Probability Primer

By
Chris Tufts
@devlin.tufts
Let's start with an example.

I wake up in one of the following states:

Angry: 80% of days
\[ p(\text{angry}) = 0.80 \]

Happy: 20% of days
\[ p(\text{happy}) = 0.20 \]
Chris seems like a very angry person.

But there is more to the story ...

\[ p(\text{construction}) = 0.85 \]

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:

\[ p(...) \rightarrow \text{probability of whatever is in the brackets} \]

\[ A \mid B \rightarrow \text{This symbol means 'given' as in "A given B"} \]

\[ A \land B \rightarrow \text{"AND" as in "A and B"} \]
Conditional Probability

- Probability of an event given another event has occurred:

\[ p(A \mid B) \Rightarrow \text{probability of Event } A \text{ given Event } B \text{ occurred} \]

\[ p(x \mid \#) \Rightarrow \text{probability of angry given construction} \]
Joint Probability

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

\[ P(A \land B) \]

\[ \Rightarrow \]

Probability of Event A and Event B occurring

\[ P(\neg A \land \neg B) \Rightarrow \]

Probability of angry construction
"Wait! Wait! Wait! Wait!"

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

- 'Given' [Conditional] implies that I **KNOW** an event occurred and I can then use this knowledge to determine the chance of the other event (A)

\[
p(A \mid B) \quad \text{if B occurred}
\]
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:

\[ p(\exists) = 0.8 \]
\[ p(\forall) = 0.85 \]
\[ p(\exists \land \forall) = 0.75 \]

How can I find out:

\[ p(\exists | \forall) = ? \]

There is a formula for that:

\[ p(A|B) = \frac{p(A \land B)}{p(B)} \]
Which we can use...

\[ P(\neg \neg A | B) = \frac{P(\neg \neg A \land B)}{P(B)} \]

\[ = \frac{0.75}{0.85} \]

\[ = 0.88 \]

**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 1!

This means:

\[ 1.0 - 0.88 = 0.12 \]

or

\[ p(\neg \text{NO STUDENT}) = 0.12 \]

[12% chance of being angry if no construction 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

<table>
<thead>
<tr>
<th></th>
<th>2 days</th>
<th>2 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>✗</td>
<td>15 days</td>
<td>1 day</td>
</tr>
</tbody>
</table>

15 + 2 + 2 + 1 = 20 days
IN OTHER WORDS

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

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

\[
\begin{align*}
  p(\text{x} \cap \text{y}) &= \frac{15}{20} = 0.75 \\
  p(\text{x} \cap \text{x}) &= \frac{1}{20} = 0.05 \\
  p(\text{y} \cap \text{y}) &= \frac{3}{20} = 0.1 \\
  p(\text{y} \cap \text{x}) &= \frac{3}{20} = 0.1
\end{align*}
\]

ALL SUM TO 1!
What about $P(\cdot)$?

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>○</td>
<td>2</td>
<td>2</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>×</td>
<td>15</td>
<td>1</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Circled values are the sum of each row/column.
- AKA: The total # of days that event occurred.
SO...

\[
\begin{align*}
p(\mathbf{\text{\ding{162}}} & ) = \frac{4}{20} = 0.2 \\
p(\mathbf{\text{\ding{161}}} & ) = \frac{16}{20} = 0.8 \\
p(\mathbf{\text{\ding{160}}} & ) = \frac{17}{20} = 0.85 \\
p(\mathbf{\text{\ding{159}}} & ) = \frac{3}{20} = 0.15
\end{align*}
\]

Each of these is a MARGINAL 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 @devlinuitts & miningthedetails.com for more primer zines about stats soon!