this time: probability;
next time: probability models for sums & means

read: DD ch. 1-8, 9, 10

AMS 7
29 Jan 09

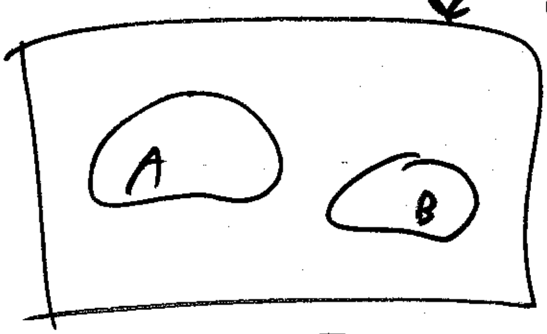
1

hwk 2 due next Tue in class

lab 2 assignment due by this Fri 5pm under my office door or in the box outside my office, except: if your lab is on Fri, assignment due by Mon 9AM

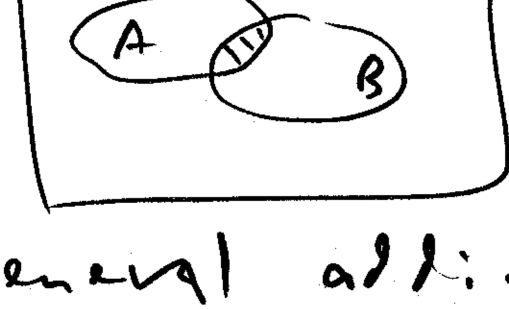
no overlap: A, B

A, B mutually exclusive



$$P(A \text{ or } B) = P(A) + P(B)$$

overlap



$$\boxed{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)}$$

general addition rule for OR

pop — at random → sample, n = 2

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow{\text{at random}} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad n = 2$$

① with replacement ②

$\boxed{\text{PLAN AHEAD}}$

P(2 on 1st draw and 2 on 2nd draw) = ?

with repl. : independent identically distributed = IID sampling

IID

	Y_2 2nd draw: 1	2	9
Y_1 1st draw: 1	(1,1)	(1,2)	(1,9)
2	(2,1)	(2,2)	(2,9)
9	(9,1)	(9,2)	(9,9)

ELM applies

P(2 on 1st draw & 2 on 2nd) = $\frac{1}{9}$

P(2 on 1st) = $\frac{3}{9} = \frac{1}{3}$; P(2 on 2nd) = $\frac{3}{9} = \frac{1}{3}$

P(2 on 1st & 2 on 2nd) = P(2 on 1st) · P(2 on 2nd)

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

theory: P(A and B) = P(A) · P(B)

at random without replacement

= simple random sampling = SRS

P(2 on 1st and 2 on 2nd draw) = 0

SRS	1	2 (4)	9
1	~~(1,1)~~	(1,2)	(1,9)
(4) 2	(2,1)	~~(2,2)~~	(2,9)
9	(9,1)	(9,2)	~~(9,9)~~

$$P(2 \text{ on } 1^{st}) = \frac{2}{6} = \frac{1}{3}$$

$$P(2 \text{ on } 2^{nd}) = \frac{2}{6} = \frac{1}{3}$$

---

with SRS $\quad P(2 \text{ on } 1^{st} \text{ and } 2 \text{ on } 2^{nd}) \neq \frac{1}{3} \cdot \frac{1}{3}$

$$0 \neq P(2 \text{ on } 1^{st}) \cdot P(2 \text{ on } 2^{nd})$$

---

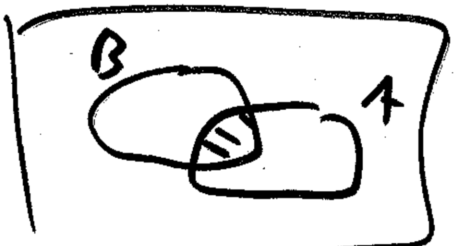
conditional probability

P(B given A) = P(B | A)

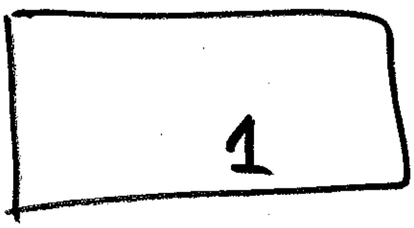
| = "given"

④

defining this concept: Rev. Thomas Bayes (≐1720)



P(B) =



P(B given A) =

B and A

def.: P(B given A) = $\frac{P(A \text{ and } B)}{P(A)}$

P(A and B)

= P(A) · P(B given A)

= P(B) · P(A given B)

general product rule for and

with sets

0 = P(2 on 1st & 2 on 2nd) = P(2 on 1st) · P(2 on 2nd | 2 on 1st)

= $\frac{1}{3}$ · 0 = 0 ✓

definition | A, B independent (in a probability sense) if information about A does not help you to predict B, & vice versa

IID draws are independent & identically distributed

~~if A, B indep = P(B) · P(A | B)~~

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

if A, B indep: $P(B \mid A) = P(B)$

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

so if A, B indep

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

T-S case study

$P(1 \text{ or more T-S in 5 kids})$

$$= 1 - P(0 \text{ TS in 5 kids})$$

$$= 1 - P(\text{not TS on 1st} \,\&\, \text{not TS on 2nd} \,\&\, \cdots \,\&\, \text{not TS on 5th})$$

independence (from biology)

$$= 1 - P(\text{not TS on 1st}) \cdot P(\text{not TS on 2nd}) \cdots P(\text{not TS on 5th})$$

iden. dist. (from biology)

$$= 1 - \left(1 - \tfrac{1}{4}\right) \cdot \left(1 - \tfrac{1}{4}\right) \cdots \left(1 - \tfrac{1}{4}\right)$$

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

UCLA case study

MLP = marijuana legalization preference

MLP	gender
N	F
N	M
Y	F
⋮	⋮

$n = 106$

7

Q: are gender & MLP associated in this dataset?

not associated = independent

MLP

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

2x2 contingency table

choose 1 person at random

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

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

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

in this dataset, there is a strong

assoc. betw. gender & MLP (this diff is large in pract. terms)

DP case study | $P(DP) = \frac{36}{326} \doteq 11\%$ (8)

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

defendant white

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

defendant black

---

outcome (Y): DP or not

treatment (X): race of defendant (W, B)

basic design: obs. study

enemy: bias from PCFs

✓ PCF (Z): race of victim

how defeat ~~by~~ bias from PCF: hold it constant

9

$P(DP \mid VW) = \frac{30}{214} \doteq 14\%$

victim white

and

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

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

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

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

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

---

direction of ~~rel~~ relationship between X & Y changes when PCF Z is controlled for: Simpson's Paradox