Queen Mary
University of London

MAS 108 — Probability I

Notes 4 — Autumn 2005

Conditional probability

Example I have four pens in my satchel; they are red, green, blue, and purple. I sample two pens. Let A be the event that the first pen is red or green, and B the event that the second pen is red or green.

We have seen that if the sampling is done with replacement then $P(A) = P(B) = 1/2$ and $P(A \cap B) = 1/4$. Informally this can be expressed as "If I know that A has happened then $P(B) = 1/2$". On the other hand, if the sampling is done without replacement then $P(A) = P(B) = 1/2$ but $P(A \cap B) = 1/6$: in informal terms, "If I know that A has happened then $P(B) = 1/3$".

Conditional probability is a way of making precise the idea that the probability of an event can appear to change if you have some extra information.

Let A be an event with non-zero probability, and let B be any event. The *conditional probability of B given A* is defined as

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

Again I emphasise that this is the definition. If you are asked for the definition of conditional probability, it is not enough to say "the probability of B given that A has occurred", although this is the best way to understand it. There is no reason why event A should occur before event B!

Note the *vertical* bar in the notation. This is $P(B \mid A)$, not $P(B/A)$ or $P(B \setminus A)$.

Note also that the definition only applies in the case where $P(A)$ is not equal to zero, since we have to divide by it, and this would make no sense if $P(B) = 0$.

To check the formula in our example:

$$\text{with replacement} \qquad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2},$$

$$\text{without replacement} \qquad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Conditional probability is used in two distinct ways:

(a) given $P(A)$ and $P(A \cap B)$, calculate the conditional probability $P(B \mid A)$;

(b) given $P(A)$ and $P(B \mid A)$, calculate $P(A \cap B)$ from the rule

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

Example Alice and Bob are going out to dinner. They toss a fair coin 'best of three' to decide who pays: if there are more heads than tails in the three tosses then Alice pays, otherwise Bob pays.

Clearly each has a 50% chance of paying. The sample space is

$$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

and the events 'Alice pays' and 'Bob pays' are respectively

$$A = \{HHH, HHT, HTH, THH\},$$
$$B = \{HTT, THT, TTH, TTT\}.$$

They toss the coin once and the result is heads; call this event E. How should we now reassess their chances? We have

$$E = \{HHH, HHT, HTH, HTT\},$$

so

$$P(A \mid E) = \frac{P(A \cap E)}{P(E)} = \frac{P(\{HHH,\ HHT,\ THH\})}{1/2} = \frac{3/8}{1/2} = \frac{3}{4}$$

while

$$P(B \mid E) = \frac{P(B \cap E)}{P(E)} = \frac{P(\{HTT\})}{1/2} = \frac{1/8}{1/2} = \frac{1}{4}.$$

Thus the new (conditional) probabilities that Alice and Bob pay for dinner are $3/4$ and $1/4$ respectively.

It may seem like a small matter, but you should be familiar enough with this formula that you can write it down without stopping to think about the names of the events. Thus, for example,

$$P(C \mid D) = \frac{P(C \cap D)}{P(D)}$$

if $P(D) \neq 0$.

Example Two fair ten-sided dice are thrown, independently of each other. Let B be 'doubles' and C be 'at least one is an odd number'. We have seen that $P(B) = 1/10$, $P(C) = 3/4$ and $P(B \cap C) = 1/20$, so

$$P(\text{at least one odd} \mid \text{doubles}) = P(C \mid B) = \frac{P(B \cap C)}{P(B)} = \frac{1/20}{1/10} = \frac{1}{2};$$

thus if you know that you have doubles then you are less likely to have at least one odd number than you would otherwise be—perhaps this is not surprising. On the other hand,

$$P(\text{doubles} \mid \text{at least one odd}) = P(B \mid C) = \frac{P(B \cap C)}{P(C)} = \frac{1/20}{3/4} = \frac{1}{15};$$

this says that if you know that you have at least one odd number then you are less likely to have doubles than you would otherwise be—perhaps this is more surprising.

There is a connection between conditional probability and independence:

Theorem 1 Let A and B be events with $P(A) > 0$ and $P(B) > 0$. Then the following three statements are equivalent.

(i) A and B are independent.

(ii) $P(B \mid A) = P(B)$.

(iii) $P(A \mid B) = P(A)$.

