4. Conditional probability

Suppose, for a random experiment, you are given some information about its outcome. For example, in a single roll of a fair 6-sided die, suppose I tell you that an even number was rolled. How do you update your probabilities in the face of such information? Adopting the frequentist philosophy, we should update probabilities so that if we were to repeat the experiment indefinitely and only pay attention to outcomes satisfying the condition “an even number occurs,” then the proportion of times an event $E'$ occurs among those trials converges to our new probability assignment of $E'$. In this section, we make this intuition formal.

Let $(\Omega, \mathcal{E}, P)$ be a probability model and define, for any $A, B \in \mathcal{E}$, the conditional probability of $B$ given $A$ by

$$P[B \mid A] := \begin{cases} \frac{P[A \cap B]}{P[A]}, & P[A] > 0 \\ \text{arbitrary}, & \text{otherwise.} \end{cases}$$

We can interpret (4) in at least two ways:

- **Frequentist interpretation**: observe outcomes repeatedly from $(\Omega, \mathcal{E}, P)$ and let $n_A$, $n_B$, and $n_{A \cap B}$ be the number of times $A$, $B$ and $A \cap B$ occur, respectively, in $n$ trials. Then, from before, we have $n_A/n \to P[A]$. Furthermore, $n_{A \cap B}/n_A$ is the fraction of times that $B$ occurred whenever $A$ occurred. Therefore,

$$\frac{n_{A \cap B}}{n_A} \approx \frac{P[A \cap B]}{P[A]} = P[B \mid A].$$

- **Bayesian interpretation**: Update your “belief” that $B$ occurred/will occur based on the information in event $A$.

**Example 4.1.** In the classic game show Let’s Make a Deal, the host Monty Hall would present the contestant with three doors, labeled 1, 2, and 3. Behind one of the doors was a grand prize, e.g., a new car; behind the other two doors was a goat.\(^1\) The game begins with the contestant choosing a door. Monty Hall then reveals a goat behind one of the unchosen doors so that now there are two unknown doors (the chosen door and the remaining unchosen door). Hall then gives the contestant the option of switching his choice of doors or keeping his original choice. After this decision is made, the contents behind each door are revealed and the contestant wins whatever is behind his chosen door (either a car or a goat). Should the contestant switch doors? Does it matter?

Initially, the contestant has three doors to choose from. Assuming the contestant has no extra information about which door is more likely to contain the car, the probability that the initially chosen door has a car behind it is $1/3$, which coincides with the probability of winning the car given that he stays with his initial choice. On the other hand, of the two initially unchosen doors, there is a probability $2/3$ that one has the car behind it and probability $1$ that there is a goat. We can formalize this by considering the possible combinations behind the unchosen doors $GG, CG, GC$, each occurring with initial probability $1/3$. Therefore,

$$P[W \mid \text{switch}] = P[W \mid \text{switch}, GG]P[GG] + P[W \mid \text{switch}, GG^c]P[GG^c]$$

$$= 0 \times \frac{1}{3} + 1 \times \frac{2}{3}$$

$$= \frac{2}{3}.\footnote{A modern version of this show airs now and is hosted by Wayne Brady. The games on that show are variations on this classic theme.}$$
Problem 4.2. Many people, even after seeing the above calculation, do not fully believe the conclusion that switching is the optimal strategy. Think about a clear, concise, intuitive way to convince someone of this without using language or tools of probability.

Example 4.3 (Equally likely outcomes). Let $\Omega$ be finite, $\mathcal{E} = 2^\Omega$, and $P$ be the uniform distribution on $\mathcal{E}$, i.e., $P[E] := #E/#\Omega$ for $E \in \mathcal{E}$. Then, assuming $A \neq \emptyset$,

$$P[B | A] = \frac{P[A \cap B]}{P[A]} = \frac{(A \cap B)/#\Omega}{#A/#\Omega} = \frac{(A \cap B)}{#A}.  \tag{1}$$

In particular,

- $P[A | A] = 1$ and
- $P[\omega | A] = 1/#\Omega$ for all $\omega \in \Omega$.

Therefore, for computing $P[\cdot | A]$, we can treat $A$ as the sample space and update the set of events to $\mathcal{E}_A := 2^A$.

Exercise 4.4. Suppose $P[A] > 0$, then $P[\cdot | A]$ induces a probability measure on subsets of $A$.

Example 4.5 illustrates a very important property of the discrete uniform distribution. Given a uniform probability model on a finite sample space $\Omega$ and a non-empty subset $A \subseteq \Omega$, the conditional probability measure $P[\cdot | A]$ induced on $A$ is the discrete uniform probability measure on sample space $A$. This property, sometimes called consistency under subsampling, is not satisfied by all probability measures. It is a very useful property to keep in mind.

4.1. Law of cases and subcases. Suppose an event $C \in \mathcal{E}$ can be partitioned into subcases $C_1, C_2, \ldots$ such that $C_1, C_2, \ldots$ are mutually exclusive and $\bigcup_i C_i = C$. Note that, in this case, $\sum_i P[C_i | C] = 1$. For an event $A \in \mathcal{E}$, how can we compute $P[A | C]$ from $P[C_i | C]$, $i \geq 1$?

