This Time: probability
Next Time: prob. models for sums & means
Read: Do CB1 Ch. 9-11', LN pp L-113 - 136
Hwk @ due next Thr in class

\[
\begin{bmatrix}
1 \\
2 \\
3 \\
4 \\
5
\end{bmatrix}
\overset{\text{random}}{\longrightarrow}
\begin{bmatrix} y \end{bmatrix}_{n=1} \quad P(y \text{ is odd}) = \frac{2}{3} = 67\% = 0.67
\]

Equally likely model applies ✓

\[
\begin{bmatrix}
0 \\
1 \\
2 \\
3 \\
4 \\
5
\end{bmatrix}
\quad \text{y = # of their children with Tay-Sachs disease}
\quad \text{if ELM applies}
\quad P(1 \text{ or more } T-s) = \frac{5}{6} = 83\% = 0.83
\]

\text{but ELM does not apply}

\[\text{if ELM} \quad \text{but not so!}\]

\[\text{expected to be highest}\]

\[\text{prob}\]

\[\text{prob}\]

\[\text{not A} \quad \text{not A}\]

therefore:
\[P(A) + P(A^c) = 1 = 100\%\]

\[P(A) = 1 - P(A^c)\]
addition rule, special case:
\[ P(A \cup B) = P(A) + P(B) \]

general addition rule for OR
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[
\begin{bmatrix}
1 \\
2 \\
q
\end{bmatrix}
\overset{\text{at random}}{\rightarrow}
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}_{m=2}
\]

\[ P(y_1 = q \text{ and } y_2 = q) = ? \]

at random \hspace{1cm} \text{with replacement} \hspace{1cm} \text{independent identically distributed (IID)}
\hspace{1cm} \text{without replacement} \hspace{1cm} \text{simple random sampling (SRS)}

\text{case 1:}
\[
\begin{bmatrix}
1 \\
2 \\
q
\end{bmatrix}
\overset{\text{IID}}{\rightarrow}
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}
\]

\[ P(y_1 = q \text{ and } y_2 = q) = ? \]

\[
\begin{array}{ccc}
1 & 2 & q \\
(1,1) & (1,2) & (1,q) \\
(2,1) & (2,2) & (2,q) \\
(q,1) & (q,2) & (q,q)
\end{array}
\]

ELM? \text{yes}
\[ P(\begin{array}{c} y_1 = q \\ y_2 = q \end{array}) = \frac{3}{3} \]

\[ P(y_1 = q) = P(A) = \frac{1}{3} = \frac{3}{3} \]

\[ P(y_2 = q) = P(B) = \frac{1}{3} = \frac{3}{3} \]

our conjecture: \[ P(A \cap B) = P(A) \cdot P(B) \]
\[ \frac{1}{3} = \frac{1}{3} \cdot \frac{1}{3} \]

\[
\begin{bmatrix}
1 \\
2 \\
q
\end{bmatrix}
\overset{\text{SRS}}{\rightarrow}
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}
\]

\[ P(y_1 = q \text{ and } y_2 = q) = 0 \]

\[
\begin{array}{ccc}
1 & 2 & q \\
(1,1) & (1,2) & (1,q) \\
(2,1) & (2,2) & (2,q) \\
(q,1) & (q,2) & (q,q)
\end{array}
\]

ELM? \text{yes}
\[ P(y_1 = q) = \frac{1}{3} = \frac{3}{6} \]

\[ P(y_2 = q) = \frac{1}{3} = \frac{3}{6} \]

by our conj.: \[ P_{\text{srs}}(y_1 = q \text{ and } y_2 = q) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \]

but actually it's 0 = \% 

our conj. fails for SRS
Conditional Probability

Bayes 1760 → Bayesians

\[ P(B|A) = ? \]

"given"

ex. \( P(y_2 = 9 \mid y_1 = 2) = ? \)

\[
P(B) = \frac{\text{area of } B}{\text{total area}} = \frac{\square}{\square}
\]

\[
P(B|A) = \frac{\text{overlap } B \text{ and } A}{A}
\]

\[ P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \quad \left(\text{provided } P(A) > 0; \quad P(B|A) \text{ is undefined if } P(A) = 0\right)
\]

\[ P(A \text{ and } B) = P(A) \cdot P(B|A) \]

general product rule for and

\[ P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \]

\[ A, B \text{ independent if information about } A \text{ does not change your predictions about } B, \text{ & vice versa.} \]

Product Rule, IID Sampling: \( P(A \text{ and } B) = P(A) \cdot P(B|A) \)

\[
\begin{align*}
\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \\
(y_1 = 2) \quad (y_2 = 9)
\end{align*}
\]

if \( A, B \) are independent, then \( P(A \text{ and } B) = P(A) \cdot P(B|A) \)

but \( P(B|A) = P(B) \)

so if \( A, B \) indep: \( P(A \text{ and } B) = P(A) \cdot P(B) \)
\[ P(1 \text{ or more } T-S \text{ in family of } 5) = \]
\[ = 1 - P(0 \text{ T-S babies}) \]
\[ = 1 - P(\text{not T-S on } 1^{st} \text{ baby and not T-S on } 2^{nd} \text{ and ... and not T-S on } 5^{th}) \]
\[ \text{independent} = 1 - P(\text{not T-S on } 1^{st}) \cdot P(\text{not T-S on } 2^{nd}) \cdot ... \cdot P(\text{not T-S on } 5^{th}) \]
\[ = 1 - (1 - \frac{1}{4}) \cdot (1 - \frac{1}{4}) \cdot ... \cdot (1 - \frac{1}{4}) \]
\[ = 1 - (\frac{3}{4})^5 = 0.76 = 76\% \]

Marijuana Legalization Preference

<table>
<thead>
<tr>
<th>MLP</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>m</td>
</tr>
<tr>
<td>N</td>
<td>f</td>
</tr>
</tbody>
</table>

\[ n = 106 \]

sorted into

<table>
<thead>
<tr>
<th>Y</th>
<th>m</th>
<th>m</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>f</th>
<th>f</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>m</th>
<th>m</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ n = 5 \]

\[ 20 \]

I now for each UCLA stat student in my class in 1992 who participated.

Q: Are gender & MLP independent in this data set?

\[ \begin{array}{c|cc}
\text{MLP} & Y & N \\
\hline
\text{m} & 29 & 26 \\
\text{f} & 52 & 57 \\
\hline
\end{array} \]

\[ \frac{81}{106} = 76\% \]

Choose a student at random...

\[ P(Y|m) = \frac{52}{57} = 91\% \]

\[ P(Y|f) = \frac{29}{49} = 59\% \]

A: No, they're strongly dependent. 91 & 59 are not close to 76

(\text{and} x \text{ or } + \text{ or } \text{not given})

Simpson's paradox discussed.