Conditional Probability Lecture 1

I am sure that by now you are acutely aware that calculating probabilities involves keeping track of all the possible universes within the sample space.
But even with such a simple experiment as tossing a coin, the true sample space is vast!

As you throw a coin, people hurry up and down the street, birds are flying in the sky, grass is breaking out of the ground without anyone to notice how it grows, planets move around the sun, and galaxies hurl away from one another.
What information is relevent to a given problem?
How is probability updated in response to new information?
To address these concerns, we must first observe how the occurence of one event F affects the probability of some other event E. In other words,

Given that F happened, how likely is E to occur? Symbolically, we express this as $P(E|F)$. What should the formula of $P(E|F)$ be in terms of the unrestricted probabilities of E and F?

<u>Ex.</u> Two coins are flipped 20 times. Let E be the event that coin #1 comes up heads. Let F be the event that at least one coin lands up tails. Suppose these were the outcomes:

(T, T)	(H, H)	(H, H)	(T, T)
(H, H)	(H, T)	(T, H)	(H, T)
(H, T)	(H, T)	(H, H)	(H, T)
(H, H)	(H, T)	(H, T)	(H, T)
(H, H)	(T, H)	(T, H)	(T, T)

- Occurance of E = # 13
- Occurance of F = # 14
- Occurance of EF = # 7

A natural idea for estimating $P(E|F)$ is to measure the frequency with which E occurs among outcomes in F. Since E can contain outcomes that are not in F, we must consider EF.

By the frequency argument, $P(E|F) \approx \frac{n(EF)}{n(F)}$ when $n(F)$ - the number of occurences of F is large. In this experiment $P(E|F) \approx \frac{7}{14} = \frac{1}{2}$. Since 14 isn't a large number, this estimate doesn't inspire too much confidence.

By the frequency interpritation of probability,

$$p(E|F) = \lim_{n(F) \to \infty} \frac{n(EF)}{n(F)} = \lim_{n \to \infty} \frac{n(EF)}{n(F)}$$

$$= \lim_{n \to \infty} \frac{\left(\frac{n(EF)}{n}\right)}{\left(\frac{n(F)}{n}\right)} = \frac{P(EF)}{P(F)}$$

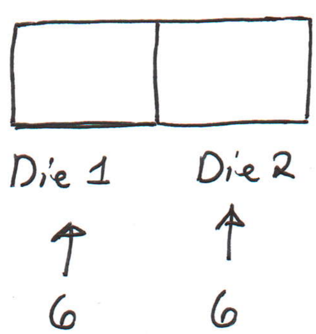
The idea here is that F is the new sample space and as we generate more and more outcomes of F

$(n(F) \to \infty)$, we'll be able to gage the relative frequency with which EF occurs. If $P(F) \neq 0$, $\lim_{n \to \infty} n(F) = \infty$. This allows us to switch to

$$\lim_{n \to \infty} \frac{n(EF)}{n(F)}.$$

Ex. If two fair dice are rolled, what is the probability that the first die lands on a 3 given that the sum of the dice is 8?

Solution: The original sample space may be modeled by the Kafka document



1. We may pick $F = \{(2,6), \boxed{(3,5)}, (4,4), (5,3), (6,2)\}$ As our new sample space and observe that any of its

5 outcomes are equally likely. Only one of these outcomes corresponds to die #1 = 3.

Thus if E - event die #1 = 3, then

$P(E|F) = \frac{1}{5}$.

2. We may also work with the original sample space:

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P((3,5))}{\left(\frac{5}{36}\right)}$$

$$= \frac{\left(\frac{1}{36}\right)}{\left(\frac{5}{36}\right)} = \frac{1}{5}.$$

Ex. A fair coin is flipped twice. What is the conditional probability that both flips land on heads, given that

(a) First flip lands on heads?

(b) At least one flip lands on heads?

Solution: Let B = event both flips land on heads

(a) Let A = event first flip lands on heads

Then $A = \{(H,H), (H,T)\}$. It seems obvious that every outcome in A is equally likely.

Working with the restricted sample space, we obtain

$$P(B|A) = \frac{1}{2}$$

(b) Let L = event at least one flip lands on heads. $L = \{(H,H), (H,T), (T,H)\}$

$$P(B|L) = \frac{1}{3}.$$

Observation: If S is a sample space with equally likely outcomes, then so is the restriction to $F \subset S$.

Proof: Assume $|S| = n$ and $|F| = m \leq n$.

If $x \in F$, $P(x|F) = \dfrac{P(xF)}{P(F)} = \dfrac{P(x)}{P(F)} = \dfrac{\left(\frac{1}{n}\right)}{\left(\frac{m}{n}\right)}$

$$= \frac{1}{m}.$$

Thus our assumptions about the restricted sample space were justified.

Ex. Student is taking a 1 hr exam. The probability the student is finished with exam in less than x hours is $\frac{x}{2}$ for all $0 \leq x \leq 1$. Given that the student is still working after 0.75 hours, what is the probability that the full hour is used?

Solution: Let F_x be the event the student is finished in less than x hours. We want

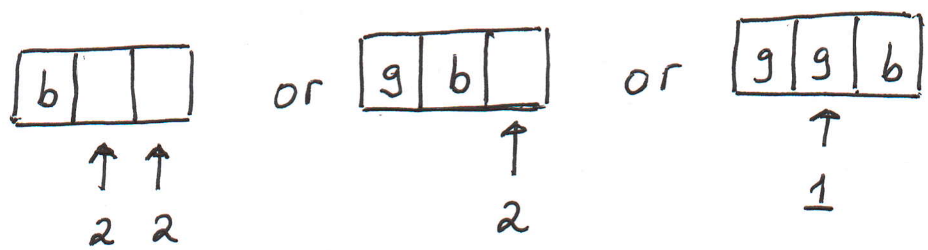
$$P(F_1^c \mid F_{0.75}^c) = \frac{P(F_1^c F_{0.75}^c)}{P(F_{0.75}^c)} = \frac{P(F_1^c)}{P(F_{0.75}^c)}$$

$$= \frac{1 - P(F_1)}{1 - P(F_{0.75})} = \frac{1 - \frac{1}{2}}{1 - \frac{3}{8}} = \frac{4}{8-3} = \frac{4}{5}$$

$$= 0.8$$

Ex. The Crown prince is always the eldest male offspring in the royal family. If the crown prince is one of 3 siblings, what is the probability that he is the oldest child?

