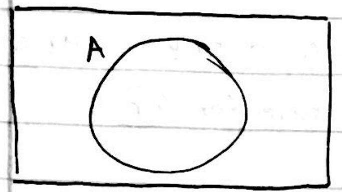
read:
DD ch. 1–3 (A)
DD ch. 1–9 (B)
today:
LN pp. 95–118
next time:
LN pp. 119–126

this time: conditional probability

next time: probability models for sums

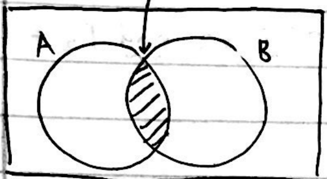
Ashley Tai
AMS 7
26 Apr 18

Conditional Probability



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

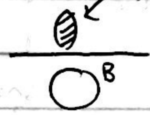
A and B



$P(A \text{ given } B) = ?$

$P(A \mid B) = ?$

A and B



(de Moivre) 1705

$\boxed{\text{definition}}$:

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

prob. chain rule

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

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

$$P(\bar{Y}_1 = 2 \text{ and } \bar{Y}_2 = 2) \overset{\text{SRS}}{=} P(\bar{Y}_1 = 2) \cdot P(\bar{Y}_2 = 2 \mid \bar{Y}_1 = 2)$$
$$= \frac{1}{3} \cdot 0 = 0 \checkmark$$

$$P(\bar{Y}_1 = 2 \text{ and } \bar{Y}_2 = 2) \overset{\text{IID}}{=} P(\bar{Y}_1 = 2) \cdot P(\bar{Y}_2 = 2 \mid \bar{Y}_1 = 2)$$
$$= P(\bar{Y}_1 = 2) \cdot P(\bar{Y}_2 = 2)$$

$\boxed{\text{definition}}$: A and B are independent if and only if information about A doesn't change chances of B & vice versa

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

$\boxed{\text{Tay-Sachs Case study}}$

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

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

$= 1 - P\left(\begin{matrix}\text{1st baby} \\ \text{not T-S}\end{matrix} \text{ and } \begin{matrix}\text{2nd baby} \\ \text{not T-S}\end{matrix} \text{ and } \ldots \text{ and } \begin{matrix}\text{5th baby} \\ \text{not T-S}\end{matrix}\right)$

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

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

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

UCLA Marijuana Case Study R-37

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

Y = outcome: yes ①, no ⓪

X = predictor: F ①, M ⓪

$n = 106$ $\longleftrightarrow$ sort

Y	X	
0 ⋮ 0	1 ⋮ 0	5
1 ⋮ 1	1 ⋮ 1	20
0 ⋮ 1	1 ⋮ 0	52
0 ⋮ 0	1 ⋮ 1	29

2x2 contingency table

	Y	N	
F	29	20	49
M	52	5	57
	81	25	106

categorical data analysis

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

$$P(Y \mid F) = \frac{29}{49} \doteq 59\%$$

$$P(Y \mid M) = \frac{52}{57} \doteq 91\%$$

Q: Are gender & MLP independent or dependent in this dataset?

A: dependent (G & MLP are associated)

↳ strongly dependent: 76% (Y) $\lessgtr$ 91% M, 59% F

91% differs from 59% by an amount that huge in practical terms (ie. highly practically significant (practsig))

Death Penalty Case Study

outcome (Y): death penalty or not

treatment (X): white vs. black (defendant)

$\boxed{\text{basic design}}$: observational study

enemy: bias from PCFs

(Z) leading PCF: ethnicity of victim (white vs. black)

- ethnic composition of jury
- state

how to defeat PCF: hold it constant in relating X & Y

death penalty

	Y	N	
white	19	141	160
black	17	149	166
	36	290	326

$$P(DP \mid DW) = \frac{19}{100} = 11.9\%$$

(death penalty, white defendant)

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

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

white victim

defendant / death penalty

defendant	Y	N	
white	19	132	151
black	11	52	63
	30	184	214

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

↑ white victim

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

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

black victim

defendant / death penalty

defendant	Y	N	
white	0	9	9
black	6	97	103
	6	106	

black victim

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

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

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

Simpson's Paradox (1950) : direction of relationship between X and Y changes when Z is accounted for

Probability Models for sums & means

Roulette (R - 52)

0, 00, 1, ⋮, 36 } 38

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

↓

$\text{split} = \frac{2}{38} = \frac{1}{19} \doteq 5\%$

population: possible outcomes on a single spin

sample: the observed spins

$\text{mean}(\mu) = \frac{37(-1) + (+35)}{38} = -\frac{2}{38} \doteq -0.05$

* On average I expect to win $\mu = -0.05$ on each \$1 bet on a single #

ELM ✓

IID ✓

your net gain ($N = 38$)

	net gain
1	-1
2	-1
3	⋮
4	-1
5	
6	+35
⋮	-1
36	⋮
0	
00	

your net gain: mostly -1, occasionally +35 ($n = 1{,}000$)