Topic 2: Introduction to Probability

Optional Reading: Chapter 4

Xiner Zhou
Department of Statistics
University of California, Davis

Part II : Conditional Probability

- Definition of conditional probability
- Bayes Rule
- Law of Total Probability

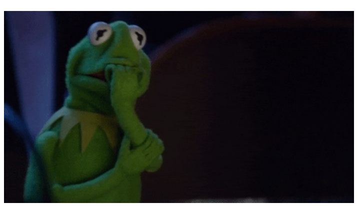
Imagine being awakened one night by the sound of your burglar alarm.

What is the probability or your degree of belief that a burglary attempt has taken place, i.e. someone has broken into your place?

A patient is tested for a disease that affects only 1% of the population.

The test result is positive, i.e. the test claims that the patient has the disease.

What is the probability that he indeed has the disease, given that the evidence provided by test result?



Bad things must be true!

We have introduced probability as a language for expressing uncertainty about events.

Probability can also be interpreted as “the degree of belief or dis-belief” of some claims or events.

Now we have a new situation:

- Whenever we observe new evidence, in the form of data/information (alarm, positive test result….)
- We acquire new information that may affect our belief about the events

How should we update our belief about an event in light of new evidence we observe?

- Hearing alarm must changed our belief about whether burglary happened or not
- Test postive must changed our belief about whether the patient has the disease or not

Conditional Probability is the concept and the tool that address this fundamental question.

"Conditional Probability is the soul of Statistics."

1. How to incorporate evidence into our understanding of the world in a logical, coherent manner

   It's the language of many scientific/ medical / legal reasoning

2. Conditioning is a very powerful problem-solving strategy

   Often make it possible to solve a complicated problem by decomposing it into manageable pieces with case-by-case or condition-by-condition reasoning.

Conditional probability

If A and B are events with P(B)>0,

Then the conditional probability of A given B, denoted by P(A|B) is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- To find the conditional probability that event A occurs given that event B occurs, divide the probability that both A and B occur by the probability that B occurs
- In general, A and B can be any events
- But it's important to interpret:
    - A is the event whose uncertainty we care about
    - B is the evidence we observe, such as burglary alarm, positive test

Suppose A is the event that a person is colorblind, B is the event that a person is men, our population composition is shown as:

	Men(B)	Women (B^C)	Total
Colorblind (A)	.04	.002	.042
Not Colorblind (A^C)	.47	.488	.958
Total	.51	.49	1.00

What is the risk of colorblind for men?

$$P(colorblind \mid man) = ?$$

$$P(colorblind \mid man) = \frac{P(\text{colorblind} \cap \text{man})}{P(\text{man})} = \frac{0.04}{0.51} = 0.078$$

What is the risk of colorblind for women?

$$P(colorblind \mid woman) = ?$$

$$P(colorblind \mid woman) = \frac{P(\text{colorblind} \cap \text{woman})}{P(\text{woman})} = \frac{0.02}{0.49} = 0.004$$

=> Risk for men is higher for women!

Here Gender provides extra information about the risk of colorblindness,

therefore is a source of evidence, that explains why we condition on gender.

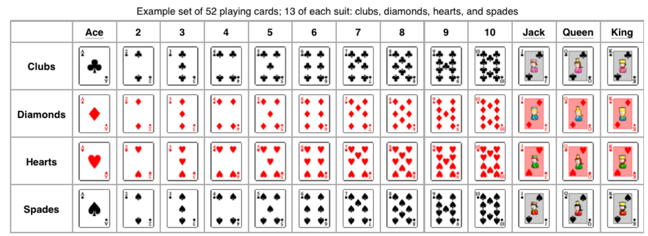
A standard deck of cards is well-shuffled.

Two cards are drawn randomly, one at a time without replacement.

Let A be the event that 1st card is a heart,

B be the event that 2nd card is a red.

Find P(B|A)



$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A \cap B) = \frac{|A \cap B|}{|S|} = \frac{13 \times (26-1)}{52 \times 51} = 25/204$$

$$P(A) = \frac{|A|}{|S|} = \frac{13}{52} = \frac{1}{4}$$

$$\Rightarrow P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{25/204}{1/4} = \frac{25}{51}$$

What the order card is actually tell you extra information!

Compare P(B) vs P(B|A)

$$P(B) = \frac{|B|}{|S|} = \frac{26}{52} = \frac{1}{2}$$

That’s why we condition on the other card as a source of evidence!

(Two Children problem from Scientific American 1950s)

Mr. Jones has 2 children, the order child is a girl,

what is the probability that both are girls?

Mr. Smith has 2 children, at least one of them is a boy,

What is the probability that both children are boys?

Intuition? Definition? True Understanding of Conditional Probability?

S={GG, GB, BG, BB}
For Mr. Jones, intuitively if we older is girl, then in order to have both girls, we only need to know what's the probability that younger is a girl, which is 1/2. Let's check our intuition:

$$P(\text{both girls}|\text{elder is girl}) = \frac{P(\text{both girls, elder is girl})}{P(\text{elder is girl})} = \frac{1/4}{1/2} = \frac{1}{2}$$

For Mr Smith,

$$P(\text{both girls}|\text{at least one is girl}) = \frac{P(\text{both girls, at least one is girl})}{P(\text{at least one is girl})} = \frac{1/4}{3/4} = \frac{1}{3}$$

=> So for both Mr Jone and Mr Smith, since they provide some extra information / data/ evidence, their P(both girls| evidence) are all different from P(both girls)=1/4 without any information.

Let's re-arrange the definition of conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability of intersection of two events

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

The probability of both events happen can also be calculated by the multiply two probabilities:

- conditional probability of A given B
- Probability of B

We will use it as a tool for Bayes' rule and other rules.

