Lecture 1: Probability on Events

Contents

1 Sample space and events

Probability is typically defined in terms of some **experiment**. The **sample space**, denoted by Ω, is the set of all possible outcomes of the experiment.

Definition 1 *An* **event**, *E, is any subset of the sample space, Ω.*

Example 2 *Consider the experiment where 2 dice are rolled. An outcome (a.k.a. sample point) is denoted by (i, j), where i is the first roll and j is the second roll. There are 36 sample points. Let*

$$E = \{ (1,3) \textit{ or } (2,2) \textit{ or } (3,1) \}$$

denote the event that the sum of the dice rolls is 4.

In general, the sample space may be *discrete*, meaning that the number of outcomes are finite or at least countable, or *continuous*, meaning that the number of outcomes are uncountable.

Example 3 *Suppose you are throwing a dart, which is equally likely to land anywhere in the interval* $[0, 1]$. *The sample space here is continuous. An example of an event here is:*

$$E = \text{ Event that dart lands in the interval } [0, 0.3]$$

One can talk of unions and intersections of events, since they are also sets: e.g., we can talk about $E \cup F$, $E \cap F$, and E^c, where E and F are events, and E^c is the complement of E, namely the set of sample points in Ω but not in E.

Question: For an experiments involving the rolling of 2 dice, consider events E_1 and E_2 defined on Ω in Figure 1. Do you think that E_1 and E_2 are independent?

$$\begin{array}{cccccc} & E_1 & & & E_2 & \\ \end{array}$$

$$S = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right.$$

Figure 1: *Illustration of two events in sample space* S.

Answer: No, they're not independent. We'll get to this later when we define independence.

We say that E_1 and E_2 are mutually exclusive.

Definition 4 *If* $E_1 \cap E_2 = \emptyset$ *then* E_1 *and* E_2 *are* **mutually exclusive**.

Definition 5 *If* $E_1, E_2, \ldots, E_n$ *are events such that* $E_i \cap E_j = \emptyset, \forall i, j$, *and such that* $\bigcup_{i=1}^{n} E_i = F$, *then we say that events* $E_1, E_2, \ldots, E_n$ **partition** *set* F.

2 Probability defined on events

Probability is defined on events.

$$\mathbf{P}\{E\} = \text{probability of event E occurring.}$$

We can think of each sample point as having some probability of occurring, and the probability that event E occurs is the sum of the probabilities of the sample points in E.

Example 6 *Consider again the experiment where 2 dice are rolled. Each outcome* (i, j) *is equally likely, and has probability* $\frac{1}{36}$*. Let*

$$E = \{\ (1,3)\ \textit{or}\ (2,2)\ \textit{or}\ (3,1)\ \}$$

denote the event that the sum of the dice rolls is 4. We say that $\mathbf{P}\{E\} = \frac{3}{36}$.

Example 7 *Suppose you are throwing a dart, which is equally likely to land anywhere in the interval* $[0,1]$*. The sample space here is continuous.*

Question: *What is the probability that the dart lands at exactly* 0.3*?*

Answer: *We define this to be 0. You will soon see why.*

Question: *What is the probability that the dart lands in* $[0, 0.3]$*?*

Answer: *This is 30%.*

Definition 8 *The probability of the union of two events is defined as follows:*

$$\mathbf{P}\{E \cup F\} = \mathbf{P}\{E\} + \mathbf{P}\{F\} - \mathbf{P}\{E \cap F\}$$

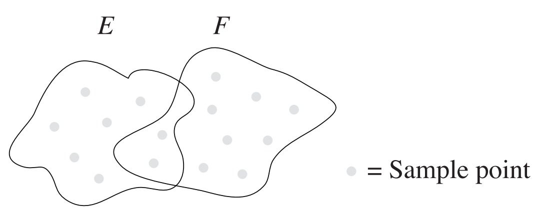

Figure 2: *Venn diagram.*

This should make sense if we think of events as sets. Observe that the subtraction of the $\mathbf{P}\{E \cap F\}$ term is necessary so that those sample points are not counted twice.

Theorem 9 $\mathbf{P}\{E \cup F\} \leq \mathbf{P}\{E\} + \mathbf{P}\{F\}$.

Proof: This follows immediately from Definition 8. ■

Question: When is Theorem 9 an equality?

Answer: When E and F are mutually exclusive.

Question: Above we wrote $\mathbf{P}\{E \cup F\}$ in terms of E, F, and their intersection. If is possible to do something similar for $\mathbf{P}\{E \cup F \cup G\}$?

