this time: probability

next time: prob. models for sums

read: DD(B) ch. 9, LN pp. L-119 – L-126

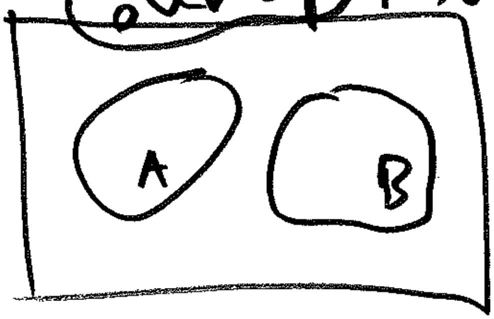
AMS 7
23 oct 18

today: LN pp. L-97 →

R-37

working with $\boxed{\text{OR}}$

no overlap 100%



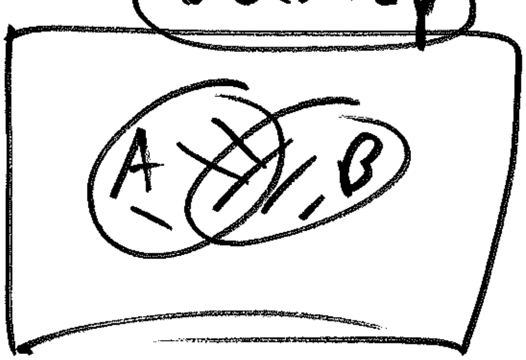
$$P(A \text{ or } B) = P(A) + P(B)$$

special case of $\boxed{\text{OR}}$

when $\boxed{A, B \text{ don't overlap}}$

$\leftrightarrow$ $\boxed{A, B \text{ are mutually exclusive}}$

overlap



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

general rule for $\boxed{\text{or}}$

working with $\boxed{\text{AND}}$ — 2 cases to consider ②

pop. — case 1 $\boxed{\text{IID}}$ — sample

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

$$P(Y_1 = 2 \text{ and } Y_2 = 2) = ?$$

case 1

at random with replacement (independent identically distributed (IID))

$\boxed{\text{IID}}$

1st draw \ 2nd draw	1	2	9
1	(1, 1)	(1, 2)	(1, 9)
2	(2, 1)	(2, 2)	(2, 9)
9	(9, 1)	(9, 2)	(9, 9)

$\boxed{\text{Q:}}$ ELM apply to these 9 possibilities?

$\boxed{\text{A:}}$ yes

$$P(Y_1 = 2 \text{ and } Y_2 = 2) = \frac{1}{9}$$

$$P(A \text{ and } B) \stackrel{?}{=} P(A) \, ? \, P(B)$$ 3

$$P_{IID}(X_1 = 2) = \frac{1}{3} = \frac{3}{9}$$

$$P_{IID}(X_2 = 2) = \frac{1}{3} = \frac{3}{9}$$

$$P(X_1 = 2 \text{ and } X_2 = 2) = \frac{1}{9}$$

$$P(X_1 = 2) \cdot P(X_2 = 2)$$

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

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

works for IID sampling

(1^{st} draw, 2^{nd} draw independent)

Case 2: simple random sampling (SRS): at random without replacement

pop

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow{\text{SRS}} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \quad n = 2$$

sample

$$P_{SRS}(I_1 = 2 \text{ and } I_2 = 2) = 0$$

$\boxed{\text{SRS}}$	2nd draw: 1	2	9
1st draw: 1	$\cancel{(1,1)}$	(1,2)	(1,9)
2	(2,1)	$\cancel{(2,2)}$	(2,9)
9	(9,1)	(9,2)	$\cancel{(9,9)}$

$\boxed{\text{Q:}}$ ELM apply to those 6 possibilities?