Solution: It is convenient to use a restricted sample space, in which at least one of the 3 children is a boy.

F: 

Thus $|F| = 4 + 2 + 1$. In words, either the crown prince is the oldest child, or the middle child, or the youngest child. There are 4 outcomes corresponding to the event where he is the oldest child.

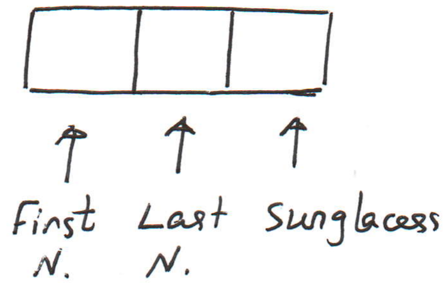
Thus, since each outcome of F is equally likely,

$p(C|F) = \frac{4}{7}$ where C- event crown prince is the oldest child.

As we go along, remember, probability is the devil's tongue. Be constantly on your guard!

Ex. Young Dolores sees a strange man - her mother's new tenant - looking at her through the living room window. Each of his first and last names is either Humbert or Quilty. She knows one of them is Humbert. What is the probability that his full name is Humbert Humbert if she is wearing one of 3 pairs of her favorite sun glasses - pair 1?

Solution: The restricted sample space is of the form $\{(H,Q,1), (Q,H,1), (H,H,1)\}$ where the original sample space was



Since we assume that each outcome in the sample space is equally likely, each of the 3 outcomes in the restricted sample space are also equally likely. Thus the desired probability is $\frac{1}{3}$.

<u>Ex.</u> Suppose your neighbor has two dogs. Each dog is either white or brown. Each of them has one of 3 favorite dishes. What is the probability that both dogs are brown if you see a brown dog eating its favorite dish in your neighbor's yard?

<u>Solution:</u> Assume the 3 possible dishes for both dogs are 1 - IAMS 2 - Kibbles 'n Bits
3 - Pedigree.

Assume also that each dog is equally likely to be either b-Brown or w-White and equally likely to have any of the 3 food choices as a first preference.

The Kafka protocol for the unrestricted space is

b	1	b	3
Old Dog	Choice	Young.	Choice
2 pos.	3 pos.	2 pos.	3 pos.

Our observation is either $(b, 1, \square, \square)$ or $(\square, \square, b, 1)$. The number of possible outcomes in this observation is $2 \cdot 3 + 2 \cdot 3 - 1 = 11$, where we subtract 1, because of counting the outcome $(b, 1, b, 1)$ twice. For both dogs to be brown, the outcome must be of the form $(b, 1, b, \square)$ or $(b, \square, b, 1)$, There are $3 + 3 - 1 = 5$ such outcomes. Hence the probability both dogs are brown is $\frac{5}{11}$.

Why are so many people making errors in computing these probabilities? A major mistake is to seek a causal relationship between the sunglasses and the person's name or between a dog's food and its color, No! The objects that probability theory counts are possible universes, and not individual features of these. Every bit of information gives some hint about the universe that will come to be. Every bit of information narrows down the set of possibilities.

Q. Why then does some information affect probability and other information does not?

A. For any event E in a sample space S

$$P(E) = \frac{\#\text{ universes with property } E}{\#\text{ total universes}}$$

(Assume for simplicity that each potential universe is equally likely). When the restricted sample space respects this proportion, the probability isn't altered.

<u>Ex.</u> If two fair dice are rolled, what is

(a) The probability the first die is a 6?

(b) The first die is a 6, given that the sum on the dice is 7?

<u>Solution:</u>

(a) The possible universes here are outcomes on both dice: $(\underset{\substack{\uparrow \\ 6\text{ pos.}}}{x}, \underset{\substack{\uparrow \\ 6\text{ pos.}}}{y})$

The desired outcomes are of the form $(6, \underset{\substack{\uparrow \\ 6\text{ pos.}}}{y})$

Thus the probability that first die is 6 is

$$\frac{6}{36} = \frac{1}{6}.$$

(b) On the other hand, if the sum on the dice is 7, the restricted sample space is

$$F = \{\boxed{(6,1)}, (5,2), (4,3), (3,4), (2,5), (1,6)\}$$

Only 1 universe corresponds (carries) to the desired outcome.

$$p(\text{Die}1 = 6 \mid F) = \frac{1}{6}$$

Do you see?

$\frac{6}{36}$ ← more desired universes / ← more possible universes

$\frac{1}{6}$ ← less desired universe types / ← fewer overall possibilities.

Remark: In part (a) it is tempting to ignore the second die and simply consider the possible universes as displaying the first die alone:

$$(x, \cancel{y}) \longrightarrow x.$$

This will lead to the correct solution, but through incorrect reasoning. The universe is

Die 1 $\leftrightarrow$ Die 2, not Die 1!

Rather, imagine that the value of Die 2 is known to you, that it is held fixed.

e.g. $(x, y) \longrightarrow (x, 1)$

You may suppress what is held constant and write x instead of $(x, 1)$.

After we have developed more machinery, we will justify this extra assumption.