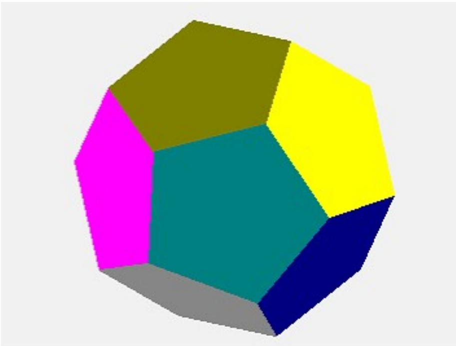
Probability Theory

An introduction and beyond

Chapter 3

Conditional probabilities

This is an article from my home page: www.olewitthansen.dk



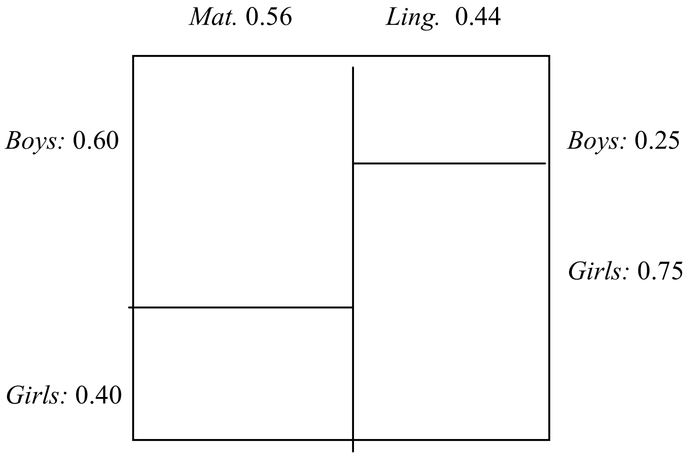
Ole Witt-Hansen (2017) 1997

Contents

1. Definition of a conditional probability

Earlier in the Danish 9 – 12 grade high school (Gymnasium) the students could choose between entering the mathematical branch or the linguistic branch.

In one Gymnasium 56% of the students are mathematicians and 44% are linguists.
Among the mathematicians 60% are boys, whereas among the linguists there are 25% boys.
This is schematically illustrated below.



Then we consider an experiment, where we at random select a student. We shall introduce some shorthand notations for various outcomes of the experiment.

M: A mathematician is selected.
L: A linguist is selected.
B: A boy is selected.
G: A girl is selected

Statistically we would expect the first two probabilities: $P(M) = 0.56$ and $P(L) = 0.44$.

However we can not directly find the probabilities $P(B)$ and $P(G)$ for selecting a boy or a girl, since the two events are distributed among the events mathematician or linguist.

However, if a boy is selected *only* among the mathematicians the probability is 0.60.
This we express in the following manner:

The conditional probability to select a boy, on the condition that he is a mathematician is 0.60

This is an example of what we in probability theory call a *conditional probability* and it is written symbolically as:

$$P(B \mid M) = 0.60$$

The vertical line between the two events reads: ”Given” or ”On the condition that”, and the probability $P(B \mid M)$ is then read as:

$P(B \mid M)$: *The probability of selecting a boy, given he is a mathematician.*

Or:

$P(B \mid M)$*:* The probability of selecting a boy on the *condition* that he is a mathematician.

In the same manner we may define the probabilities:

$$P(G \mid M) = 0.40\ ,\quad P(B \mid L) = 0.25\ ,\quad P(G \mid L) = 0.75$$

We shall then look into, how we may establish conditional probabilities in a symmetric outcome space, using the probability $P(B \mid M)$ as an example.
As usual we use the notation $n(H)$ for the number of elements (outcomes) in an event (subset) H, and $n(U)$ for the total number of outcomes.

The event "A mathematician boy" is then written $B \cap M$, and the probability $P(D \mid M)$ of this event is as usual calculated as *favourable outcomes/possible outcomes.* That is, the number of mathematical boys divided by the number of mathematicians.