Answer:

$$\mathbf{P}\{E \cup F \cup G\} = \mathbf{P}\{E\} + \mathbf{P}\{E^c \cap F\} + \mathbf{P}\{E^c \cap F^c \cap G\}$$

This is easy to see from the Venn diagram.

3 The 3 Axioms of Probability

Let Ω represent the sample space, and let E and F represents any two *mutually exclusive* events. Then:

1. $\mathbf{P}\{E\} \geq 0, \quad \forall E$
2. $\mathbf{P}\{E \cup F\} = \mathbf{P}\{E\} + \mathbf{P}\{F\}$
3. $\mathbf{P}\{\Omega\} = 1$

Question: Can you now understand why we had to define:

$$\mathbf{P}\{\text{ Event that dart lands at exactly } 0.3\} = 0$$

Answer: Suppose this were $\epsilon > 0$. Then we could take a countable number of points between 0 and 1 (e.g. the rationals) and their probability would be > 1, which is not allowed because the whole sample space only has probability 1.

4 Conditional probabilities and events

Definition 10 *The* **conditional probability of event** E **given event** F *is written as* $\mathbf{P}\{E \mid F\}$ *and is given by the following, where we assume* $\mathbf{P}\{F\} > 0$*:*

$$\mathbf{P}\{E \mid F\} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}} \tag{1}$$

This can be thought of as the probability that event E occurs, given that we have narrowed our sample space to points in F. To see this, consider Figure 3:

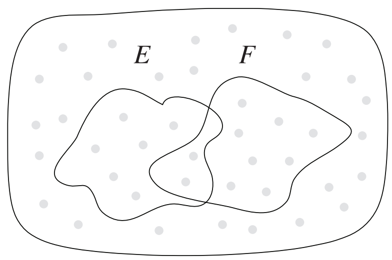

Figure 3: *Sample space with 42 sample points.*

$$\begin{aligned}
\mathbf{P}\{E\} &= \frac{8}{42} \\
\mathbf{P}\{F\} &= \frac{10}{42} \\
\mathbf{P}\{E \mid F\} &= \frac{2}{10} = \frac{\frac{2}{42}}{\frac{10}{42}} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}}
\end{aligned}$$

Thus $\mathbf{P}\{E \mid F\}$ represents the probability of being in subset E, given that we have narrowed our space of possible sample points to only those 10 points in F.

Example 11 *Table 1 shows my sandwich choices each day. Let's define the "first half of the week" to be Monday through Wednesday (inclusive). Let's define the "second half of the week" to be Thursday through Sunday (inclusive).*

Mon	Tue	Wed	Thu	Fri	Sat	Sun
Jelly	Cheese	Turkey	Cheese	Turkey	Cheese	None

Table 1: *My sandwich choices*

Question: *What is* $\mathbf{P}\{Cheese \mid Second\ half\ of\ week\}$*?*

Answer: *One way to do this is to imagine that we're restriced to the 2nd half of the week. In that case, the fraction of days that I eat a cheese sandwich is* $\frac{2}{4}$*. A different way to do this is to use conditioning as follows:*

$$\begin{aligned}\mathbf{P}\{Cheese\ \&\ Second\ half\} &= \frac{2}{7}\\ \mathbf{P}\{Second\ half\} &= \frac{4}{7}\\ \mathbf{P}\{Cheese \mid Second\ half\ of\ week\} &= \frac{\frac{2}{7}}{\frac{4}{7}} = \frac{2}{4}\end{aligned}$$

Example 12 *A couple has two children.*

Question: *What is* $\mathbf{P}\{both\ are\ boys \mid at\ least\ one\ is\ a\ boy\}$?

Answer:

$$\begin{aligned}&\mathbf{P}\{both\ are\ boys \mid at\ least\ one\ is\ a\ boy\}\\ &= \frac{\mathbf{P}\{both\ are\ boys\ and\ at\ least\ one\ is\ a\ boy\}}{\mathbf{P}\{at\ least\ one\ is\ a\ boy\}}\\ &= \frac{\mathbf{P}\{both\ are\ boys\}}{\mathbf{P}\{at\ least\ one\ is\ a\ boy\}}\\ &= \frac{\frac{1}{4}}{\frac{3}{4}}\\ &= \frac{1}{3}\end{aligned}$$

Question: *How might the question read if you wanted the answer to be* $\frac{1}{2}$*?*

Answer: *The question would ask what is* $\mathbf{P}\{\text{both are boys} \mid \text{first is a boy}\}$*?*

Here's an example that uses conditioning in the other direction, to get $\mathbf{P}\{E \cap F\}$.

