AMS 7
Lecture 6
7/1/16

Back to T-S example

T-S babies
0
1
2
3
4
5

~~E.L.M. applied to list of possible outcomes~~ $\cancel{\frac{5}{6}}$ = 83%

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

$P(A \boxed{\text{or}} B)$ $P(A)$ $P(B)$

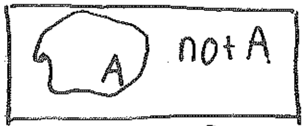
$A = \{1 \text{ or more T-S babies}\} = \boxed{\text{not}} \{\text{exactly 0 babie}$

If have 0 T-S babies, baby$_1$ healthy $\boxed{\text{and}}$ baby$_2$ $\boxed{\text{and}}$ ... baby$_5$.

$P(A \boxed{\text{or}} B)$	$P(A)$	$P(B)$
$P(A)$	$P(\text{not } A)$	
$P(A \boxed{\text{and}} B)$	$P(A)$	$P(B)$

Venn Diagrams

Visual for the ELM



$P(A) = \frac{A}{\boxed{\text{Area } 1}}$

total area of box $= 1 = 100\%$

$$\boxed{\begin{array}{l}\text{For any event } A,\ 0\% = 0 \le P(A) \le 1 = 100\%. \\ P(A) + P(\text{not } A) = 1 \quad (= 100\%)\end{array}}$$

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

Summary of Prob. Rules

1. The easy rule

2. Addition Rule working with or

OR

Now lets look at $\boxed{\text{or}}$ 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 $\boxed{\text{or}}$

If (A <u>and</u> B) can't happen, A + B are <u>mutually exclusive</u>

AND

Now lets look at $\boxed{\text{and}}$ problems

$$\underset{\text{pop}}{\begin{bmatrix}1\\2\\9\end{bmatrix}} \xrightarrow{\text{at random}} \underset{\text{sample}}{\begin{bmatrix}y_1\\y_2\end{bmatrix}}$$

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

1. <u>w/ Replacement</u>

		y_2		
		1	2	9
y_1	1	1,1	1,2	1,9
	2	2,1	2,2	2,9
	9	9,1	9,2	9,9

Can we use ELM w/ this list of 9 possibilities? yes

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

$P(y_1 = 9) = \frac{1}{3}$ $\qquad$ $P(y_2 = 9) = \frac{3}{9} = \frac{1}{3}$

$P(y_1 = 9 \text{ and } y_2 = 9) = \frac{1}{9}$
$P(y_1 = 9) = \frac{1}{3}$ $\quad$ $P(y_2 = 9) = \frac{1}{3}$

$$\boxed{P(A \text{ and } B) = P(A) \cdot P(B)}$$

2. w/o replacement

Now impossible to get 1,1 2,2 and 9,9. The 6 remaining combos are equally likely, ELM applies.

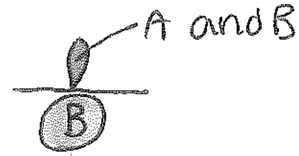
$P(Y_1 = 9) = \frac{2}{6} = \frac{1}{3}$ $\quad \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$ BUT

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

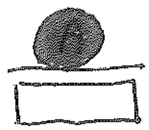
$\frac{1}{9} \neq 0$

Conditional Probability

P(A given B) =



P(A) =



Bayes Definition

(A given B) = $P(A \mid B)$ (means given) $= \begin{cases} \dfrac{P(A \text{ and } B)}{P(B)} & \text{if } P(B) > 0 \\ \text{undefined} & P(B) = 0 \end{cases}$

Multiply by P(B)

Product Rule for And

$\boxed{P(A \text{ and } B) = P(B) \cdot P(A \mid B)}$

$P(B \mid A) = \dfrac{P(B \text{ and } A)}{P(A)} = \dfrac{P(A \text{ and } B)}{P(A)}$ equiv.

$\cdot P(A)$

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

W/O replacement; $P(\underbrace{Y_1 = 9}_{A}$ and $\underbrace{Y_2 = 9}_{B}) =$

$P(Y_1 = 9) \cdot P(Y_2 = 9 \mid Y_1 = 9)$
$= \frac{1}{3} \cdot 0 = 0$ This is correct!

