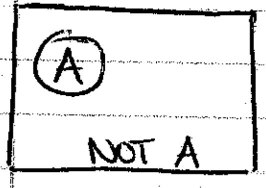
MS

01/29/2009

PROBABILITY RULES

"NOT"



$0\% \leq P(A) \leq 100\%$

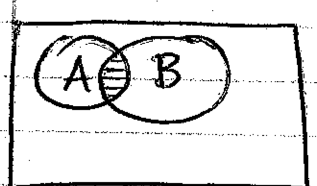
$P(A) + P(\text{not } A) = 100\%$

$$PA = 100\% - P(\text{not } A)$$

"OR"

There can either be overlap between A & B, or there can Be no overlap between A & B

- A & B being <u>mutually exclusive</u> (no overlap) is a special case of the following rule:

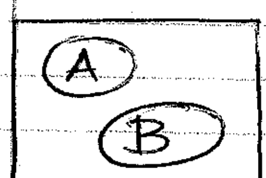


When there is overlap:

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

$$= A + B - \underbrace{\quad}_{\text{We need the rule for working w/ "AND"}}$$

Special Case: A & B are mutually exclusive:



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

There is no piece that would get counted twice like there was in the general rule ( ), so nothing needs to be subtracted.

The special case for "OR" is also simple for more variables,

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

But the rule gets very complicated if there is overlap, and the equation changes on a case by case basis.

01/29/2009

PROBABILITY RULES (CONT.)

"AND"

- use sample population to better understand:

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow[\text{RANDOM}]{@} \text{2 Independent draws} \quad \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad n=2$$

What is the probability of drawing a 2 on both the 1st draw and the 2nd draw?

P(2 (on 1st draw) AND 2 (on 2nd draw))

This probability depends on whether the draws are made with replacement (IID) or without replacement (SRS)

ASSUME IID:

	1	2	9
1	1,1	2,1	9,1
2	1,2	2,2	9,2
9	1,9	2,9	9,9

3

P(2 on 1st draw)

ELM (Equally Likely Model) applies.

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

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

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

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

W/ IID: $p(A \& B) = p(A) \times p(B)$

ASSUME SRS:

Intuitively, P(2 on 1st draw & 2 on 2nd draw) = 0

getting same # twice is impossible.

	1	2	9
1	~~1,1~~	1,2	1,9
2	2,1	~~2,2~~	2,9
9	9,1	9,2	~~9,9~~

p(2 on 1st draw) = $\frac{2}{6} = \frac{1}{3}$

You don't know what the probability of 2 on the 2nd draw is unless you know the result of the 1st draw

MS

01/29/2009

PROBABILITY RULES (Cont.)

If you assume that you did not draw a 2 in the 1st draw, then:

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

$$P(2 \text{ on } 1^{st} \text{ draw } \& \ 2 \text{ on } 2^{nd} \text{ draw}) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} \neq 0$$

OUR RULE FOR IID DOES NOT WORK W/ SRS.

W/ SRS, the simple product rule of IID fails because the probability of the 2nd draw having the desired outcome depends on the outcome of the 1st draw.

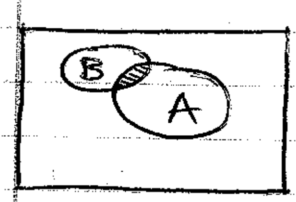
- To complete the "AND" rule for SRS, we need to introduce "GIVEN".

CONDITIONAL PROBABILITY ("GIVEN")

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

(| = "given" (say "bar")

Concept defined by Bayes:



$P(B) = \frac{B}{\Box}$ ← the box stands for all the ways it could come out

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

Saying "given" means that both A & B are true.

- You know that the rock you dropped fell in A; what are the chances that it also fell in B?

$$P(2 \text{ on } 1^{st} \text{ draw } \& \ 2 \text{ on } 2^{nd} \text{ draw}) = P(2 \text{ on } 1^{st}) \times P(2 \text{ on } 2^{nd} \mid 2 \text{ on } 1^{st})$$
$$= \frac{1}{3} \times 0$$
$$= 0$$

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

- The product rule of IID is just a special case of the product rule above.

MS

01/29/2009

In a probability sense, if A and B are independent, then information about A does not help predict information about B.

- IID draws are INDEPENDENT and identically distributed.
  - If A & B are independent, then $P(B|A) = P(B)$
    - $p(A)\,p(B \text{ given } A) = p(A)\,p(B)$

SUMMARY: PROBABILITY RULES →

→ "NOT"

$$P(\text{NOT } A) = 100\% - P(A)$$

→ "OR"

$$P(A \text{ OR } B) = p(A) + p(B) - P(A \text{ AND } B)$$

→ "AND"

$$P(A \text{ AND } B) = p(A) \cdot p(B \text{ GIVEN } A)$$
$$= p(B) \cdot p(A \text{ GIVEN } B)$$