Example 13 *A bomb detector alarm lights up with probability* 0.99 *if a bomb is present. If no bomb is present, the bomb alarm still (incorrectly) lights up with probability* 0.05.*Suppose that a bomb is present with probability* 0.1.

Question: *What is the probability that there is no bomb and the alarm lights up?*

Answer:

$$\begin{aligned}\mathbf{P}\{\textit{No bomb} \ \& \ \textit{Alarm}\} &= \mathbf{P}\{\textit{Alarm} \mid \textit{No bomb}\} \cdot \mathbf{P}\{\textit{No bomb}\} \\ &= 0.05 \cdot 0.9 \\ &= 0.045\end{aligned}$$

Equation (1) can also be used to prove the following more general theorem involving conditioning:

Theorem 14 *Suppose there are n events: $E_1, E_2, \ldots, E_n$, each with positive probability. Then*

$$\mathbf{P}\{E_1 \cap E_2 \cap \cdots \cap E_n\} = \mathbf{P}\{E_1\} \cdot \mathbf{P}\{E_2 \mid E_1\} \cdot \mathbf{P}\{E_3 \mid E_1 \cap E_2\} \cdots \mathbf{P}\{E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}\}$$

5 Independent events

Definition 15 *Events E and F are* **independent** *if*

$$\mathbf{P}\{E \cap F\} = \mathbf{P}\{E\}\,\mathbf{P}\{F\}$$

Question: If E and F are independent, what is $\mathbf{P}\{E \mid F\}$?

Answer:

$$\mathbf{P}\{E \mid F\} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}} \stackrel{indpt}{=} \frac{\mathbf{P}\{E\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{F\}} = \mathbf{P}\{E\}$$

That is, $\mathbf{P}\{E\}$ is not affected by whether F is true or not.

Question: Can two mutually exclusive (non-null) events ever be independent?

Answer: No. In this case, $\mathbf{P}\{E \mid F\} = 0 \neq \mathbf{P}\{E\}$.

Question: Suppose one is rolling 2 dice. Which of these pairs of events are independent?

1. E_1 = "First roll is 6" and E_2 = "Second roll is 6"
2. E_1 = "Total of the two rolls is 7" and E_2 = "First roll is 4"

Answer: They're both independent!

Question: Suppose we had defined: E_1 = "Total of the two rolls is 8" and E_2 = "First roll is 4" Are E_1 and E_2 independent now?

Answer: No.

6 Law of total probability

We now derive the Law of Total Probability.

Observe:

$$E = (E \cap F) \cup \left(E \cap F^C\right)$$

That is, the set E can be viewed as the union of the set $E \cap F$ and the set $E \cap F^C$, because any point in E is either in F or *not* in F.

Now observe: $E \cap F$ and $E \cap F^C$ are mutually exclusive. Thus:

$$\begin{aligned}
\mathbf{P}\{E\} &= \mathbf{P}\{E \cap F\} + \mathbf{P}\left\{E \cap F^C\right\} \\
&= \mathbf{P}\{E \mid F\}\,\mathbf{P}\{F\} + \mathbf{P}\left\{E \mid F^C\right\}\mathbf{P}\left\{F^C\right\}
\end{aligned}$$

where $\mathbf{P}\left\{F^C\right\} = 1 - \mathbf{P}\{F\}$

The theorem below is a generalization:

Theorem 16 (Law of total probability) *Let $F_1, F_2, \ldots, F_n$ partition the state space Ω. Then,*

$$\begin{aligned}
\mathbf{P}\{E\} &= \sum_{i=1}^{n} \mathbf{P}\{(E \cap F_i)\} \\
&= \sum_{i=1}^{n} \mathbf{P}\{E \mid F_i\} \cdot \mathbf{P}\{F_i\}
\end{aligned}$$

Proof:

$$E = \bigcup_{i=1}^{n} (E \cap F_i)$$

Now, since the events $E \cap F_i$, $i = 1 \ldots n$, are mutually exclusive, we have that:

$$\mathbf{P}\{E\} = \sum_{i=1}^{n} \mathbf{P}\{(E \cap F_i)\} = \sum_{i=1}^{n} \mathbf{P}\{E|F_i\} \cdot \mathbf{P}\{F_i\}$$

■

Question: Suppose we are interested in the probability that a certain type of transaction fails. We know that if there is a caching failure, then the transaction will fail with probability 5/6. We also know that if there is a network failure then the transaction will fail with probability 1/4. Suppose that the network fails with probability 1/100, and the cache fails with probability 1/100. Is this enough to tell us the probability that the transaction will fail?

