AMS 7
15 Oct. 15

This Time: probability

Next Time: prob. models for sums & means

Read: DD CBI Ch. 9-11 , LN pp L-119-136

Hwk 2 due next Thr in class

$$\overset{\text{pop}}{\begin{bmatrix}1\\2\\9\end{bmatrix}} \xrightarrow[\text{random}]{\text{at}} \begin{bmatrix} y \end{bmatrix} n=1$$

equally likely model applies ✓

$P(y \text{ is odd}) = 2/3 \doteq 67\% \doteq 0.67$

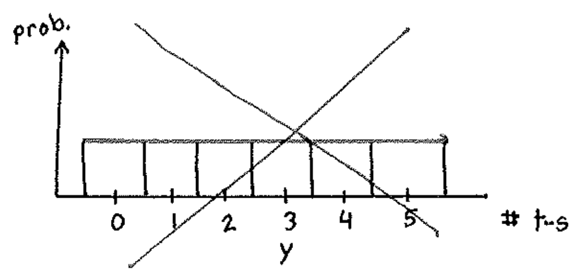
$$\begin{bmatrix}0\\ \hline 1\\2\\3\\4\\5\end{bmatrix}$$

y = # of their children with tay-sachs disease

if ELM applies

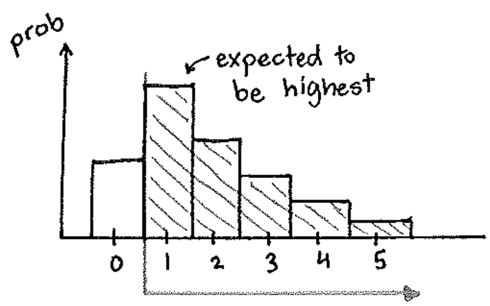
$p(1 \text{ or more T-s}) = 5/6 \doteq 83\% \doteq 0.83$

but ELM does <u>not</u> apply



if ELM

but not so!

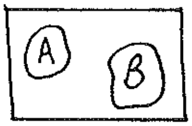


A | not A

therefore:

$P(A) + P(\text{not } A) = 1 = 100\%$

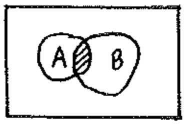
$\boxed{P(A) = 1 - P(\text{not } A)}$



no overlap ↔ A, B are mutually exclusive

addition rule, special case:

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



overlap

general additional rule for or

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

$$\overset{\text{pop}}{\begin{bmatrix}1\\2\\9\end{bmatrix}} \xrightarrow[\text{random}]{\text{at}} \begin{bmatrix}Y_1\\Y_2\end{bmatrix} n=2 \qquad P(Y_1=9 \text{ and } Y_2=9) = ?$$

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

at random → without replacement | simple random sampling (SRS)

case 1:

$$\begin{bmatrix}1\\2\\9\end{bmatrix} \xrightarrow{\text{IID}} \begin{bmatrix}Y_1\\Y_2\end{bmatrix} \qquad P(Y_1=9 \text{ and } Y_2=9) = ?$$

IID

Y_1 \ Y_2	1	2	9
1	(1,1)	(1,2)	(1,9)
2	(2,1)	(2,2)	(2,9)
9	(9,1)	(9,2)	(9,9)

ELM? yes

$P(\overbrace{Y_1=9}^{A} \text{ and } \overbrace{Y_2=9}^{B}) = 1/9$ ✓

$P(Y_1=9) = P(A) = 1/3 = 3/9$

$P(Y_2=9) = P(B) = 1/3 = 3/9$

our conjecture: $P(A \text{ and } B) = P(A) \cdot P(B)$ this works for IID

$1/9 = 1/3 \cdot 1/3$

$$\begin{bmatrix}1\\2\\9\end{bmatrix} \xrightarrow{\text{SRS}} \begin{bmatrix}Y_1\\Y_2\end{bmatrix} \qquad P(Y_1=9 \text{ and } Y_2=9) = 0 \leftarrow \text{common sense}$$

SRS

Y_1 \ Y_2	1	2	9
1	~~(1,1)~~	(1,2)	(1,9)
2	(2,1)	~~(2,2)~~	(2,9)
9	(9,1)	(9,2)	~~(9,9)~~

ELM? yes

$P(Y_1=9) = 1/3 = 2/6$

$P(Y_2=9) = 1/3 = 2/6$

$\boxed{\text{by our conj.:}}$

$P_{SRS}(Y_1=9 \text{ and } Y_2=9) = 1/3 \cdot 1/3 = 1/9$

but actually it's $0 = 0/6$

our conj. fails for SRS

Conditional Probability

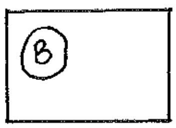
Bayes 1760 → Bayesians

and or not — Pascal, Fermat 1650

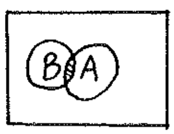
$P(B|A) = ?$

"given"

ex. $P(Y_2 = 9 \mid Y_1 = 2) = ?$



$$P(B) = \frac{\text{area of } B}{\text{total area } (1)} = \frac{B}{\square}$$



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

def: $P(B|A) = \dfrac{P(A \text{ and } B)}{P(A)}$ (provided $P(A) > 0$; $P(B|A)$ is undefined if $P(A) = 0$)

so $P(A \text{ and } B) = P(A) \cdot P(B|A)$ general product rule for and

$= P(B) \cdot P(A|B)$

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

def: A, B independent if information about A does not change your predictions about B, & vice versa

Product Rule, IID Sampling: $P(\underset{(Y_1=9)}{A} \text{ and } \underset{(Y_2=9)}{B}) = P(A) \cdot P(B|A)$

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

if A, B are independent, then $P(A \text{ and } B) = P(A) \cdot P(B|A)$

but $P(B|A) = P(B)$

so if A, B indep: $P(A \text{ and } B) = P(A) \cdot P(B)$

$P(\text{1 or more T-S in family of 5}) = ?$

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

$$= 1 - P\left(\text{not T-S on 1st baby} \text{ and } \text{not T-S on 2nd} \text{ and } \dots \text{ and } \text{not T-S on 5th}\right)$$

$$\text{indep.} = 1 - P\binom{\text{not T-S}}{\text{on 1st}} \cdot P\binom{\text{not T-S}}{\text{on 2nd}} \cdot \ldots \cdot P\binom{\text{not T-S}}{\text{on 5th}}$$

$$\text{identical distribution fact} = 1 - (1 - 1/4) \cdot (1 - 1/4) \cdot \ldots \cdot (1 - 1/4)$$

$$= 1 - (3/4)^5 = 0.76 = 76\%$$

Marijuana Legalization Preference

mlp	gender
yes	m
no	f
⋮	⋮
Y	X

$n = 106$

1 row for each UCLA stat student in my class in 1992 who participated

sorted into

Y ⋮ Y	m ⋮ m	52
Y ⋮ Y	f ⋮ f	29
N ⋮ N	m ⋮ m	5
N ⋮ N	f ⋮ f	20

$\boxed{\text{Q:}}$ Are gender & mlp independent in this data set?

G \ mlp	Y	N	
f	29	26	49
m	52	5	57
	81	25	106

choose a student at random...

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

$P(y \mid m) = \frac{52}{57} \doteq 91\%$ $\quad P(y \mid f) = \frac{29}{49} \doteq 59\%$

$\boxed{\text{A:}}$ No, they're strongly dependent. 91 & 59 are not close to 76

$$\begin{pmatrix} \text{and } \times \\ \text{or } + \\ \text{not} \\ \text{given} \end{pmatrix}$$

Simpson's paradox discussed