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

Dividing the numerator and the denominator of the right hand side of the equation with $n(U)$, we get:

$$P(B \mid M) = \frac{n(B \cap M)}{n(M)} = \frac{\dfrac{n(B \cap M)}{n(u)}}{\dfrac{n(M)}{n(U)}} = \frac{P(B \cap M)}{P(M)}$$

This leads to the definition:

For two arbitrary events A and $H \neq \emptyset$ in any (not necessarily symmetric) outcome space, we define the conditional probability of A given H (that is, H has occurred) as:

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

The probability for *A: given H*, is the probability that both A and H occur, divided by the probability that H occur.

Often the formula is used when multiplying with $P(H)$

$$P(A \cap H) = P(A \mid H)\, P(H)$$

Since $A \cap H = H \cap A$, then we must also have $P(A \cap H) = P(H \cap A)$ and consequently:

$$P(A \mid H)\, P(H) = P(H \mid A)\, P(A) \quad \Leftrightarrow \quad P(A \mid H) = \frac{P(H \mid A) P(A)}{P(H)}$$

This (controversial) formula is called Bayes formula. Mathematically it is completely straight, but in some cases, when applied to probabilities of events in real life, it seem to violate the principle of causality, in the sense that it can predict the probability of an event that has occurred, which does not make sense.

If $P(A)$ is the probability of an accident when driving on the highway in winter times, and $P(H)$ is the probability of an icy road, then $P(A|H)$ is the probability of having an accident if the road is icy.

So far so well, but if we use Bayes' formula without reservations, then when it is turned the other way round to calculate $P(H|A)$, which is mathematically perfectly legal, it predicts the probability of a icy road, given there has been an accident. The crucial point is of course, that when the accident has happened, we know whether the road was icy or not, and therefore the notion of probability is meaningless.
Statistically, however, it will probably be true that the number of accidents on icy roads corresponds to the statistically percentage of days with icy roads.

It is true that Bayes' theorem has caused much (mostly philosophical) debate, of which we shall not participate, but in probability theory it is established as a most fundamental theorem.

Examples:

1. A card is drawn from a deck of cards. You are informed that it is a card of hearts (H).
Find the probability that it is a picture card (B).
The question may be answered in two different ways.
The number of picture cards in hearts is 3 and there are 13 cards of hearts. From the definition of probability the answer is therefore: $P(\mathrm{B} \mid \mathrm{H}) = \frac{3}{13}$.
But the question may also be calculated from the formula defining conditional probabilities:

$$P(B|H) = \frac{P(B \cap H)}{P(H)} = \frac{\frac{3}{52}}{\frac{13}{52}} = \frac{3}{13}$$

2. For families having 3 children, we want to find the probabilities that:
a) Among the children there are both a girl and a boy (A).
For the gender of 3 children there are $2^3 = 8$ equal possibilities, since there are 2 possibilities for the first child and so on.
There are two outcomes (all boys and all girls), where this is not the case, so $P(\mathrm{A}) = \frac{6}{8} = \frac{3}{4}$.
b) Same question, but where it is informed that they have a girl. (G).
Thus we must determine the probability $P(\mathrm{A} \mid \mathrm{G})$, where we notice that $A \cap P = A$, $n(A) = 6$ and $n(\mathrm{G}) = 7$ (8 minus the outcome of 3 boys).

$$P(A \mid G) = \frac{P(A \cap G)}{P(G)} = \frac{\frac{6}{8}}{\frac{7}{8}} = \frac{6}{7}$$

3. This example does not really concern conditional probabilities, but erroneously is ascribed to conditional probability. On the contrary, it is a notorious riddle, which also have appeared the magazine for teachers in the Danish 9-12 grade high school, where it rather surprisingly has given rise to lengthy debates!
The problem can for example be stated as:
There are three boxes. The two of them are empty, while the third one contains a gold coin.
A person is asked to select one of the boxes. The box is not opened, but after the choice has been made, one of the remaining boxes (which does not have the gold coin) is removed.
The person is asked if he wants to redo the selection?
Within the framework of probability: Should he undo the selection or not, or is it without importance?
Most people are perhaps inclined to answer that it is does not matter, since the gold coin may be in any of the three boxes. But this is erroneous for the following reasons. By the first choice a box was chosen with probability 1/3. The probability the coin is in one of the other two boxes are therefore 2/3. These probabilities

can not be changed, by removing a box, so the probability that the coin is in one of the two remaining boxes is still 2/3. Since the coin is not in the box that has been removed, then the probability that the coin is in the remaining box must be 2/3. Consequently, one should always redo the choice, since the probability of selecting the box with the coin is doubled from 1/3 to 2/3.

