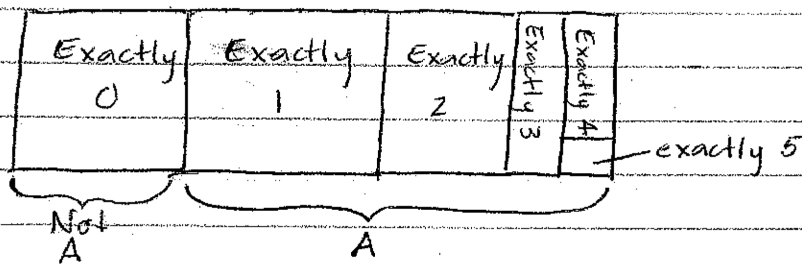
Jan. 29, 2009

Probability (con't)

- Rule for $\boxed{\text{Not}}$

Ex: Tay-Sachs babies (probability of having 1 or more)

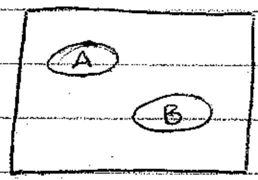


$$* \boxed{P(\text{1 or more T-S babies}) = 1 - P(\text{no T-S babies})} *$$

- Rule for $\boxed{\text{Or}}$

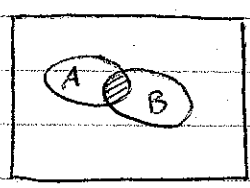
- 2 cases to consider:

1) No overlap (A & B are mutually exclusive)



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

2) Overlap



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

- Rule for $\boxed{\text{And}}$

Population

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow[\text{random}]{\text{at}} \overset{\text{sample}}{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}} \quad n = 2$$

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

Jan. 29, 2009

1) With replacement

- Independent Identically Distributed = IID sampling

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 applies

$P(2 \text{ on } 1^{st} \text{ draw}) = 3/9 = 1/3$
$P(2 \text{ on } 2^{nd} \text{ draw}) = 3/9 = 1/3$
$P(2 \text{ on } 1^{st} \text{ \& } 2^{nd} \text{ draw}) = 1/3(1/3) = \boxed{1/9}$

* Theory: With IID sampling, $P(A \text{ and } B) = P(A) \cdot P(B)$

2) At random without replacement

- simple random sampling = 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)~~

*can't draw same # twice

$P(2 \text{ on } 1^{st} \text{ draw}) = 2/6 = 1/3$
$P(2 \text{ on } 2^{nd} \text{ draw}) = 2/6 = 1/3$
? $P(2 \text{ on } 1^{st} \text{ \& } 2^{nd} \text{ draw}) = 1/3(1/3) = 1/9$?

- Not true

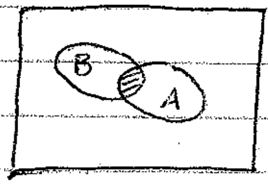
We know that without replacement, it is impossible to draw 2 twice.
$P(2 \text{ on } 1^{st} \text{ \& } 2^{nd} \text{ draw}) = 0$

* Our previous theory doesn't work with SRS because the 2^{nd} draw depends on the 1^{st} draw. With IID, the 2 draws are independent.

- Conditional Probability

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

Jan. 29, 2009



$P(B) = \frac{(B)}{\square}$

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

· Definition: $P(B \text{ given } A) = \frac{P(A \,\&\, B)}{P(A)}$

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

· How does this apply to SRS?
back to probability of drawing a 2 on both draws:

$P(2 \text{ on } 1^{st} \,\&\, 2 \text{ on } 2^{nd} \text{ draw}) = P(2 \text{ on } 1^{st}) \cdot P(2 \text{ on } 2^{nd} | 2 \text{ on } 1^{st})$
$= 1/3 \cdot 0 = \boxed{0}$

· Conclusion: A and B are independent (in a probability sense) if information about A does not help you to predict B, and vice versa.
- IID draws are independent
- SRS draws are dependent

* If A and B are independent: $P(A \text{ and } B) = P(A) \cdot P(B|A)$
· where $P(B|A) = P(B)$
$P(A|B) = P(A)$
so, $P(A \text{ and } B) = P(A) \cdot P(B)$

- Tay-Sachs Case Study (con't)
· $P(1 \text{ or more T-S babies}) = 1 - P(\text{no T-S babies})$
$= 1 - P(\text{not T-S on first and} \ldots \text{and not T-S on } 5^{th})$
· Independent (from Biology)
$= 1 - P(\text{not T-S on } 1^{st}) \cdot P(\text{not T-S on } 2^{nd}) \cdot \ldots \cdot (\text{not T-S on } 5^{th})$

Jan. 29, 2009

- Identical Distribution (from biology)

$= 1-(1-1/4)\cdot\ldots\cdot(1-1/4)$

$= 1-(1-1/4)^5 = .76 =$ $\boxed{76\%}$ chance of having 1 or more

- UCLA Case Study: Legalizing Marijuana

MLP = marijuana legalization preference (yes or no)

$$\begin{array}{c|c} MLP & Gender \\ \end{array}\left[\begin{array}{c|c} N & F \\ N & M \\ Y & F \\ \vdots & \vdots \end{array}\right] \quad n = 106$$

- Question: Are gender and MLP associated in this data set?

(not associated = independent)

MLP

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

- choose 1 person at random:
  - $P(Y) = \frac{81}{106} = 76\%$
  - $P(Y \mid \text{given female}) = \frac{29}{49 \text{ (total F)}} = 59\%$
  - $P(Y \mid \text{male}) = \frac{52}{57 \text{ (total M)}} = 91\%$

* In this data set, there is a strong association between gender and MLP

Jan. 29, 2009

Example: Death Penalty Case Study

P(death penalty) = $\frac{36}{326}$ = 11%

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

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

* very slightly associated

- Enemy = bias from PCF's
  - race of victim = PCF = z
  - how to defeat bias from PCF? hold it constant

→ white victim:

P(DP | victim white) = $\frac{30}{214}$ = 14%

P(DP | victim white & defendent white) = $\frac{19}{151}$ = 12.6%

P(DP | victim white & defendent black) = $\frac{11}{63}$ = 17.5%

→ black victim:

P(DP | victim black) = $\frac{6}{112}$ = 5.4%

P(DP | victim black & defendent white) = $\frac{0}{9}$ = 0%

P(DP | victim black & defendent black) = $\frac{6}{103}$ = 5.8%

* Direction of relationship between y (DP or not) and x (race of defendent) changes when PCF z (race of victim) is controlled for = <u>Simpson's Paradox</u>