AMS 7 - Lecture 4.26.18

This Time: Conditional Probability
Next Time: Prop. Models for sums

* Read DD ch. 1-3 (A) & DD ch. 1-9 (B)
  - Today → IN pg. 95-118
  - Next Time → LN pg. 119-126

* What's the probability of A?
  \[ P(A) = \frac{\text{Area of } A}{\text{Area of } A + B} \approx 100\% \]

* What's the probability of A given B?
  \[ P(A | B) = \frac{\text{Overlap of } A \& B}{\text{Area of } B} \]

* Definition (A. de Moivre) - 1705
  \[ P(A | B) = \begin{cases} 
  \frac{P(A \text{ and } B)}{P(B)} & \text{if } P(B) > 0 \\
  \text{Undefined} & \end{cases} \]

R-37
  \[ P(A | B) = \frac{P(A \text{ and } B)}{P(B)} \rightarrow P(B) \cdot P(A | B) = P(A \text{ and } B) \]
  \[ P(B | A) = \frac{P(A \text{ and } B)}{P(A)} \rightarrow P(A) \cdot P(B | A) = P( A \text{ and } B) \]
  \[ P( Y_1 = 2 \text{ and } Y_2 = 2 ) \stackrel{\text{SRS}}{=} P( Y_1 = 2 ) \cdot P( Y_2 = 2 | Y_1 = 2 ) \]
  \[ = \frac{1}{3} \cdot 0 = 0 \]
\[ * P(y_1 = 2 \text{ and } y_2 = 2) \xrightarrow{\text{I.I.D.}} P(y_1 = 2) \cdot P(y_2 = 2 | y_1 = 2) \]
\[ = P(y_1 = 2) \cdot P(y_2 = 2) \]

**Definition**

Let \( A \) and \( B \) are independent if and only if information about \( A \) does not change chances of \( B \), and vice versa: \( P(A \& B) = P(A) \cdot P(B) \)

* Tay-Sachs Case Study*

\[ P(1 \text{ or more T-S babies in family of 5, both parents carriers}) \]
\[ = 1 - P(0 \text{ T-S babies}) \]
\[ = 1 - P(\text{1st baby Not T-S AND 2nd baby Not T-S AND ... AND 5th baby Not T-S}) \]
\[ \text{by independence:} \]
\[ = 1 - P(\text{Not T-S on 1st}) \cdot P(\text{Not T-S on 2nd}) \cdot \ldots \cdot P(\text{Not T-S on 5th}) \]
\[ = 1 - (1 - \frac{1}{4}) \cdot (1 - \frac{1}{4}) \cdot \ldots \cdot (1 - \frac{1}{4}) \]
\[ = 1 - (1 - \frac{1}{4})^5 = 0.760 = 76.0\% \]

* UCLA Marijuana Case Study*

\[
\begin{array}{c|c|c|c|c}
Y & X & \text{X = Predictor (Female or Male)} & \text{Female (1), Male (0)} & \text{Yes (1), No (0)} \\
0 & 0 & & & \\
1 & 0 & & & \\
\vdots & \vdots & & & \\
1 & 1 & & & \\
\end{array}
\]

\[
\begin{array}{c|c|c}
Y & X & \text{MLP} \\
0 & 0 & \text{5} \\
0 & 1 & \text{20} \\
1 & 0 & \text{52} \\
1 & 1 & \text{29} \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{2x2 contingency table} & \text{Female} & \text{Male} \\
\text{Yes} & 29 & 20 \\
\text{No} & 52 & 5 \\
\text{n} & 81 & 25 \\
\end{array}
\]

\[ \Delta \text{2x2 contingency table} \]
\[ \rightarrow \text{Categorical data analysis} \]

\[ \text{Female} \ 29 + 20 = 49 \]
\[ \text{Male} \ 52 + 5 = 57 \]
\[ n = 106 \]
P(Y) = 81/100 = 76%
P(Y|F) = 29/49 = 59.1%
P(Y|M) = 52/57 = 91%

Question: Are gender & MLP independent or dependent in this data set? DEPENDENT

* Gender & MLP are associated because

- 76%: Yes
- 59%: Male
- 91%: Female

huge difference in practical terms.

* Practically Significant (highly PractSig)

Death Penalty Case Study
Y (outcome): death penalty or not
X (treatment): white vs. black defendant

→ Basic design: observational study

- enemy: bias from PCF

L → Leading PCF

- Ethnicity of victim
- (white vs. black)

* How to defeat the PCF?

L → Hold it constant in relating X & Y

<table>
<thead>
<tr>
<th></th>
<th>Death Penalty</th>
<th>Defendant</th>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>DP</td>
<td>White</td>
<td>19/100</td>
</tr>
<tr>
<td>Black</td>
<td>DP</td>
<td>White</td>
<td>17/100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Black</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P(DP) = 36/320 = 11%
**Middle Table**

\[ P(DP|VW) = \frac{30}{214} \approx 14.0\% \]

White victim

\[ P(DP|DW \text{ AND } VW) = \frac{19}{151} \approx 12.6\% \]

\[ P(DP|DB \text{ AND } VW) = \frac{11}{103} \approx 17.5\% \]

**Bottom Table**

\[ P(DP|VB) = \frac{6}{112} \approx 5.4\% \]

Black victim

\[ P(DP|VB, DW) = \frac{0}{9} = 0\% \]

\[ P(DP|VB, DB) = \frac{6}{103} \approx 5.8\% \]

> Why did the Simpson's Paradox arise?

1. Victim usually knows killer
2. White people mostly hang around white people & black people with black, therefore, whites mostly kills whites & blacks kill blacks but if victim is white, much more likely to get death penalty, therefore, it will look like white defendants get DP more than "they really are"

---

Probability models for sums & means

* Roulette (38 possibilities) ELM \(\sqrt{\&} \gg \text{IID} \ll \)

\[ P(\text{Win on a single play, single #}) = \frac{1}{38} \approx 2.5\% \]

Split = 2/38 = 5%
Possible outcome on a single spin

Population

\[
\begin{pmatrix}
-1 \\
-1 \\
\vdots \\
+35 \\
-1 \\
\vdots \\
-1
\end{pmatrix}
\]

Sample observed spins

\[
\begin{pmatrix}
\text{mostly -1} \\
\vdots \\
\text{occasionally +35}
\end{pmatrix}
\]

n = 1,000

Single # 6

\[
38 = \frac{37(-1) + (+35)}{38}
\]

\[\text{mean } M = \frac{-2}{38} = -0.05\]

\[\text{negative}
\]

\[\sigma = \text{sigma}
\]

\[\text{On average, I expect to win } M = -0.05 \text{ on each $1.00 bet for a single #}
\]