This time: conditional probability

read: DD ch. 1–3 (A), DD (B) ch. 1–9

AMS 7
26 Apr 18

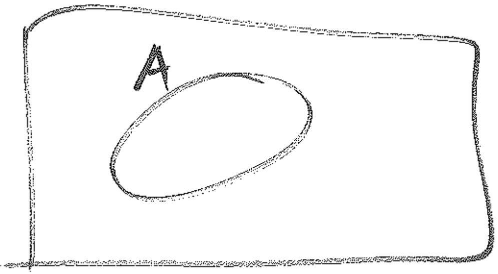
(1)

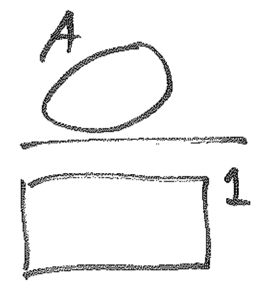
next time: prob. models for sums

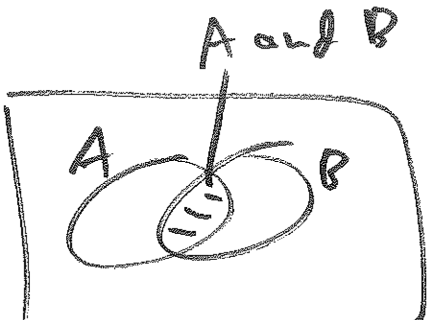
today: LN pp. 95–118;

next time: LN pp. 119–126

conditional probability



P(A) = 

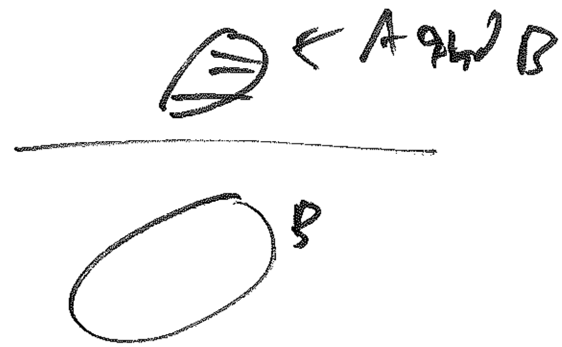


$P(A \text{ given } B) = \,?$

$\quad\,\, \|$

$P(A \mid B) = \,?$

def: (de Moivre 1705)

= 

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

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

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

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

(prob. chain rule)

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

$$P(I_1 = 2 \text{ and } I_2 = 2) \overset{\text{SRS}}{=}$$

$$P(I_1 = 2)\cdot P(I_2 = 2 \mid I_1 = 2)$$

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

$$P(Y_1 = 2 \ \& \ Y_2 = 2) \overset{IID}{=}$$

$$P(Y_1 = 2) \cdot \boxed{P(Y_2 = 2 \mid Y_1 = 2)}$$

$$=$$

$$= P(Y_1 = 2) \cdot P(Y_2 = 2)$$

---

def. A, B are independent if & only if information about A doesn't change chances of B & vice versa:

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

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

Tay Sachs case study

④

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

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

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

$\overset{\text{(II)}}{=} 1 - \left(1 - \frac{1}{4}\right) \cdot \left(1 - \frac{1}{4}\right) \cdots \left(1 - \frac{1}{4}\right)$

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

UCLA marijuana case study

Ⓡ - 37

Y = outcome — yes (1), no (0) MLP (5)

X = predictor — F (1), M (0)

$$\begin{array}{cc} Y & X \\ \end{array}$$
$$\left[\begin{array}{c|c} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{array}\right]$$

$n = 106$

$\longleftrightarrow$ sort

$$\begin{array}{c} Y \quad X \\ \left[\begin{array}{c|c} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \hline 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ \hline 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ \hline 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{array}\right] \end{array} \quad \begin{array}{c} 5 \\ \\ 20 \\ \\ 52 \\ \\ 29 \end{array}$$

2x2 contingency table

categorical data analysis

MLP

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

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

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

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

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

A: dependent (G & MLP <u>are</u> associated)

<u>strongly</u> dependent: 76% (Y) → 91% M, → 59% F

91% differs from 59% by an amount that's large in practical terms (i.e., (highly) practically significant (practsig))

---

death penalty | case study |: R-51

7

outcome (Y): death penalty or not

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

basic design | obs. study

enemy: bias from PCFs

leading PCF: ethnicity of victim (white vs. black) | ethnic composition of jury | state

Z

how defeat PCF | held it constant in relating X & Y

(8)

top table

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

death penalty

white defendant

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

defendant black

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

middle table

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

victim white

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

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

bottom table

$P(\text{DP} \mid V\underline{B}) = \frac{6}{112} \doteq 5.4\%$ (9)

victim black

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

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

---

direction of relationship between $\underline{X}$ & $\underline{Y}$ <u>changes</u> when $\underline{Z}$ is accounted for: <u>Simpson's</u> (1950) <u>Paradox</u>

Roulette (R-52)

Probability models for sums & means

(10)

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

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

0
00
1
⋮
36

$\}$ 38

ELM? ✓

IID ✓

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

(A) single #: $\boxed{6}$

pop

possible outcomes on a single spin

your net gain

-$1	1
-1	2
⋮	⋮
-1	5
+35	6
-1	7
⋮	⋮
	36
	0
	00

N = 38

"new" mean $\mu = -\$0.05$

SD σ = "sigma"

sample

the observed spins

your net gain

mostly -$1, occasionally +$35

n = 1,000

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