Answer: It is tempting to write:

(WRONG)

$$\begin{aligned}
\mathbf{P}\{\text{transaction fails}\} &= \mathbf{P}\{\text{transaction fails} \mid \text{caching failure}\} \cdot \frac{1}{100} \\
&\quad + \mathbf{P}\{\text{transaction fails} \mid \text{network failure}\} \cdot \frac{1}{100} \\
&= \frac{5}{6} \cdot \frac{1}{100} + \frac{1}{4} \cdot \frac{1}{100}
\end{aligned}$$

Question: What is wrong with the above?

Answer: The two events we conditioned on, a network failure and a caching failure, do not necessarily partition the space. First, they may not be mutually exclusive, since a network failure may cause a caching failure, and thus the probability that they both occur is non-zero. Second, there may be many *other* reasons why the transaction failed. We have not expressed the probability that the transaction fails for some reason unrelated to either the cache or the network. The sum of the probabilities of the events that we condition on needs to total 1.

One needs to be very careful that the events are *both* (i) mutually exclusive and (ii) sum to the whole sample space.

Here is a nice example of using the Law of Total Probability repeatedly:

Example 17 *A computer switches between 2 modes: working and broken. Each day the computer has some probability of staying in the same mode and some probability of switching, as shown below.*

Question: *Suppose that the computer starts out in the working state. What is the probability that after 3 days the computer is in the working state?*

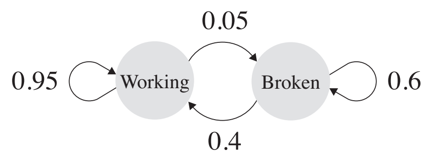

Figure 4: *Two modes of machine.*

Answer: *We will write W_i (respectively, B_i) to denote the event that the computer is working (respectively, broken) after i days.*

We start by expressing W_3 in terms of W_2 and B_2:

$$\begin{aligned}\mathbf{P}\{W_3\} &= \mathbf{P}\{W_3 \mid W_2\} \cdot \mathbf{P}\{W_2\} + \mathbf{P}\{W_3 \mid B_2\} \cdot \mathbf{P}\{B_2\} \\ &= 0.95 \cdot \mathbf{P}\{W_2\} + 0.4 \cdot \mathbf{P}\{B_2\}\end{aligned}$$

Now we express W_2 and B_2 in terms of W_1 and B_1:

$$\begin{aligned}\mathbf{P}\{W_2\} &= \mathbf{P}\{W_2 \mid W_1\} \cdot \mathbf{P}\{W_1\} + \mathbf{P}\{W_2 \mid B_1\} \cdot \mathbf{P}\{B_1\} \\ &= 0.95 \cdot \mathbf{P}\{W_1\} + 0.4 \cdot \mathbf{P}\{B_1\}\end{aligned}$$

$$\begin{aligned}\mathbf{P}\{B_2\} &= \mathbf{P}\{B_2 \mid W_1\} \cdot \mathbf{P}\{W_1\} + \mathbf{P}\{B_2 \mid B_1\} \cdot \mathbf{P}\{B_1\} \\ &= 0.05 \cdot \mathbf{P}\{W_1\} + 0.6 \cdot \mathbf{P}\{B_1\}\end{aligned}$$

Finally we observe that, given that we're starting out in the working state,

$$\begin{aligned}\mathbf{P}\{W_1\} &= 0.95 \\ \mathbf{P}\{B_1\} &= 0.05\end{aligned}$$

The above equations easily yield $\mathbf{P}\{W_3\}$. Likewise, it is easy to write a computer program to derive $\mathbf{P}\{W_n\}$, for any n.

7 Sneak Preview of next time ...

Sometimes, one needs to know $\mathbf{P}\{F|E\}$, but all one knows is the reverse direction: $\mathbf{P}\{E|F\}$. Is it possible to get $\mathbf{P}\{F|E\}$ if all one knows is: $\mathbf{P}\{E|F\}$, $\mathbf{P}\{E\}$, and $\mathbf{P}\{F\}$? It turns out that it is! See if you can figure it out.