Back to T-S example

<table>
<thead>
<tr>
<th># T-S babies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

E.L.M. applied to list of possible outcomes gets 5 out of 6 as 83.3%

ELM does not apply bc chance of having 5 is less likely than 0.

\[ P(A \text{ or } B) \quad P(A) \quad P(B) \]

\[ A = \{1 \text{ or more T-S babies}\} = \text{not } \{\text{exactly 0 babies}\} \]

If have 0 T-S babies, baby1 healthy \underline{and} baby2 \underline{and} ... baby6.

<table>
<thead>
<tr>
<th>( P(A \text{ or } B) )</th>
<th>( P(A) )</th>
<th>( P(B) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(A) )</td>
<td>( P(\text{not } A) )</td>
<td>( P(B) )</td>
</tr>
<tr>
<td>( P(A \text{ and } B) )</td>
<td>( P(A) )</td>
<td>( P(B) )</td>
</tr>
</tbody>
</table>
Venn Diagrams. Visual for the ELM

\[ P(A) = \frac{\text{Area } A}{\text{Area } \Box} \]

Total area of box = 1 = 100%

For any event A, 0% = 0 ≤ P(A) ≤ 1 = 100%.

\[ P(A) + P(\text{not } A) = 1 \quad (= 100\%) \]

\[ P(A) = 1 - P(\text{not } A) \]

Summary of Prob. Rules:

1. The easy rule

2. Addition Rule working with or
Now let's look at **OR** problems.

1. \( P(A \text{ or } B) = P(A) + P(B) \)
   - Special case when no overlap

2. \( P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \)
   - General rule: addition for **OR**

If \((A \text{ and } B)\) can't happen, \(A + B\) are mutually exclusive.

AND

Now let's look at **AND** problems.

Picking from 1, 2, 9 cards, pick 1 = \( y_1 \), pick 2 = \( y_2 \).

### 1. **W/ Replacement**

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>1</th>
<th>2</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,1</td>
<td>1,2</td>
<td>1,9</td>
</tr>
<tr>
<td>2</td>
<td>2,1</td>
<td>2,2</td>
<td>2,9</td>
</tr>
<tr>
<td>9</td>
<td>9,1</td>
<td>9,2</td>
<td>9,9</td>
</tr>
</tbody>
</table>

Can we use ELM w/ this list of possibilities? Yes

\[
P(y_1 = 9 \text{ and } y_2 = 9) = \frac{1}{9}
\]

\[
P(y_1 = 9) = \frac{1}{3} \quad P(y_2 = 9) = \frac{1}{9}
\]

\[
P(y_1 = 9 \text{ and } y_2 = 9) = \frac{1}{9}
\]

\[
P(y_1 = 9) = \frac{1}{3} \quad P(y_2 = 9) = \frac{1}{3}
\]

**P(A \text{ and } B) = P(A) \cdot P(B)**
2. **W/O replacement**

Now impossible to get 1,1, 2,2 and 9,9.

The 6 remaining combos are equally likely, ELM applies.

\[
P(Y_1 = 9) = \frac{2}{6} = \frac{1}{3}, \quad \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \quad \text{BUT}
\]

\[
P(Y_2 = 9) = \frac{2}{6} = \frac{1}{3}, \quad 9,9 \text{ impossible, } = 0
\]

\[
\frac{1}{9} \neq 0
\]

**Conditional Probability**

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

\[
P(A) = \frac{1}{6} \quad \text{Bayes Definition}
\]

\[
(A \text{ given } B) = P(A \mid B) = \begin{cases} 
\frac{P(A \text{ and } B)}{P(B)} \text{ if } P(B) > 0 \\
\text{undefined} \quad P(B) = 0
\end{cases}
\]

Multiply by \(P(B)\)

**Product Rule for AND**

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

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

\[
P(A \text{ and } B) = P(A) \cdot P(B \mid A)
\]
Without replacement: \( P(Y_1 = a \text{ and } Y_2 = a) = \) 
\[
P(Y_1 = a) \cdot P(Y_2 = a | Y_1 = a) \] 
\[
= \frac{1}{3} \cdot 0 = 0 \] This is correct!

