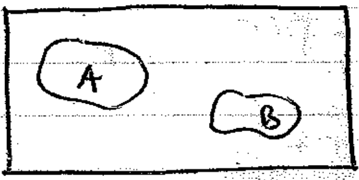
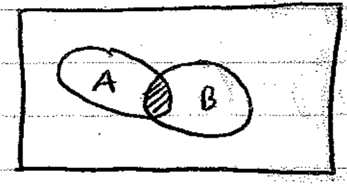
1) Lecture # 7 Probability — 29 Jan, 2009



no overlap : A, B → A, B mutually exclusive

P(A (or) B) = P(A) + P(B)



With overlap

P(A (or) B) = P(A) + P(B) − P(A (and) B)

* General Addition Rule for (or) ↗

$$\begin{array}{c} \text{population} \\ \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \end{array} \xrightarrow{\text{at random}} \begin{array}{c} \text{sample} \\ \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \end{array} \quad n = 2$$

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

— depends on with or without replacement

① With Replacement : Independent Identically Distributed (IID)

y_1 (1st draw ↓	y_2 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)

ELM applies (Equally Likely Model)

P(2 on 1st draw and 2 on 2nd draw) = 1/9

P(2 on 1st draw) = 3/9 = 1/3

P(2 on 2nd draw) = 3/9 = 1/3

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

1/9 = 1/3 · 1/3

CD

29 Jan, 2008

2) Lecture #7

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

② * At Random without Replacement
= Simple Random Sampling = SRS

$P(2 \text{ on } 1^{st} \text{ and } 2 \text{ on } 2^{nd} \text{ draw}) = 0$

SRS

		y_2		
		1	2	9
y_1	1	~~(1,1)~~	(1,2)	(1,9)
	2	(2,1)	~~(2,2)~~	(2,9)
	9	(9,1)	(9,2)	~~(9,9)~~

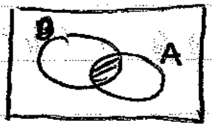
$P(2 \text{ on } 1^{st}) = \frac{2}{6} = 1/3$

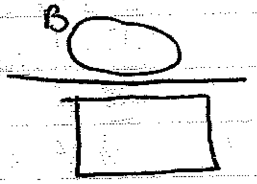
$P(2 \text{ on } 2^{nd}) = \frac{2}{6} = 1/3$

With SRS $P(2 \text{ on } 1^{st} \text{ and } 2 \text{ on } 2^{nd} \neq 1/3 \cdot 1/3$
$0 \neq P(2 \text{ on } 1^{st}) \cdot P(2 \text{ on } 2^{nd})$

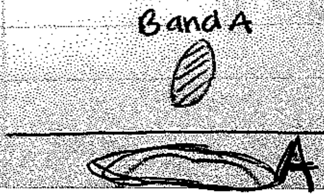
* Additional Probability $P(B \text{ given } A) = P(B|A)$
$| =$ "given"

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



P(B) = 

P(B given A) =



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

3) Lecture #7 — 29 Jan, 2009

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

^ general product rule for and

With SRS $P(2 \text{ on } 1^{st} \text{ and } 2 \text{ on } 2^{nd}) =$

$$P(2 \text{ on } 1^{st}) \cdot P(2 \text{ on } 2^{nd} \mid 2 \text{ on } 1^{st}) \checkmark$$
$$\tfrac{1}{3} \cdot 0 \qquad = 0$$

^ Definition: A, B independent (in a probability sense) if information about A does not help you to predict B, and visa versa.

→ IID draws are Independent and identically distributed

→ if A, B independent:

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

→ If A, B indep: $P(B|A) = P(B)$
and $P(A|B) = P(A)$

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

$$= 1 - P(0 \text{ T-S in 5 kids})$$
$$= 1 - P\left(\substack{\text{not T-S} \\ \text{on } 1^{st}} \text{ and } \substack{\text{not T-S} \\ \text{on } 2^{nd}} \text{ and} \ldots \text{and } \substack{\text{not T-S} \\ \text{on } 5^{th}}\right)$$

independence (from biology) $= 1 - P\left(\substack{\text{not T-S} \\ \text{on } 1^{st}}\right) \cdot P\left(\substack{\text{not T-S} \\ \text{on } 2^{nd}}\right) \cdot \ldots \cdot P\left(\substack{\text{not T-S} \\ \text{on } 5^{th}}\right)$

identical distribution (from biology) $= 1 - (1-\tfrac{1}{4}) \cdot (1-\tfrac{1}{4}) \cdot \cdots (1-\tfrac{1}{4})$

$$= 1 - (1-\tfrac{1}{4})^5 = .76 = 76\%$$

CD

4) Lecture #7 — 29 Jan, 2009

UCLA Case Study: mlp: marijuana legalization preference

n=106

MLP	Gender
N	F
N	M
Y	F
⋮	⋮

Q: Are gender and mlp associated in this data set?

- not associated = independent

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

2x2 Contigency Table

Choose 1 person at random $P(Y) = \frac{81}{106} \doteq 76\%$

$P(Y|F) = \frac{29}{49} \doteq 59\%$ $\quad P(Y|M) = \frac{52}{57} = 91\%$

^ In this data set there is a strong association between gender and mlp.

59% & 91% ⇐ difference is large in practical terms

CD

pg 5 of 5

5) Lecture #7 29 Jan, 2009

Death Penalty Case Study $P(DP) = \frac{36}{326} = 11\%$

dW: defendent white $P(DP|DW) = \frac{19}{160} = 11.9\%$

DB: defendent black $P(DP|DB) = \frac{17}{166} = 10.2\%$

Ⓨ Outcome: **DP** or not

Ⓧ Treatment: Race of dependent (W, B)

Basic Design: Observational study

Enemy: Bias from PCFs

Ⓩ ✓PCF: Race of the Victim

How defeat bias from PCF: hold it constant

~ $P(DP|VW)$ $\frac{30}{214} \doteq 14\%$ ~ $P(DP|VW \text{ and } DW)$ $\frac{19}{151} \doteq 12.6\%$

victime white

~ $P(DP|VW \text{ and } DB) = \frac{11}{63} \doteq 17.5\%$

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

~ $P(DP|VB \text{ and } DW) = \frac{0}{9} = 0\%$

~ $P(DP|VB \text{ and } DB) = \frac{6}{103} = 5.8\%$

* Direction of relationship between Ⓧ and Ⓨ changes when PCF Ⓩ is controlled for: Simpson's Paradox