2. Developing probabilities on events

We still need to answer the question in the leading example on conditional probabilities from the preceding section concerning the high school students. What is the probability to select a boy or a girl, since this probability is masked in some conditionally probabilities.

So, we set out for finding the probability $P(B)$ that the selected student is a boy.
We know the conditional probabilities $P(B \mid M) = 0.60$ and $P(B \mid S) = 0.25$, together with the probabilities that the student is mathematician or linguist: $P(M) = 0.56$ and $P(L) = 0.44$.
We then "expand" the subset B (*boy*) on the subsets M (*mathematician*) and L (*linguist*).

$$B = (B \cap M) \cup (B \cap L)$$

A boy is either a mathematician or linguist, so the two sets $(B \cap M)$ and $(B \cap S)$ are disjoint. (having no common elements), and we may apply the simplified form for adding the probability of the union of the two sets.

$$P(B) = P(B \cap M) + P(B \cap L)$$

We then apply the formula for conditional probability: $P(A \cap B) = P(A \mid B) \cdot P(B)$ on each term.

$$P(B) = P(B \mid M) \cdot P(M) + P(B \mid L) \cdot P(L)$$

Generally this formula can only be applied when $M \cup L = U$ and $M \cap L = \emptyset$ in which case M and L are said to form a class division of U.
The formula has an obvious generalization to a class division having n subsets.

The subsets $H_1, H_2, H_3, \dots, H_n$ are said to form a class division of a set U, if it upholds the two conditions;

$$U = H_1 \cup H_2 \cup H_3 \cup \dots \cup H_n \quad \text{and} \quad H_i \cap H_j = \emptyset \quad \text{for every } i \neq j.$$

It then follows:

$$A = (A \cap H_1) \cup (A \cap H_2) \cup \dots \cup (A \cap H_n) \quad \Rightarrow \quad P(A) = P(A \cap H_1) + P(A \cap H_2) + \dots + P(A \cap H_n)$$

And using the definition of conditional probability, we obtain a general and very important formula.

$$P(A) = P(A \mid H_1) \cdot P(H_1) + P(A \mid H_2) \cdot P(H_2) + P(A \mid H_3) P(H_3) + \dots + P(A \mid H_n) P(H_n)$$

Using this formula, brings us in a position to calculate the announced probability $P(B)$ of selecting a boy in the class, using the probabilities from the scheme.

$$P(B) = P(B \mid M) \cdot P(M) + P(B \mid L) \cdot P(L) \quad \Leftrightarrow \quad P(B) = 0.60 \cdot 0.56 + 0.25 \cdot 0.44 = 0.446$$

Examples

1. We shall elaborate a little on the example of selecting a student. It is stated that the student is a boy. We wish to find the probability that he is linguist. What we wish to find is the probability $P(L \mid B)$. Since we only know the inverse probability, we apply Bayes' formula:

$$P(L|B)=\frac{P(B|L)\cdot P(L)}{P(B)}=\frac{0.25\cdot 0.44}{0.446}=0.247$$

2. The next two exercises are notorious ones, appearing in a mathematical textbook, that was used in the Danish Gymnasium until the 1988 reform. Very few, (including teachers) succeeded in giving the right answer.