Bayes' Rule

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Proof: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B \mid A)P(A)}{P(B)}$

Why the trouble? Why substitute $P(A \cap B)$ by $P(B \mid A)P(A)$?

- In many applications, we are asked to calculate conditional probability P(AIB)
- So by definition we need $P(A \cap B)$
- But $P(A \cap B)$ is often very hard to calculate directly, unless you're all-knowing God
- It turns out the other way of conditional probability $P(B \mid A)$ is often very easy!

Bayes' rule is extremely famous, one of most useful result!

One more problem with Bayes' rule:

Often, $P(B)$ in Bayes' rule is still hard to calculate directly

➡ The second main theorem of this part:

The Law of Total Probability (LOTP) will complete the main goal of calculating any conditional probability

"Conditional Probability is the soul of Statistics."

1. How to incorporate evidence into our understanding of the world in a logical, coherent manner

   It's the language of many scientific/ medical / legal reasoning

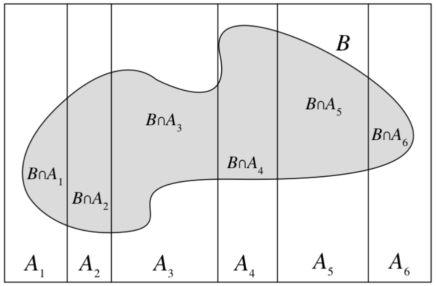
2. Conditioning is a very powerful problem-solving strategy

   Often make it possible to solve a complicated problem by decomposing it into manageable pieces with case-by-case or condition-by-condition reasoning.

The Law of Total Probability

Let n events $A_1 \ldots A_n$ be a partition of sample space S,

i.e. partition: $A_1 \ldots A_n$ are disjoint events, but their union is the whole sample space S



Then:

$$P(B) = \sum_{i=1}^{n} P\left(B, A_i\right) = \sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right) = P\left(B \mid A_1\right) P\left(A_1\right) + P\left(B \mid A_2\right) P\left(A_2\right) + \cdots + P\left(B \mid A_n\right) P\ (A$$

When direct calculation of the probability of some event B is difficult,
we can divide the sample space into slices,
Find the conditional probability in each slices,
Then take a weighted sum of these conditional probabilities, weights being how large the slices are

The choice of how to slice S is crucial: a well-chosen partition will reduce a complicated problem into a simple one.

It's often used together with Bayes' rule, esp. for Bayes' denominator calculation.

Combining:

- the definition of conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- probability of intersection of two events

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

- Law of Total Probability

$$P(B) = \sum_{i=1}^{n} P\left(B, A_i\right) = \sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right)$$

The final version/most useful version of Baye's rule:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{\sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right)}$$

Imagine being awakened one night by the sound of your burglar alarm.

What is the probability or your degree of belief that a burglary attempt has taken place, i.e. someone has broken into your place?

Data has been collected and some evidence provided as:

- There is a 95% chance that a burglary will trigger that alarm

$P(\text{alarm}|\text{Burglary})=0.95$

(This can be easily obtained by testing the alarm system,

On the other hand, $P(\text{burglary}|\text{alarm})$ is not easily testable…)

- There is a slight 1% change that alarm will be triggered by a mechanism other than burglary, such as a bug

$P(\text{alarm}|\text{no burglary})=0.01$

- Previous crime pattern in the region indicate that there is 0.0001 chance of burglary

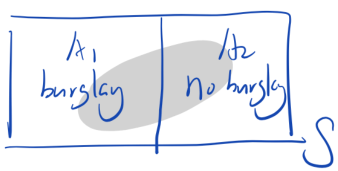
$P(\text{burglary})=0.0001$

Solution:

By Bayes' rule $P(burglary \mid alarm) = \dfrac{P(alarm \mid burglary)P(burglary)}{P(alarm)}$

How to know P(alarm)?

LOTP:



divide sample space into A_1 = burglary , A_2 = no burglary

=> P(alarm)=P(alarm|burglary)P(burglary)+ P(alarm|no burglary)P(no burglary)
= 0.95 * 0.0001 +0.01*(1-0.0001)
= 0.010094

=> $P(burglary \mid alarm) = \dfrac{P(alarm \mid burglary)P(burglary)}{P(alarm)} = \dfrac{0.95 \times 0.0001}{0.010094} = 0.00941$

SO, still low probability of burglary, even though the alarm triggered
The evidence provided by alarm triggered has increased the probability of burglary almost 1000-fold!

A patient is tested for a disease that affects only 1% of the population.

The test result is positive, i.e. the test claims that the patient has the disease.

What is the probability that he indeed has the disease, given that the evidence provided by test result?

We know additional evidence that:

$$P(D) = 0.01$$

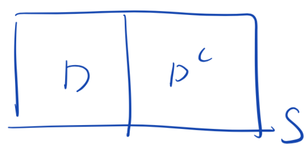
$$P(T \mid D) = 0.95$$

$$P(T^C \mid D^C) = 0.95$$

Solution:

By Bayes' rule $P(D \mid T) = \dfrac{P(T \mid D)P(D)}{P(T)}$

LOTP : $P(T) = P(T|D)P(D) + P(T|D^c)P(D^c)$



$$\Rightarrow P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T)} = \frac{P(T \mid D)P(D)}{P(T \mid D)P(D) + P(T \mid D^C)P(D^C)} = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16$$

There is only a 16% chance that the patient has the disease, given that he tested positive, even though the test seems to be very reliable with accuracy 95%!

For rare disease, most people who tested positive don't actually have the disease!