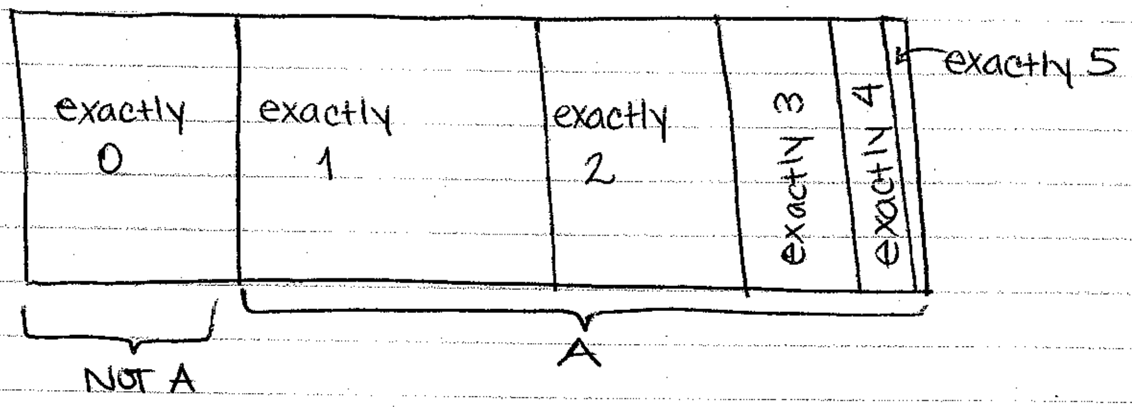
→ "GIVEN"

$$P(B \text{ GIVEN } A) = \frac{P(B \text{ AND } A)}{P(A)}$$

01/29/2009

TS CASE STUDY

P(1 or more TS babies in 5 kids) = 1 − p(0 TS babies)



$1 - p(0 \text{ TS babies}) = 1 - p\left(\begin{matrix}\text{not TS}\\ \text{on } 1^{st}\end{matrix} \;\&\; \begin{matrix}\text{not TS}\\ \text{on } 2^{nd}\end{matrix} \;\&\; \cdots \;\&\; \begin{matrix}\text{not TS}\\ \text{on } 5^{th}\end{matrix}\right)$

Whatever the result of the 1^{st} conception is does not influence the outcomes of the following conceptions → IID
- Biology justifies this assumption of independence.

P(having child w/ TS) =

	H	h
H	HH	Hh
h	Hh	hh

$= \frac{1}{4}$

$P(\text{not TS}) = 1 - \frac{1}{4} = \frac{3}{4}$

$P(1 \text{ or more TS babies}) = 1 - P(0 \text{ TS babies}) = 1 - \left(\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4}\right)$
$= 1 - \left(\frac{3}{4}\right)^5 = 0.76$

76% chance of having 1 or more TS babies.

MS

6/7
01/29/2009

CASE STUDY

UCLA → Are gender & marijuana legalization preference associated in this data set?

Marijuana Legalization Preference	Gender
NO	F
NO	M
YES	F
⋮	⋮
YES	F

$n=106$

2X2 Contingency Table

Mari. Leg. Preference

gender	Yes	No	
female	29	20	49
male	52	5	57
	81	25	106

$P(\text{YES}) = 81/106 = 76\%$

$P(\text{Yes} \mid \text{female}) = 29/49 = 59\%$

$P(\text{Yes} \mid \text{male}) = 52/57 = 91\%$

Strong association between gender & marijuana legalization preference
- dependent variables

Because $91\% \gg 59\%$, there seems to be a strong association between the 2 variables.

MS

7/7
01/29/2009

CASE STUDY → DEATH PENALTY
Does race influence whether the defendant is given the death penalty?

	Death Penalty		
Defendant	Yes	No	
White	19	141	160
Black	17	149	166
	36	290	326

$P(DP) = \frac{36}{326} = 11\%$
$P(DP|DW) = \frac{19}{160} = 11.9\%$
$P(DP|DB) = \frac{17}{166} = 10.2\%$

Seems like gov't not racist
What about PCF's?
RACE OF VICTIM

White Victim

	Death Penalty		
Defendant	Y	N	
White	19	132	151
Black	11	52	63
	30	184	214

Black Victim

	Death Penalty		
Defendant	Y	N	
White	0	9	9
Black	6	97	103
	6	106	112

PCF defeated by keeping it constant.

P(death penalty | Victim White) = $\frac{30}{214}$ = 14%
P(death penalty | Victim white & Defendant White) = $\frac{19}{151}$ = 12.6%
P(death penalty | Victim white & Defendant Black) = $\frac{11}{63}$ = 17.5%

Now it is a lot more visible that race influences the death penalty → more blacks given death penalty when victim white.

P(death penalty | Victim black) = $\frac{6}{112}$ = 5.4%
P(death penalty | Victim black & defendant white) = $\frac{0}{9}$ = 0%
P(death penalty | Victim black & defendant black) = $\frac{6}{103}$ = 5.8%

Here, the data shows the black victim's life is not as important as a white victions or a white defendants life because no white murderers got the death penalty

SIMPSON'S PARADOX: direction of relationship between X & Y changes when the PCF is held constant.