   There is 75% probability that a player will lie, when giving an answer. He rolls a dice, and he is asked if it showed 6 eyes.
   a) What is the probability that he answers yes?
   b) What is the probability that it showed 6 eyes, if he answered yes?
   c) Which answer (yes or no) gives the greatest probability that the dice actually showed 6 eyes.
   To find the answer we expand the probability on the events "6 eyes" or "not 6 eyes", and use Bayes' formula:

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

a) $P(\text{yes}) = P(\text{yes} \mid \text{not } 6)\, P(\text{not } 6) + P(\text{yes} \mid 6)\, P(6)$

$$P(yes)=\frac{3}{4}\cdot\frac{5}{6}+\frac{1}{4}\cdot\frac{1}{6}=\frac{16}{24}=\frac{2}{3}$$

b)

$$P(6|yes)=\frac{P(6\cap yes)}{P(yes)}=\frac{P(yes\cap 6)}{P(yes)}=\frac{P(yes|6)P(6)}{P(yes)}=\frac{\frac{1}{4}\cdot\frac{1}{6}}{\frac{2}{3}}=\frac{1}{16}$$

c)

$$P(6|no)=\frac{P(6\cap no)}{P(no)}=\frac{P(no\cap 6)}{P(no)}=\frac{P(no|6)P(6)}{P(no)}=\frac{\frac{3}{4}\cdot\frac{1}{6}}{1-\frac{2}{3}}=\frac{3}{8}$$

3. Olga makes a certain game of patience, which statistically comes out every twentieth time.
But she has noticed that if she makes a little certain harmless cheat there are equal chances that it comes out, as the opposite.
Olga's game of patience on the average comes out every tenth time.
What is the probability that Olga cheats the nest time she sets up a game of patience?

The answer to this riddle is a bit tricky, because you may not write the answer as the left side of an equation, as in the pervious examples.
First we write down the probabilities that appears from the formulation.
$P(\textit{comes out} \mid \text{no cheat}) = 0.05$. $P(\textit{comes out} \mid \text{cheat}) = 0.5$, $P(\textit{comes out}) = 0.1$. We want to find the probability $P(\textit{cheat})$. So we write down an equation which contains the information we have, except for one unknown.

$$P(comes\ out) = P(comes\ out\ |cheat)P(cheat) + P(comes\ out\ |no\,cheat)P(no\,cheat)$$
$$0.1 \quad = 0.5P(cheat) + 0.05P(no\ cheat) \quad \Leftrightarrow \quad 0.1 \quad = 0.5P(cheat) + 0.05(1 - P(cheat))$$

The last equation is solved to give: $P(\textit{cheat}) = 1/9$.

Exercises

1. Two cards are drawn from a pile consisting of aces and king from a normal deck of cards. Find the probability that both cards are aces, when it is informed that
   a) On card is an ace,
   b) One card is a black ace,
   c) One card is an ace of spades.

2. A physician thinks that there are 80% probability that a patient suffers from the disease S. He submits the patient to a clinical test, where there are 90% probability of a positive reaction if the patient has the disease S and 15% probability of a positive reaction if the patient does not have the disease S.

a) What probability does the physician thinks there are for a positive reaction of the test? b) The patient has a positive reaction. With what probability do the physician now thinks that the patient suffer from the disease.

3. **20% OF THE POPULATION ARE LEFT HANDED. FROM THE LEFT HANDED 45% ARE RIGHT/LEFT CONFUSED WHILE AMONG THE RIGHT HANDED PERSONS, IT IS ONLY 15%. FIND THE POSSIBILITY THAT A PERSON SELECTED AT RANDOM IS RIGHT/LEFT CONFUSED. A PERSON SELECTED AT RANDOM TURNS OUT TO BE RIGHT/LEFT CONFUSED. FIND THE PROBABILITY THAT HE IS RIGHT HANDED.**

3. Independent events

We have not yet presented a general formula for the "*both…and*" event, that is, to calculate the probability $P(A \cap B)$, when $P(A)$ and $P(B)$ are known. Without resort to probability theory most people would probably multiply the two probabilities.