With replacement: \( P(Y_1 = a \text{ and } Y_2 = a) = \frac{1}{9} \)
\[
= P(Y_1 = a) \cdot P(Y_2 = a | Y_1 = a) 
= \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \]

\( Y_1 \) and \( Y_2 \) are independent (\( Y_1 \) does not affect \( Y_2 \) when \( w/ \) replacement)

\[A \text{ and } B \text{ are probabilistically independent} \]
\[\text{if information about } B \text{ doesn't change chances for } A \text{ and vice versa.}\]

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

Product rule when \( A, B \) indep,
\[P(A \text{ and } B) = P(A) \cdot P(B | A) = P(A) \cdot P(B) + \text{ vice versa}\]

At Random \( w/ \) Replacement \( \leftrightarrow \) Independent (\( \text{IID} \) Identically Distributed)
Gender + Legalization Preference

<table>
<thead>
<tr>
<th>Pref</th>
<th>gender</th>
<th>Y</th>
<th>M</th>
<th>n=100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

Preference: Variable
Qual, Nominal, dich
Gender: Variable
Qual, Nom, dich

Two-by-Two Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>gender</th>
<th>Yes</th>
<th>No</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y/F</td>
<td>20</td>
<td></td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>N/M</td>
<td>29</td>
<td></td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>Y/M</td>
<td>52</td>
<td></td>
<td></td>
<td>106</td>
</tr>
</tbody>
</table>

Q: Are gender + marijuana legal preference independent (probability) in this dataset?

\[
P(Y) \quad \begin{cases} \text{if all 3 probabilities are similar, gender + MLP are approx. independent} \\
p(Y|F) = \frac{51}{106} = 76\% \\
p(Y|M) = \frac{29}{49} = 59\% \\
p(Y|F) = \frac{52}{57} = 91\% \\
\end{cases}
\]

76% ≠ 59% ≠ 91% so variables (gender+MLP) are (strongly) dependent in this dataset.
This problem will help with the midterm.

Now that we know **and** and **or**, we can finish the T-S problem.

\[
P(A) = P(1 \text{ or more T-S in } 6) = 1 - P(\text{not } A) = 1 - P(\text{exactly } 0 \text{ T-S}) = 1 - \prod \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \left( \frac{1}{2} \right)
\]

Independent

\[
= 1 - \left[1 - P(1\text{st T-S}) \cdot P(2\text{nd T-S}) \cdots P(g^{th} T-S)\right] = 1 - \left(1 - \frac{1}{4}\right)^5 \approx 76.5%
\]

We can see that the probability of T-S increases with the number of kids.
K-5 Death Penalty Case Study

\[ P(DP) = \frac{90}{320} = 11\% \]

\[ P(DP|DW) = \frac{19}{160} = 11.9\% \quad \text{Surprising... more whites than blacks get death} \]

\[ P(DP|DB) = \frac{17}{160} = 10.2\% \quad \text{penalty.} \]

Potential PCFs: Race of victim

Given white victim vs black victim

\( Y \) (outcome) \( DP \) or not
\( X \) (tr+mnt) Race defendant W,B
Design Observational Study
\( Z \) (PCF) Race of victim

How to control PCF? Hold it constant, look at relationship bw Y and X separately for white and black victims.

\[ P(DP|VW) = \frac{30}{214} = 14.0\% \quad \text{Victim White,} \]

\[ P(DP|VW \text{ and } DW) = \frac{19}{161} = 12.6\% \]

\[ P(DP|VW, DB) = \frac{14}{63} = 17.6\% \]

\[ P(DP|VB) = \frac{62}{112} = 5.4\% \quad \text{Victim Black,} \]

\[ P(DP|VB, DW) = \frac{8}{8} = 0\% \]

\[ P(DP|VB, DB) = \frac{6}{83} = 5.8\% \]

This is a paradox (Simpson's Paradox) bc relationship bw \( X+Y \) goes in one direction, but is reversed when \( Z \) accounted for.
We get a clearer picture of the true effect of ethnicity of defendant on death penalty by controlling for ethnicity of victim than by not doing so.

Why did the Simpson's paradox arise?

a. victim usually knows killer
b. whites hang around whites (vice versa)
c. therefore whites mostly kill whites (vice versa)
d. if victim white, more likely to get death penalty
e. therefore it will look like more white def get DP than "they really are"