Chapter 4

Conditional Probability & Independence

References: Pruim 2.2.8, Larsen & Marx 2.4

So far, the probabilities we have calculated have all been unconditional probabilities. A sample space was defined and probabilities were calculated relative to this set. In many instances, we may have partial information about the outcome of the experiment which could help narrow down the sample space and refine our probability calculation further.

As an example, suppose we toss a fair coin two times and I tell you that there is at least one head. What is the probability that the other is a tail?

Originally, there are four equally likely outcomes: HH, HT, TH, TT. But only three of them have at least one H, so our sample space is really $\{HH, HT, TH\}$. Of these, two outcomes have a head and also a tail. So the probability is 2/3.

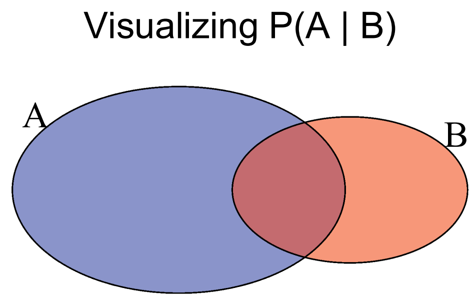
We can also think of this in a different way. In our original sample space of equally likely outcomes:

$$\begin{aligned} P(\text{at least 1 head}) &= \frac{3}{4}, \\ P(\text{ a tail and at least 1 head}) &= \frac{2}{4}, \qquad \text{and} \frac{2/4}{3/4} \qquad\qquad = 2/3; \end{aligned}$$

so 2/3 of the time when there is at least 1 head, there is also a tail.

We will denote this probability as $P(\text{tail}|\text{at least one head})$ and read this as the probability of a tail given there is at least one head.

The venn diagram illustrates conditional probability. We are given that an element in B has occurred, and we wish to calculate the probability that it also belongs to the event A. The conditional probability is the ratio of the probability of $A \cap B$ to the probability of B. The intuition is that B has become our new sample space and all further occurrences are then calibrated with respect to their relation to B.



Definition 4.1. If A and B are events in S and $P(B) > 0$ then the **conditional probability** of A given B, written $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{4.1}$$

. . . .

Example 4.1. The following table contains the prediction record of a TV weather forecaster for 100 days:

Actual \ Forecast	Sunny	Cloudy	Total
Sunny	25	10	35
Cloudy	14	51	65
Total	39	61	100

Suppose we select a day at random[1]. Let A be the event that the forecast is for sunny weather and B the event that the actual weather is sunny.

Determine the value of each of these probabilities.

[1] this means each day is equally likely to be selected

a. $P(A)$

b. $P(A^c)$

c. $P(B|A)$

d. $P(B^c|A)$

e. $P(B|A^c)$

. . . .

Note that

$$P(A^c|B) + P(A|B) = 1,$$

but

$$P(B|A) + P(B|A^c) \neq 1$$

Re-expressing the definition of a conditional probability

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

gives us a useful form for calculating intersection probabilities:

$$P(A \cap B) = P(A|B) \times P(B). \tag{4.2}$$

This is called the **chain rule for probabilities.** There is nothing special about the fact that we chose to re-express $P(A|B)$ to obtain equation (4.2). We could just as easily have used $P(B|A)$ since

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

which would have given the following expression for the intersection probability:

$$P(A \cap B) = P(B|A) \times P(A). \tag{4.3}$$

Both equations (4.2) and (4.3) should yield the same final answer and the choice between which to use will depend on the information given in a problem.

Equating equations (4.2) and (4.3) gives a popular result known as Bayes' theorem.

Theorem 4.1. *For events A and B:*

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

Bayes' theorem is useful for calculating inverse probabilities as illustrated in the example below.

. . . .

Example 4.2. Suppose the probability of snow is 20% and that the probability of an accident on a snowy day is 40%, but only 2.5% on a non-snowy day. We select a day at random and learn there was an accident. What is the probability that there was snow involved?