- *But this is only true if the two events A and B are independent of each other.*

Example

If you draw a card from a deck of cards, and look at the events: *A*: A picture card. *B*: A red picture card. *C:* A card of hearts. Then both events $A \cap C$ and $B \cap C$ are a picture card of hearts.

$$P(A \cap C) = P(B \cap C) = 3/52$$

However:

$$P(A) \cdot P(C) = 12/52 \cdot 13/52 = 12/52\ ¼ = 3 \cdot 1/52 = P(A \cap C),$$

while

$$P(B) \cdot P(C) = 6/52 \cdot 13/52 = 3/26 \cdot 1/4 = 3/2 \cdot 1/52 \neq P(B \cap C)$$

So the two events *A* and *C* are independent of each other, while *B* and *C* are not. This is quite reasonable, since there is the same number of picture cards in all suits, whereas only half of the red cards are hearts.

We may get a better understanding of this, if we review the definition of conditional probability.

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

Which is the correct method of calculating $P(A \cap B)$ in all cases. The concept of *independent events* rely on the following definition..

- *An event A is said to be independent of an event B,* if and only if: $P(A) = P(A \mid B)$.

Replacing this expression for $P(A)$ in the definition of the conditional probability we have:

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

So two events are (in the mathematical sense) independent if

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

Sometimes it can be quite difficult to decide, whether two events are independent or not.
The question can however be settled from calculating $P(A \cap B)$ and compare it to $P(A) \cdot P(B)$.

However, if the aim was to calculate $P(A \cap B)$ from the knowledge of $P(A)$ and $P(B)$, you must find another way to settle the question of independence.

Since the formula $P(A \cap B) = P(A)\,P(B)$ is symmetric in A and B, then if A is independent of B then B is also independent of A. We hereafter say that the two events are independent (of each other).

Examples

1. It is a common fallacy (among non mathematicians) to assume that if you for example throw a coin, and if you have had 5 occurrences of tails, then there is a greater probability to throw heads in the 6th throw.
   But this is not the case, (if you don't believe in the existence of a godness for luck). Any throw with a coin is independent of the previous throws, and it has the same probability.

   A quite similar fallacy occurs, when playing at the roulette or just play on red and black. It is not very likely that red will come out 8 times in a row, but it happens, (also for players using the Martingale system, doubling up on red each time black comes out). The probability is easily calculated as $1/2 \cdot 1/2 \cdot \ldots \cdot 1/2 = 1/2^8 = 1/256$.
   Similarly it is not unusual to presume that the probability of getting a boy child is greater, if you already have 4 girls. (The opposite is perhaps even more likely due to genetic causes).
   It is certainly unlikely to throw tails 5 times in a row, but it is just as likely as throwing any other sequence.
   More precisely we have $P(5 \text{ tails}) = 1/2^5 = 1/32$, $P(\text{at least 1 head}) = 1 - P(5 \text{ tails}) = 1 - \frac{1}{32} = \frac{31}{32}$

2. The probability of getting 6 eyes, when rolling a dice is 1/6. On the average one should therefore expect to obtain at least 6 eyes in 6 rolls. We shall determine the probability not to roll 6 eyes once in 6 attempts. The probability of not throwing 6 eyes is = 5/6. Since the 6 throws are assumed to be independent of each other the answer must be: $5/6 \cdot 5/6 \cdot 5/6 \cdot 5/6 \cdot 5/6 \cdot 5/6$. So the result is: $P(\text{no 6 in 6 rolls}) = \left(\frac{5}{6}\right)^6 = 0{,}335$

3. Some times it can be quite difficult logically to decide, whether two events are dependent or not. In connection with rolling two dices we consider the following events:

   A: First throw shows 4 eyes.
   B: The sum of the eyes on the dices is 7.
   C: The sum of the eyes on the dices is 5.

   Then we find:

$$P(A)=\frac{1}{6} \quad P(B)=\frac{6}{36}=\frac{1}{6} \quad P(C)=\frac{4}{36}=\frac{1}{9} \quad P(A \cap B)=\frac{1}{36} \quad P(A \cap C)=\frac{1}{36}$$

From this we conclude that A and B are independent, but not A and C
This can be explained as, no matter what the first throw is, it will be possible the make the sum 7 eyes, so they are independent, but if the sum must be five it will exclude 5, and 6 in the first throw, so they are dependent.