AMS 7 - Lecture 4·26·18

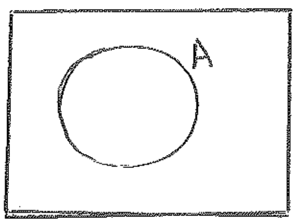
▷ This Time: Conditional Probability

▷ Next Time: Prob. Models for sums

* Read DD ch. 1-3 (A) & DD ch. 1-9 (B)

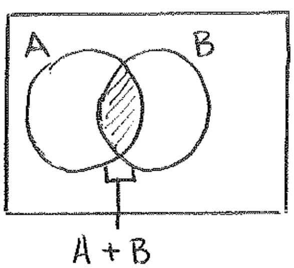
Today → LN pg. 95-118

Next Time → LN pg. 119-126



• What's the probability of A?

$$P(A) = \frac{A}{\square_1} \leftarrow 100\%$$



• What's the probability of A given B?

$$P(A \mid B) = \frac{\text{overlap of } A \,\&\, B}{B}$$

"given"

• DEFINITION (A. de Moivre) - 1705

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

R-37 $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)} \longrightarrow P(B) \cdot P(A \mid B) = P(A \text{ and } B)$ prob. chain rule

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

* $P(Y_1 = 2 \text{ and } Y_2 = 2) \overset{SRS}{=} P(Y_1 = 2) \cdot P(Y_2 = 2 \mid Y_1 = 2)$

$$\frac{1}{3} \cdot 0 = 0 \checkmark$$

$* P(y_1 = 2 \text{ and } y_2 = 2) \overset{\text{IID}}{=} P(y_1 = 2) \cdot P(y_2 = 2 \mid y_1 = 2)$

$= P(y_1 = 2) \cdot P(y_2 = 2)$

- DEFINITION

↳ A & B are independent if and only if information about A does not change chances of B & vice versa : $P(A \,\&\, B) \overset{\text{indep.}}{=} P(A) \cdot P(B)$

* Tay-Sachs Case Study

→ P(1 or more T-S babies in family of 5, both parents carriers)

$= 1 - P(0 \text{ T-S babies})$

$= 1 - P\left(\begin{array}{c}\text{1st baby}\\ \text{Not T-S}\end{array} \underline{\text{AND}} \begin{array}{c}\text{2nd baby}\\ \text{Not T-S}\end{array} \underline{\text{AND}} \ldots \underline{\text{AND}} \begin{array}{c}\text{5th baby}\\ \text{Not T-S}\end{array}\right)$

∴ by independence:

$= 1 - P\left(\begin{array}{c}\text{NOT T-S}\\ \text{on 1st}\end{array}\right) \cdot P\left(\begin{array}{c}\text{NOT T-S}\\ \text{on 2nd}\end{array}\right) \cdot \ldots \cdot P\left(\begin{array}{c}\text{NOT T-S}\\ \text{on 5th}\end{array}\right)$

$= 1 - (1 - \frac{1}{4}) \cdot (1 - \frac{1}{4}) \ldots (1 - \frac{1}{4})$

$= 1 - (1 - \frac{1}{4})^5 = 0.76 = 76\%$

2-37 * UCLA Marijuana Case Study

MLP

n = 106

$$\begin{array}{cc} Y & X \\ \end{array}$$
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}$$

Y = outcome (Yes or No)

X = Predictor (Female or male)

$\xrightarrow{\text{SORT}}$

Female (1), male (0)

Yes (1), No (0)

$$\begin{array}{cc} Y & X \\ \end{array}$$
$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \hline 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ \hline 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ \hline 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\ \vdots\\0\end{array}}\right\} 5 \\ \left.\vphantom{\begin{array}{c}0\\0\\ \vdots\\0\end{array}}\right\} 20 \\ \left.\vphantom{\begin{array}{c}0\\0\\ \vdots\\0\end{array}}\right\} 52 \\ \left.\vphantom{\begin{array}{c}0\\0\\ \vdots\end{array}}\right\} 29 \end{array}$$

MLP

▷ 2x2 contingency table

→ catagorical data analysis

	Yes	No		
	29	20	Female	29+20=49
	52	5	Male	52+5=57
	81	25		n = 106

HW #2 $P(Y) = 81/106 \doteq 76\%$

$P(Y|F) = 29/49 \doteq 59\%$

$P(Y|M) = 52/57 \doteq 91\%$

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

* gender & MLP are associated & are Strongly dependent because 76% yes → 91% male, → 59% Female; huge difference in practical terms.

* Practically significant (highly Practsig)

R-51 * Death Penalty Case Study

Y (outcome): death penalty or not

X (treatment): white vs. black defendent

→ Basic design: observational study

- enemy: bias from PCFs

↳ leading PCF

ethnicity of victom (white vs. black)

- ethnic composition of the jury
- state

* How to defeat the PCF?

↳ Hold it constant in relating X & Y

TOP TABLE

$P(DP|DW) = \frac{19}{160} \doteq 11.9\%$ — white defendant; death penalty

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

$P(DP|DB) = \frac{17}{166} = 10.2\%$ — black defendant

MIDDLE TABLE

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

white victim

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

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

BOTTOM TABLE

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

black victim

AND

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

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

the direction of relationship between X & Y changes & when Z is accounted for

*Simpson's Paradox (1950)

▷ Why did the simpson's paradox arise?

↳ Victim usually knows killer

↳ White people mostly hang around white people & black 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 defendents get DP more than "they really are"

R-52 • Probability Models for sums & Means

*Roulette (38 possibilities) ELM ✓ & >> IID <<

$P(\text{win on a single play}, \underbrace{\text{single \#}}_{\text{Split}}) = \frac{1}{38} \doteq 2.5\%$

Split $= 2/38 = 5\%$

Possible outcome on a single spin

Your net gains → Population

$$38 = N \quad \begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \\ +35 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$$

Sample Observed spins ← Your net gains

$$\begin{bmatrix} \text{mostly } -1 \\ \vdots \\ \text{occasionally} \\ +35 \end{bmatrix} \quad n = 1{,}000$$

Single # 6

$$\text{mean } \mu = \frac{37(-1) + (+35)}{38}$$

"mew"

$$= \frac{-2}{38} \doteq \$0.05$$

↳ negative

σ = sigma

*on average, I expect to win $\mu = -\$0.05$ on each \$1.00 bet for a single #