$\boxed{\text{A:}}$ yes

$$P_{SRS}(I_1 = 2) = \frac{1}{3} = \frac{2}{6}$$

$$P_{SRS}(I_2 = 2) = \frac{1}{3} = \frac{2}{6}$$

$$P_{SRS}(\mathbb{I}_1 = 2 \text{ and } \mathbb{I}_2 = 2) = 0$$ ⑤

$$\neq \underbrace{P_{SRS}(\mathbb{I}_1 = 2)}_{} \cdot P_{SRS}(\mathbb{I}_2 = 2)$$

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

our theory doesn't work for SRS

IID 2nd draw doesn't depend on 1st draw

SRS 2nd draw depends on 1st draw

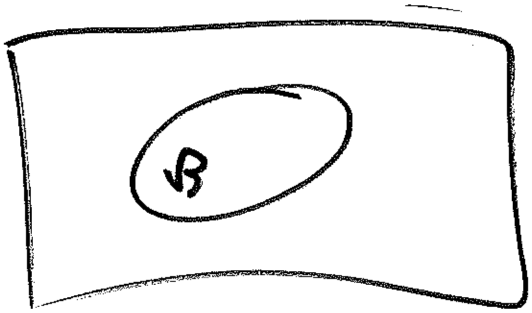
IFD Pascal, Fermat 1650

SRS de Moivre (1710),
Bayes (1760)

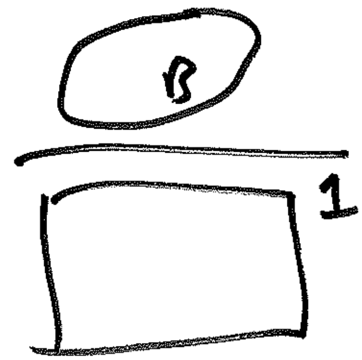
conditional probability

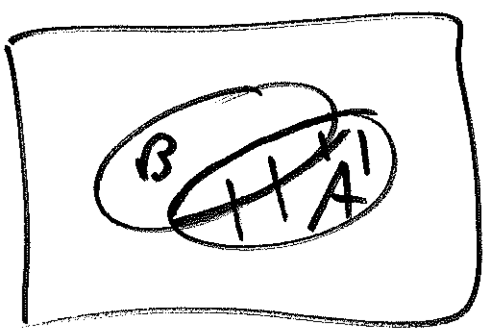
$P(B \text{ given } A) = ?$

$P(B|A)$



$P(B) =$





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

$=$

def:

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

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

7

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

mult. by $P(A)$:

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

$\rightarrow P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$

mult. by $P(B)$.

general product rule for and

$$\boxed{\begin{aligned} P(A \text{ and } B) &= P(B) \cdot P(A|B) \\ &= P(A) \cdot P(B|A) \end{aligned}}$$

(chain rule for prob)

special case

(IID) **def.** A, B are **independent** iff

information about A does not change chances for B, & vice versa

$X_1 = 2$ $\quad$ $X_2 = 2$ $\quad\quad$ if A, B indep

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

IID

$$= P(X_1 = 2) \cdot P(X_2 = 2 \mid X_1 = 2)$$

---

if A, B are indep then

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

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

$$P_{IID}(Y_1 = 2 \text{ and } Y_2 = 2)$$

$$= P_{IID}(Y_1 = 2) \cdot P_{IID}(Y_2 = 2 \mid Y_1 = 2)$$

$$= \frac{1}{3} \cdot P_{IID}(Y_2 = 2)$$

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

$$P_{SRS}(Y_1 = 2 \text{ and } Y_2 = 2)$$

$$= P_{SRS}(Y_1 = 2) \cdot P_{SRS}(Y_2 = 2 \mid Y_1 = 2)$$

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

(12.41)

MLP	G
Y	F
Y	M
N	F
⋮	⋮

$n = 106$

1 row for each individual (person)

(sort)

Y	F	29
⋮	⋮	
Y	F	
N	F	20
⋮	⋮	
N	F	
Y	M	52
⋮	⋮	
Y	M	
N	M	5
⋮	⋮	
N	M	
		106

(dich) categorical data analysis

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

2 x 2 contingency table

Q: Are G & MLP independent in this dataset?

A: Imagine choosing a person at random from 106 people (ELM? yes)

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

$$P(\text{yes} \mid \text{female}) = \frac{29}{49} \doteq 59\%$$

$$P(\text{yes} \mid \text{male}) = \frac{52}{57} \doteq 91\%$$

---

since $91\% \neq 59\% \neq 76\%$,

G & MLP are strongly dependent (in this data set)

(are associated)

$\boxed{\text{R-51}}$ outcome (Y): { Y death penalty (DP) / N no dp

"treatment" (X): defendant { W white / B black (nonwhite)

L-114