Proof To prove that three conditions are equivalent, we shall prove that (iii) $\Rightarrow$ (i), (i) $\Rightarrow$ (ii), and (ii) $\Rightarrow$ (iii). The we can deduce that any one of the conditions implies both the others by following the 'implies' arrow ($\Rightarrow$) round.

(iii) $\Rightarrow$ (i) We assume that (iii) is true. Then

$$\begin{aligned} P(A \mid B) = P(A) \quad &\Rightarrow \quad \frac{P(A \cap B)}{P(B)} = P(A), && \text{by definition of conditional probability,} \\ &\Rightarrow \quad P(A \cap B) = P(A) \times P(B), && \text{by multiplying up by } P(B), \\ &\Rightarrow \quad A \text{ and } B \text{ are independent}, && \text{by definition of independence.} \end{aligned}$$

(i) $\Rightarrow$ (ii) Now we assume that (i) is true. Then

$$\begin{aligned} A \text{ and } B \text{ are independent} \quad &\Rightarrow \quad P(A \cap B) = P(A) \times P(B), && \text{by definition of independence,} \\ &\Rightarrow \quad \frac{P(A \cap B)}{P(A)} = P(B), && \text{because } P(A) \neq 0, \\ &\Rightarrow \quad P(B \mid A) = P(B), && \text{by definition of conditional probability.} \end{aligned}$$

(ii) ⇒ (iii) Finally we assume that (ii) is true. Then

$$\begin{aligned} P(B \mid A) = P(B) \quad &\Rightarrow \quad \frac{P(A \cap B)}{P(A)} = P(B), \qquad \text{by definition of conditional probability,} \\ &\Rightarrow \quad \frac{P(A \cap B)}{P(B)} = P(A), \qquad \text{because } P(B) \neq 0, \\ &\Rightarrow \quad P(A \mid B) = P(A), \qquad \text{by definition of conditional probability.} \quad \blacksquare \end{aligned}$$

This theorem is most likely what people have in mind when they say 'A and B are independent means that B has no effect on A'.

Genetics

Here is a simplified version of how genes code eye colour, assuming only two colours of eyes.

Each person has two genes for eye colour. Each gene is either B or b. A child receives one gene from each of its parents. The gene it receives from its father is one of its father's two genes, each with probability $1/2$; and similarly for its mother. The genes received from father and mother are independent.

If your genes are BB or Bb or bB, you have brown eyes; if your genes are bb, you have blue eyes.

Example Suppose that John has brown eyes. So do both of John's parents. His sister has blue eyes. What is the probability that John's genes are BB?

Solution John's sister has genes bb, so one b must have come from each parent. Thus each of John's parents is Bb or bB; we may assume Bb. So the possibilities for John are (writing the gene from his father first)

$$\mathrm{BB}, \mathrm{Bb}, \mathrm{bB}, \mathrm{bb}$$

each with probability $1/4$. (For example, John gets his father's B gene with probability $1/2$ and his mother's B gene with probability $1/2$, and these are independent, so the probability that he gets BB is $1/4$. Similarly for the other combinations.)

Let X be the event 'John has BB genes' and Y the event 'John has brown eyes'. Then $X = \{\mathrm{BB}\}$ and $Y = \{\mathrm{BB}, \mathrm{Bb}, \mathrm{bB}\}$. The question asks us to calculate $P(X \mid Y)$. This is given by

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{1/4}{3/4} = 1/3.$$

Iterated conditional probability

The conditional probability of an event C, given that both A and B have occurred, is just $P(C \mid A \cap B)$. Sometimes instead we just write $P(C \mid A,B)$. It is given by

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

so

$$P(A \cap B \cap C) = P(C \mid A,B)P(A \cap B).$$

Now we also have

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

so finally (assuming that $P(A \cap B) \neq 0$), we have

$$P(A \cap B \cap C) = P(C \mid A,B)P(B \mid A)P(A).$$

This generalizes to any number of events:

Theorem 2 Let $E_1, \ldots, E_n$ be events. If none of the probabilities involved is zero then

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = \\ P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \cdots \times P(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

Proof The proof is by *induction*, which you may not have met before.

Step 1 (getting started) When $n = 2$, the statement is

$$P(E_1 \cap E_2) = P(E_1) \times P(E_2 \mid E_1),$$

which is true by definition of $P(E_2 \mid E_1)$.