4.1 Independence of events

In some cases it may happen that the occurrence of a particular event, B, has no effect on the probability of another event A. In other words, the event A is independent of the event of event B. Symbolically we are saying:

$$P(A|B) = P(A). \tag{4.4}$$

Moreover, since

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

it then follows that

$$P(A \cap B) = P(A) \times P(B). \tag{4.5}$$

when A and B are independent.

We take this as our definition of statistical independence. Note that independence could equivalently be defined by either equation (4.4) or (4.5) or even by

$$P(B|A) \quad = \quad P(B).$$

The advantage of equation (4.5) over the other two is that it treats the events symmetrically and will be easier to generalize to more than two events.

Definition 4.2. If

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

then we say that A and B are **independent events**.

. . . .

Example 4.3. A fair six sided dice is rolled twice. Suppose A is the event that the first throw yields a 2 or a 5 and B is the event that the sum of the two throws is 7.

a. Are A and B disjoint?

b. Are A and B independent?

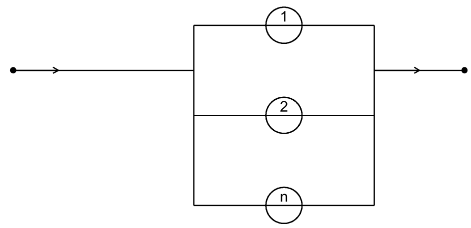
If two events are disjoint, then they are definitely not independent!

Independence of a collection of events requires an extremely strong condition as defined below.

Definition 4.3. A collection of events $A_1, A_2, \dots, A_n$ are **mutually independent** if for <u>any</u> sub-collection $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ we have:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \quad = \quad P(A_{i_1}) \times P(A_{i_2}) \times \cdots \times P(A_{i_k}).$$

Example 4.4. A system composed of n separate components is said to be a parallel system if it functions when at least one of the components function. For such a system, if component i, which is independent of the other components, functions with probability $p_i, i = 1, \dots, n$, what is the probability that the system fails to function?



. . . .

4.2 Practice Problems

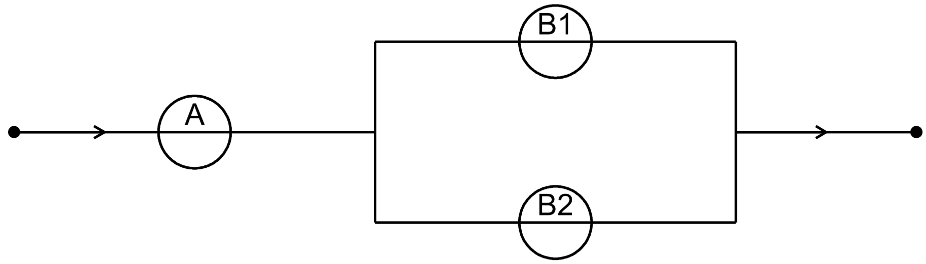
1. Give the values for the following conditional probabilities.
   a. $P(B|B)$
   b. $P(B|B^c)$
   c. $P(B|S)$
   d. $P(B|\emptyset)$
   e. $P(B|A)$ when A is a subset of B.

2. If $P(A) = 1/4$, $P(B|A) = 1/2$ and $P(A|B) = 1/3$. Find $P(A \cup B)$.

3. A high school senior is looking into college options. He really wants to live on campus, and is considering two choices: Evergreen State College, which admits 96% of the applicants and has 20% of its students living on campus, or Western Washington University, which admits 78% of the applicants and has 29% of its students living on campus. Everything else being equal, which choice is better? (Hint: define two events - A: being admitted and B: getting campus housing. You are given $P(A)$ and $P(B|A)$ at each college. Use them to find $P(A \cap B)$ and then compare. The one with the bigger value for $P(A \cap B)$ is better)

4. An electrical circuit is composed of relays of types A, and B in the configuration pictured below. Suppose that individual relays of types A and B function properly with probabilities .9, .8, respectively, and that the

relays operate independently.



What is the probability that the current will flow (from one black dot to the other) when the relays are activated. Hint: In order for the current to flow, the single type A relay must function properly AND at least one of the type B relays must function properly.