Lecture 2

Consequences of the probability rules

Example 1 (Complement rule)**.** Recall that A^c, the complement of A, is the collection of all points in Ω that are not in A. Thus, A and A^c are disjoint. Because $\Omega = A \cup A^c$ is a disjoint union, Rules 2 and 3 together imply then that

$$\begin{aligned} 1 &= \mathrm{P}(\Omega) \\ &= \mathrm{P}(A \cup A^c) \\ &= \mathrm{P}(A) + \mathrm{P}(A^c). \end{aligned}$$

Thus, we obtain the physically–appealing statement that

$$\mathrm{P}(A) = 1 - \mathrm{P}(A^c).$$

For instance, this yields $\mathrm{P}(\varnothing) = 1 - \mathrm{P}(\Omega) = 0$. "Chances are zero that nothing happens."

Example 2. If $A \subseteq B$, then we can write B as a disjoint union: $B = A \cup (B \cap A^c)$. Therefore, $\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(B \cap A^c)$; here $E \cap F$ denotes the intersection of E and F [some people, including the author of your textbook, prefer EF to $E \cap F$; the meaning is of course the same]. The latter probability is ≥ 0 by Rule 1. Therefore, we reach another physically-appealing property:

$$\text{If } A \cup B, \text{ then } \mathrm{P}(A) \leq \mathrm{P}(B).$$

Example 3 (Difference rule)**.** The event that "B happens but not A" is in math notation $B \cap A^c$. Note that $B = (B \cap A) \cup (B \cap A^c)$ is the union of two disjoint sets. Therefore, $\mathrm{P}(B) = \mathrm{P}(B \cap A) + \mathrm{P}(B \cap A^c)$, which can be rewritten as

$$\mathrm{P}(B \cap A^c) = \mathrm{P}(B) - \mathrm{P}(B \cap A).$$

If, in addition, $A \subseteq B$ [in words, whenever A happens then so does B] then $B \cap A = A$, and we have $\mathrm{P}(B \cap A^c) = \mathrm{P}(B) - \mathrm{P}(A)$. This is the "difference rule" of your textbook.

Example 4. Suppose $\Omega = \{\omega_1, \ldots, \omega_N\}$ has N distinct elements ("N distinct outcomes of the experiment"). One way of assigning probabilities to every subset of Ω is to just let

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{N},$$

where $|E|$ denotes the number of elements of E. Let us check that this probability assignment satisfies Rules 1–4. Rules 1 and 2 are easy to verify, and Rule 4 holds vacuously because Ω does not have infinitely-many disjoint subsets. It remains to verify Rule 3. If A and B are disjoint subsets of Ω, then $|A \cup B| = |A| + |B|$. Rule 3 follows from this. In this example, each outcome ω_i has probability $1/N$. Thus, these are "equally likely outcomes."

Example 5. Let

$$\Omega = \Big\{ (H_1\,, H_2)\,, (H_1\,, T_2)\,, (T_1\,, H_2)\,, (T_1\,, T_2) \Big\}.$$

There are four possible outcomes. Suppose that they are equally likely. Then, by Rule 3,

$$\begin{aligned}
\mathrm{P}(\{H_1\}) &= \mathrm{P}\Big(\{H_1\,, H_2\} \cup \{\, H_1\,, T_2\} \Big) \\
&= \mathrm{P}(\{H_1\,, H_2\}) + \mathrm{P}(\{H_1\,, T_2\}) \\
&= \frac{1}{4} + \frac{1}{4} \\
&= \frac{1}{2}.
\end{aligned}$$

In fact, in this model for equally-likely outcomes, $\mathrm{P}(\{H_1\}) = \mathrm{P}(\{H_2\}) = \mathrm{P}(\{T_1\}) = \mathrm{P}(\{T_2\}) = 1/2$. Thus, we are modeling two fair tosses of two fair coins.

Example 6. Let us continue with the sample space of the previous example, but assign probabilities differently. Here, we define $\mathrm{P}(\{H_1\,, H_2\}) = \mathrm{P}(\{T_1\,, T_2\}) = 1/2$ and $\mathrm{P}(\{H_1\,, T_2\}) = \mathrm{P}(\{T_1\,, H_2\}) = 1/2$. We compute, as we did before, to find that $\mathrm{P}(\{H_1\}) = \mathrm{P}(\{H_2\}) = \mathrm{P}(\{H_3\}) = \mathrm{P}(\{H_4\}) = 1/2$. But now the coins are not tossed fairly. In fact, the results of the two coin tosses are the same in this model.

The following generalizes Rule 3, because $\mathrm{P}(A \cap B) = 0$ when A and B are disjoint.

Lemma 1 (Inclusion-exclusion rule)**.** *If A and B are events (not necessarily disjoint), then*

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B).$$

Proof. We can write $A \cup B$ as a disjoint union of three events:

$$A \cup B = (A \cap B^c) \cup (A^c \cap B) \cup (A \cap B).$$

By Rule 3,

$$\mathrm{P}(A \cup B) = \mathrm{P}(A \cap B^c) + \mathrm{P}(A^c \cap B) + \mathrm{P}(A \cap B). \tag{1}$$

Similarly, write $A = (A \cap B^c) \cup (A \cap B)$, as a disjoint union, to find that

$$\mathrm{P}(A) = \mathrm{P}(A \cap B^c) + \mathrm{P}(A \cap B). \tag{2}$$

There is a third identity that is proved the same way. Namely,

$$\mathrm{P}(B) = \mathrm{P}(A^c \cap B) + \mathrm{P}(A \cap B). \tag{3}$$

Add (2) and (3) and solve to find that

$$\mathrm{P}(A \cap B^c) + \mathrm{P}(A^c \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - 2\mathrm{P}(A \cap B).$$

Plug this in to the right-hand side of (1) to finish the proof. □

Examples

Example 7 (Rich and famous, p. 23 of Pitman)**.** In a certain population, 10% of the people are rich $[R]$, 5% are famous $[F]$, and 3% are rich and famous $[R \cap F]$. For a person picked at random from this population:

(1) What is the chance that the person is not rich $[R^c]$?

$$\mathrm{P}(R^c) = 1 = \mathrm{P}(R) = 1 - 0.1 = 0.9.$$

(2) What is the chance that the person is rich, but not famous $[R \cap F^c]$?

$$\mathrm{P}(R \cap F^c) = \mathrm{P}(R) - \mathrm{P}(R \cap F) = 0.1 - 0.03 = 0.07.$$

(3) What is the chance that the person is either rich or famous [or both]?

$$\mathrm{P}(R \cup F) = \mathrm{P}(R) + \mathrm{P}(F) - \mathrm{P}(R \cap F) = 0.1 + 0.05 - 0.03 = 0.12.$$

Example 8. A fair die is cast. What is the chance that the number of dots is 3 or greater? Let N denote the number of dots rolled. N is what is called a "random variable." Let N_j denote the event that $N = j$. Note that $N \geq 3$ corresponds to the event $N_3 \cup N_4 \cup N_5 \cup N_6$. Because the N_j's are disjoint [can't happen at the same time],

$$\mathrm{P}\{N \geq 3\} = \mathrm{P}(N_3) + \mathrm{P}(N_4) + \mathrm{P}(N_5) + \mathrm{P}(N_6) = \frac{4}{6} = \frac{2}{3}.$$