rightarrow$ 0.16246
18 2 11 9	$\rightarrow$ 0.01380	12 8 17 3	$\rightarrow$ 0.06212
17 3 12 8	$\rightarrow$ 0.06212	11 9 18 2	$\rightarrow$ 0.01380
16 4 13 7	$\rightarrow$ 0.16246	10 10 19 1	$\rightarrow$ 0.00160
15 5 14 6	$\rightarrow$ 0.25994	9 11 20 0	$\rightarrow$ 0.00007

Fisher's exact test: example 1

Observed data

	N	Y	
A	18	2	20
B	11	9	20
	29	11	40

P-value $\approx$ 3.1%

In R: `fisher.test()`

Recall:

$\longrightarrow$ χ^2 test: P-value = 1.3%

$\longrightarrow$ LRT: P-value = 1.1%

Fisher's exact test: example 2

Observed data

	I-B	NI-B	
I-A	9	9	18
NI-A	20	62	82
	29	71	100

P-value $\approx$ 4.4%

Recall:

$\longrightarrow$ χ^2 test: P-value = 3.0%

$\longrightarrow$ LRT: P-value = 3.7%

Fisher's exact test

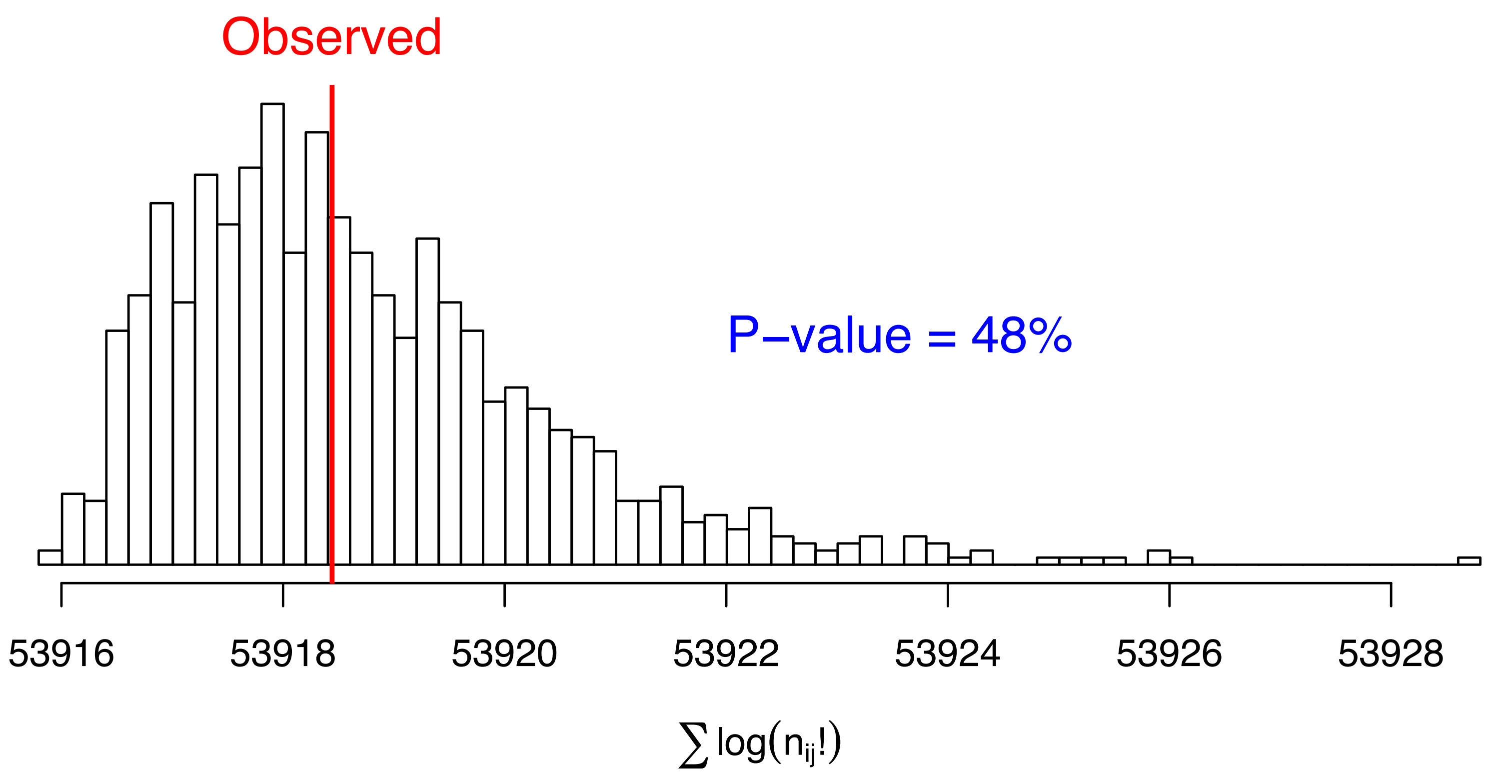
Observed data

	1	2	$\cdots$	k	
1	n_{11}	n_{12}	$\cdots$	n_{1k}	n_{1+}
2	n_{21}	n_{22}	$\cdots$	n_{2k}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
r	n_{r1}	n_{r2}	$\cdots$	n_{rk}	n_{r+}
	n_{+1}	n_{+2}	$\cdots$	n_{+k}	n

- Assume H_0 is true.
- Condition on the marginal counts
- Then Pr(table) $\propto \; 1/\prod_{ij} n_{ij}!$

- Consider all possible tables with the observed marginal counts
- Calculate Pr(table) for each possible table.
- P-value = the sum of the probabilities for all tables having a probability equal to or smaller than that observed.

Fisher's exact test: the example



⟶ Since the number of possible tables can be very large, we often must resort to computer simulation.