Hypothesis Testing
Tea Tasting (Fisher's Exact Test)

April 6, 2023
Ronald Fisher

Lady tea + milk : tea first or milk first

8 cups : 4 cups tea first
4 cups milk first

Two views

1. Null Hypothesis
2. She can tell the difference.

Some random guess
The probability of "all correct" is calculated as follows:

$$P_e("all\ correct") = \frac{\text{number of successful guesses}}{\text{total number of guesses}} = \frac{\binom{8}{4}}{\binom{12}{4}} \approx 0.014$$

The table and diagram illustrate the possible outcomes and their probabilities.
Truth

<table>
<thead>
<tr>
<th></th>
<th>Milk</th>
<th>Tea</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-k</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypergeometric distribution

\[ P(k) = \binom{4}{k} \binom{4}{4-k} \frac{8!}{(4^k)} = \frac{1}{70} \cdot \binom{4}{k}^2 \]

\[
\begin{array}{c|c|c|c|c|c|c}
   & 0 & 1 & 2 & 3 & 4 & 5 \\
\hline
P(k) & \frac{1}{70} & \frac{16}{49} & \frac{36}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} \\
\hline
P_r(k \geq k) & 1 & \frac{69}{70} & \frac{33}{70} & \frac{13}{70} & \frac{1}{70} & \frac{1}{70} \\
\hline
\end{array}
\]

\[ P_r(k \geq 2) = 0.241 \]

\[ P_r(k = 2) = 0.019 \]
The goal is to get a high score.

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(x=2) )</td>
<td>( \frac{2}{10} )</td>
<td>( \frac{1}{20} )</td>
<td>( \frac{7}{10} )</td>
<td>( \frac{2}{10} )</td>
<td>( \frac{2}{10} )</td>
<td>( \frac{1}{20} )</td>
<td>( \frac{1}{10} )</td>
</tr>
<tr>
<td>( P(x \geq 2) )</td>
<td>( \frac{1}{20} )</td>
<td>( \frac{1}{20} )</td>
<td>( \frac{1}{20} )</td>
<td>( \frac{1}{20} )</td>
<td>( \frac{3}{20} )</td>
<td>( \frac{1}{10} )</td>
<td></td>
</tr>
</tbody>
</table>

\[ P_1 \left[ x = 5 \right] = \frac{1}{20} \]

\[ P_1 \left[ x = 6 \right] = \frac{1}{10} \]
Summary of Hypothesis Test

1. Define Null Hypothesis

2. Assuming Null Hypothesis is true → determine probability of outcomes

3. Collect Data

4. Compute Probability of data outcome or something more extreme.
Quiz Review

Sample $X_1, \ldots, X_n$ iid $f(\theta) = \mathcal{N}(\mu, \sigma^2)$ R.V.

Statistic $T(X_1, \ldots, X_n)$ R.V.

E.g. $\alpha = 0.05 \equiv 95\%$

$1 - \alpha \times 100\%$ - confident interval

$[\bar{X}_n, R_0]$ so $\Pr(\bar{X}_n \leq \mu \leq R_0) = 1 - \alpha$

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

if we know $\sigma^2$

$Z_{\alpha/2} = \frac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1)$

if we do not know $\sigma^2$

$Z_{\alpha/2} = \frac{\bar{X}_n - \mu_0}{s / \sqrt{n}} \sim t_{\alpha/2, n-1}$
\[ S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

Sample Variance

\[ \sqrt{S_n^2} = S_n = \text{sample std. dev.} \]

\[ t\text{-distribution} \quad X_i \sim t(n-1) \]

2 degrees of freedom

\[ \tau_{1/2} = 1 - \alpha \quad \text{quantile of } N(0,1) \]

\[ t_{1/2} = 1 - \alpha \quad \text{quantile of } t\text{-distribution} \]
PDF $f_{\mu, \sigma^2}$ from $N(\mu, \sigma^2)$

$$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x - \mu)^2}{2 \sigma^2}}.$$