Probability Primer

By

Chris Tufts

@devlintufts

Let's start with an example

I wake up in one of the following states:

Angry — 80% of days OR $p(\text{angry}) = 0.80$

Happy — 20% of days OR $p(\text{happy}) = 0.20$

Chris seems like a very angry person

But there is more to the story....

p(construction) = .85
OR
85% of days there is construction
...
It starts 2 hours before Chris has to be up!!!

I wonder if the construction is connected to his anger???

Let's go over some notation:

$p(---)$ → probability of whatever is in the brackets

$A | B$

↳ This symbol means 'given' as in "A given B"

$A \cap B$

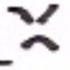
↳ 'AND' as in "A and B"

Conditional Probability

- Probability of an event given another event has occurred

p(A|B) ⇒ probability of Event A **Given** Event B occurred

⋮

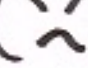
p( | ) ⇒ probability of angry Given construction

Joint Probability

- Probability of 2 [or more] events co-occurring

$p(A \cap B)$ ⇒ Probability of Event A and Event B occurring

⋮

p( ∩ ) ⇒ Probability of angry + construction

{Wait! Wait! Wait!}

> What is the difference between 'Given' and 'both occurring' ???

☐ 'Given' [Conditional] implies that I {KNOW} an event (B) occurred and I can then use this Knowledge to determine the chance of the other event (A)

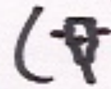
$p(A|B)$
- ↳ B occurred
- ↳ We are guessing A

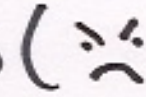
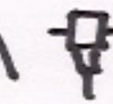
☐ 'Both occurring' [Joint] implies we are estimating the chance of these 2 events occurring. Unlike conditional, we have no Knowledge of either occurring.

So let's say I know the following:

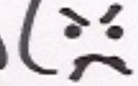
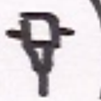
---

p() = 0.8

p() = 0.85

p( ∩ ) = 0.75

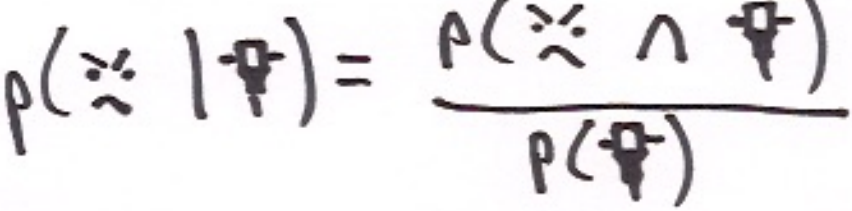
How can I find out:

p( | ) = ?

There is a formula for that!

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Which we can use...



$$= \frac{0.75}{0.85}$$

$$= 0.88$$

WHAT DOES THIS MEAN??

- There is an 88% chance I will wake up angry if construction is happening that morning

BUT There is More ...

- Conditional probabilities sum to <u>1</u>!

This means:

$$1.0 - 0.88$$
$$= .12$$

or

p(☹ | NO ) = .12

[12% chance of being angry <u>if</u> <u>no</u> <u>construction</u> is going on]

Where are these probabilities coming from?!?

- Collect data over 20 day period
- Each day log:

  Construction: yes/no

  Wake up state: Angry / Happy

	2 days	2 days
	15 days	1 days

15+2+2+1 = 20 days

IN OTHER WORDS

- Each cell in the table is the # of days event x and y co-occurred...
  {x ∩ y}

- To find p(x ∩ y) divide by total # of days [20]

  p(☹ ∩ ☂) = 15/20 = 0.75
  p(☹ ∩ ☀̸) = 1/20 = 0.05
  p(☺ ∩ ☂) = 2/20 = .1
  p(☺ ∩ ☀̸) = 2/20 = .1

→ ALL SUM TO 1!

WHAT ABOUT P(☹)?

			
☺	2	2	(4)
☹	15	1	(16)
	(17)	(3)	

- Circled values are the sum of each row/column

* AKA: The total # of days that event occurred

SO ...

$p(☺) = \frac{4}{20} = .2$

$p(☹) = \frac{16}{20} = .8$

} Sums to **1**

$p(💡) = \frac{17}{20} = 0.85$

$p(✱) = \frac{3}{20} = 0.15$

} Sums to **1**

→ Each of these is a <u>MARGINAL</u> Probability

(hence the summing on the table margins...)

Conclusion

- This is a generalized example, but hopefully helps you learn a thing or 2.
- Don't stop here... EXPLORE ALL THE STATS!
- CHECK BACK @devlintufts & miningthedetails.com

For more primer zines about stats SOON!