This time: Bayesian inference with \( \theta \in (0, 1) \)

Red: \( \Box \) ch 1-2
Appendices A, B;
Next time: ch 1-2

AMS 206
30 Jan 18

\( \{(\text{red}, \text{not red})\} \)

Sample of spins

Possible outcomes on a single spin:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

Sample size: \( N = 38 \)

Counts:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>62</td>
<td>50</td>
</tr>
<tr>
<td>0</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>57</td>
</tr>
<tr>
<td>1</td>
<td>63</td>
<td></td>
</tr>
</tbody>
</table>

Weak law of large numbers

\( \hat{\theta} = \frac{18}{38} = 0.474 \)
\[ \bar{X} \pm 1.96 \text{SE} \]  

95% confidence interval (CI) for \( \theta \)
Here \( \hat{\theta} = 0.18 \)

\[ \text{SE}(\hat{\theta}) = 0.019 \]

\[ \pm 1.96 (0.019) \]

\[ 0.18 \pm 1.96 (0.019) \approx 0.22 \]

\[ \hat{\theta} = \theta \]

\[ 100(1 - \alpha)\% \text{ CI:} \]

\[ \hat{\theta} \pm z(1 - \frac{\alpha}{2}) \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \]

\[ \hat{\theta} \pm z(1 - \frac{\alpha}{2}) \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \]

\( z(y) = \) standard normal \( \text{cdf} \) at \( y \)

\( z^{-1}(1 - \frac{\alpha}{2}) \) is quantile
\[ P \left( 0.142 \leq \theta \leq 0.218 \right) \approx 0.95 \]

Neyman: about 95% of the CIs will be hits

Your confidence is in the process of CI-drawing, not in the outcome.

Inference about population quantities (parameters) is typically not verifiable.
Fisher

maximum likelihood estimation

(\xi_1, \xi_2, \ldots, \xi_n)

\theta = \frac{1}{n} \sum_{i=1}^{n} \xi_i

\text{marginal distributions}

\mathbb{P}(\xi_i = 1) = \frac{1}{2}

\mathbb{P}(\xi_i = 0) = \frac{1}{2}

\text{write down joint distribution of data}

\prod_{i=1}^{n} \mathbb{P}(\xi_i = 1 | \theta) = \mathbb{P}(\xi_1 = 1, \xi_2 = 1, \ldots, \xi_n = 1 | \theta)
\[ P ( Y_1 = y_1, \ldots, Y_n = y_n \mid \theta) = \begin{cases} \theta & \text{if } y_i = 1 \\ 1 - \theta & \text{if } y_i = 0 \end{cases} \]

\[ = \theta y_1 (1 - \theta)^{1 - y_1} \]

\[ = \prod_{i=1}^{n} \theta y_i (1 - \theta)^{1 - y_i} \]

\[ = \theta y_1 (1 - \theta)^{1 - y_1} \theta y_2 (1 - \theta)^{1 - y_2} \cdots \theta y_n (1 - \theta)^{1 - y_n} \]

\[ = \theta^{y_1 + \ldots + y_n} (1 - \theta)^{n - (y_1 + \ldots + y_n)} \]
let's let \( S = \sum_{i=1}^{n} X_i \) then

\[
(\theta | \chi) = \theta^5 \left( 1 - \theta \right)^{n - 5} (0 \leq \theta \leq 1)
\]

Joint sampling dist \( \propto \) \( f(n) \) of \( y \) for fixed \( \theta \)

2. Define likelihood function as a positive constant multiple of

Joint sampling dist \( \propto \), but

thought of as a function of \( \theta \) for fixed \( y \):

\[
\ell(\theta | y) = c \cdot \prod_{i=1}^{n} \left( \frac{\theta}{\theta + \beta} \right) \]

Likelihood for \( \theta \) for given \( y \) = \( c \cdot \theta^5 \left( 1 - \theta \right)^{n-5} \)
def: if \( \ell(\theta; y) \) depends on \( y \) only through a summary of \( y \) (e.g., \( s = \sum y_i \)), then the summary is said to be (a sufficient statistic) for \( \theta \) with miss likelihood

\[ s \rightarrow \text{vector of length } n \begin{pmatrix} n \\ 1 \end{pmatrix} \]

\( n = 400 \)

\( s = 72 \)

\( \theta \sim (1 - \theta) \frac{400 - 72}{72} \)