Step 2 (inductive step) Now assume that the statement is true for $n-1$, so that

$$P(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) = \\ P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \cdots \times P(E_{n-1} \mid E_1 \cap E_2 \cap \cdots \cap E_{n-2}). \quad (*)$$

Put $D = E_1 \cap E_2 \cap \cdots \cap E_{n-1}$. Then

$$
\begin{aligned}
&P(E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap E_n) \\
&= P(D \cap E_n) \\
&= P(D) \times P(E_n \mid D) \qquad \text{by definition of } P(E_n \mid D) \\
&= P(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \times P(E_n \mid D) \\
&= [P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \cdots \times P(E_{n-1} \mid E_1 \cap E_2 \cap \cdots \cap E_{n-2})] \times P(E_n \mid D) \\
&\qquad \text{using } (*) \\
&= P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \cdots \times P(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}),
\end{aligned}
$$

so the statement is true for n. ■

Example There are 5 people in a room. Assuming that all months of the year are equally likely for birthdays, what is the probability that the 5 people were all born in different months of the year?

Solution Let E_2 be the event "second person has a different birth month from the first", E_3 = "third person has a different birth month from the first two", and so on. Then $P(E_2) = 11/12$. Also $P(E_3 \mid E_2) = 10/12$, so

$$P(E_2 \cap E_3) = P(E_2) \times P(E_3 \mid E_2) = \frac{11}{12} \times \frac{10}{12}.$$

Similarly, $P(E_4 \mid E_2 \cap E_3) = 9/12$ and $P(E_5 \mid E_2 \cap E_3 \cap E_4) = 8/12$ so

$$P(E_2 \cap E_3 \cap E_4) = P(E_2 \cap E_3) \times P(E_4 \mid E_2 \cap E_3) = \frac{11}{12} \times \frac{10}{12} \times \frac{9}{12}$$

and

$$P(\text{all different}) = P(E_2 \cap E_3 \cap E_4 \cap E_5) = \frac{11}{12} \times \frac{10}{12} \times \frac{9}{12} \times \frac{8}{12} \approx 0.38.$$

A more complicated version of this argument gives the birthday paradox. The *birthday paradox* is the following statement:

> If there are 23 or more people in a room, then the chances are better than even that two of them have the same birthday.

Can you prove this?

Sampling revisited

In random sampling, we assume that, each time we choose, all objects left are equally likely. In sampling with replacement all objects are available every time, so

$$P(r\text{-th object is } \omega) = \frac{1}{N}$$

no matter what was chosen before. So the r-th object and the s-th object are independent if $r \neq s$. Moreover,

$$P((\omega_1, \omega_2, \ldots, \omega_n)) = \frac{1}{N} \times \frac{1}{N} \times \cdots \times \frac{1}{N},$$

so all outcomes are equally likely.

In sampling without replacement,

$$P(r\text{-th object is } \omega \mid \omega \text{ is already chosen}) = 0,$$

$$P(r\text{-th object is } \omega \mid \omega \text{ is not already chosen}) = \frac{1}{N-r+1}.$$

So

$$\begin{aligned} P((\omega_1, \omega_2, \ldots, \omega_n)) &= P(\text{first is } \omega_1) \times P(\text{second is } \omega_2 \mid \text{first is } \omega_1) \\ &\quad \times P(\text{third is } \omega_3 \mid \text{first is } \omega_1 \text{ and second is } \omega_2) \times \cdots \\ &= \frac{1}{N} \times \frac{1}{N-1} \times \frac{1}{N-2} \times \cdots \times \frac{1}{N-n+1} \end{aligned}$$

for all ordered n-tuples of distinct elements, so again all outcomes are equally likely.

Example There are 3 wallets, 5 notebooks and 4 toy soldiers on the counter in a shop. I choose 3 of these objects, one after the other, without replacement. Let A be 'first is a wallet', B be 'second is a wallet', and C be 'third is a toy soldier'.

Then

$$P(\text{first two are both wallets}) = P(A \cap B) = P(A) \times P(B \mid A) = \frac{3}{12} \times \frac{2}{11} = \frac{1}{22}.$$

$$P(\text{neither of the first two is a wallet}) = P(A' \cap B') = P(A') \times P(B' \mid A') = \frac{9}{12} \times \frac{8}{11} = \frac{6}{11}.$$

$$P(\text{two wallets then a toy soldier}) = P(A \cap B \cap C) = P(A \cap B) \times P(C \mid A \cap B) = \frac{1}{22} \times \frac{4}{10} = \frac{1}{55}.$$