W/ replacement; $P((Y_1 = 9)$ and $(Y_2 = 9)) = \frac{1}{9}$
$= P(Y_1 = 9) \cdot P(Y_2 = 9 \mid Y_1 = 9)$
$P(Y_2 = 9)$

$1/3$ $1/3$ $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$

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

A and B are probabilistically <u>independent</u> if information about B doesn't change chances for A and vice versa.

$P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$

Product Rule when A, B indep,

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

+ vice versa

At Random W/ Replacement $\longleftrightarrow$ Independent (IID) Identically Distributed

R-39 Gender + Legalization Preference

Pref gender   1 row for each student

$$\begin{bmatrix} Y & M \\ N & F \\ Y & F \\ \ldots & \ldots \end{bmatrix} \quad n = 106$$

outcome   predictor

Preference: variable
Qual, Nominal, dich

Gender: variable
Qual; Nom, dich

Sort $\begin{bmatrix} N & F \\ Y & F \\ N & M \\ Y & M \end{bmatrix}$ $\begin{array}{c} 20 \\ 29 \\ 5 \\ 52 \end{array}$

gender	Yes	NO	total
F	29	20	49
M	52	5	57
total	81	25	106

<u>Two-By-Two Contingency Table</u>

Q: Are gender + marajuana legal. preference - independent (probability) in this dataset

$$\left.\begin{array}{l} P(Y) \\ P(Y|F) \\ P(Y|M) \end{array}\right\}$$ if all 3 probabilities are <u>similar</u>, gender + MLP are approx. <u>independent</u>

$P(Y) = \frac{81}{106} = 76\%$

$P(Y|F) = \frac{29}{49} = 59\%$ (bc given female)

$P(Y|F) = \frac{52}{57} = 91\%$

$76\% \neq 59\% \neq 91\%$ so variables (gender + MLP) are (strongly) dependent in this dataset.

☆ This problem will help with the midterm ☆

Now that we know $\boxed{\text{and}}$ and $\boxed{\text{or}}$ we can finish T-S problem.

$$P(A) = P(1 \text{ or more T-S in } 5)$$
$$= 1 - P(\text{not } A)$$
$$= 1 - P(\text{exactly } 0 \text{ T-S})$$
$$= 1 - P\left(\overset{1st}{\text{not T-S}} \;\boxed{\text{and}}\; \overset{2nd}{\text{not T-S}} \;\boxed{\text{and}} \ldots \overset{5th}{\text{not T-S}}\right)$$

Independent

$$= 1 - P\left(\overset{1st}{\text{not T-S}}\right) \cdot P\left(\overset{2nd}{\text{not T-S}}\right) \cdots \cdot P\left(\overset{5th}{\text{not T-S}}\right)$$
$$= 1 - \left[1 - P\left(\overset{1st}{\text{T-S}}\right)\right] \cdot \left[1 - P\left(\overset{2nd}{\text{T-S}}\right)\right] \cdots \cdot \left[1 - P\left(\overset{5th}{\text{T-S}}\right)\right]$$

ind. dist

$$= 1 - \left(1 - \tfrac{1}{4}\right)^5 \approx 76\%$$

We can see that the probability of T-S increases w/ number of kids.

R-5 Death Penalty Case Study

$P(DP) = \frac{36}{326} = 11\%$

$P(DP \mid DW) = \frac{19}{160} = 11.9\%$

$P(DP \mid DB) = \frac{17}{166} = 10.2\%$

Surprising... more whites than blacks get death penalty.

Potential PCFs: Race of victim

Given White victim vs Black victim

Y (outcome)	DP or not
X (trtmnt)	Race defendent W, B
Design	Observational Study
Z (PCF)	Race of victim

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

Victim White,

$P(DP \mid VW) = \frac{30}{214} = 14.0\%$

$P(DP \mid VW \text{ and } DW) = \frac{19}{151} = 12.6\%$

$P(DP \mid VW, DB) = \frac{11}{63} = 17.5\%$

Victim Black,

$P(DP \mid VB) = \frac{6}{112} = 5.4\%$

$P(DP \mid VB, DW) = \frac{0}{9} = 0\%$

$P(DP \mid VB, DB) = \frac{6}{103} = 5.8\%$

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"