Note that the events $A \cap C_1, A \cap C_2, \ldots$ are mutually exclusive (because $C_1, C_2, \ldots$ are) and $\bigcup_{i=1}^\infty (A \cap C_i) = A \cap C$ (because $\bigcup_{i=1}^\infty C_i = C$). Thus,

$$P[A | C] = \frac{P[A \cap C]}{P[C]} = \sum_{i=1}^\infty \frac{P[A \cap C_i]}{P[C]} = \sum_{i=1}^\infty \frac{P[C_i | C]P[A | C_i]}{P[C]} = \sum_{i=1}^\infty P[C_i | C]P[A | C_i].  \tag{2}$$

That is, the conditional probability of $A$ in case $C$ is the weighted average of conditional probability of $A$ in each subcase. When $C = \Omega$, this is called the law of cases, or law of total probability:

$$P[A] = \sum_i P[A | \Omega_i]P[\Omega_i],  \tag{3}$$

where $\Omega_1, \Omega_2, \ldots$ is a partition of $\Omega$.

Example 4.5 (Craps). Craps is a game played with 2 6-sided dice. The outcome of a given roll of the dice is the sum of the number of dots. The rules are as follows:

- On the first roll:
– you lose ("crap out") if roll a total of 2, 3, or 12;
– you win if roll a total of 7 or 11;
– otherwise, the roll total (4, 5, 6, 8, 9, or 10) becomes the point and
• you continue to roll until
  – you roll the point before a 7, you win;
  – otherwise, you lose.

What is the probability that you win at craps?

The strategy we use for computing this quantity is to condition on the first roll. This is a standard technique for analyzing processes that involve independent trials, but the rules of the game allow for an indefinite number of repetitions. Let \( W \) be the event you win and let \( T_i \) be the event that you get a total of \( i \) on the first roll, \( i = 2, 3, \ldots, 12 \). Note that the events \( T_i \) are mutually exclusive (\( T_i \cap T_j = \emptyset \) for \( i \neq j \)) and exhaustive (\( \bigcup_{i=2}^{12} T_i = \Omega \)). Hence, \( T_2, \ldots, T_{12} \) is a partition of \( \Omega \). Now, by the law of cases,

\[
P[W] = \sum_{i=2}^{12} P[T_i]P[W \mid T_i].
\]

Recall the \( P[T_i], i = 2, \ldots, 12 \), from Example 2.2, for which we must fill in the row with \( P[W \mid T_i] \):

<table>
<thead>
<tr>
<th>( i )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P[T_i] )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{4}{36} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{6}{36} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{4}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{1}{36} )</td>
</tr>
<tr>
<td>( P[W \mid T_i] )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{3}{10} )</td>
<td>( \frac{1}{11} )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The table is completing by conditioning on the first step. For example,

\[
P[W \mid T_4] = \frac{3}{36} + (1 - \frac{3}{36} - \frac{9}{36})P[W \mid T_4] = \frac{3}{9}.
\]

The complete table is given above.

Alternatively, we can notice that once the point is established at \( i = 4, 5, 6, 8, 9, \) or 10, we reduce the set of important outcomes to \( \{i\} \) is rolled} and \( \{7\} \) is rolled}. Intuitively, we can consider the new sample space \( \Omega_i = \{i, 7\} \) on which we assign probability

\[
P([i] \mid \{i \text{ or } 7\}) = \frac{P[T_i]}{P[T_i] + P[T_7]} = P[W \mid T_i], \quad i = 4, 5, 6, 8, 9, 10.
\]

It is important to realize that this is not a complete argument, because we have not accounted for the fact that we might never roll \( i \) or 7. To complete the argument, we would have to prove that the number of rolls before \( i \) or 7 shows up is finite with probability one. This amounts to summing a geometric series, and is left as an exercise.

4.2. Independent and exchangeable events. The rolls in the above model for craps are assumed to be independent.

Definition 4.6 (Independent Events). For a probability model \((\Omega, \mathcal{E}, P)\), events \( A, B \in \mathcal{E} \) are independent if

\[(6) \quad P[A \cap B] = P[A]P[B].\]

Remark 4.7. Drawing with replacement from the same set (usually) assumes independence, while drawing without replacement assumes exchangeability.
Example 4.8 (Exchangeability). Suppose two balls are drawn without replacement from a box containing balls labeled 1, . . . , 6. We assume that, on each draw, all possible outcomes are equally likely. Then we have
\[ P[\text{first ball a 2}] = \frac{1}{6} = P[\text{second ball a 2}] \]
but
\[ P[\text{first ball a 2} \mid \text{second ball a 1}] = \frac{1}{5}. \]

In general, when drawing without replacement \( n \) times from \( \{1, \ldots, N\} \), for any distinct indices \( n_1, \ldots, n_k \), the chances associated with draws \( n_1, n_2, \ldots, n_k \) are the same as that chances associated with draws 1, . . . , \( k \). The draws are said to be exchangeable.

Example 4.9 (Card shuffling). What is the probability that all 4 aces appear next to each other in a well-shuffled deck of 52 cards? By exchangeability,
\[ P[4 \text{ aces next to each other}] = 49 \times P[4 \text{ aces on the top}]. \]

Let \( A, B \in \mathcal{E} \) be events. By definition of \( P[B \mid A] \) in (4), we have
\[ P[A \cap B] = P[B \mid A]P[A]. \]

For three events \( A, B, C \in \mathcal{E} \), we have
\[ P[A \cap B \cap C] = P[C \mid A \cap B]P[A \cap B] = P[C \mid A \cap B]P[B \mid A]P[A]. \]

Proposition 4.10 (Multiplication rule for conditional probabilities). For a probability model \((\Omega, \mathcal{E}, P)\), let \( A_1, A_2, \ldots, A_n \in \mathcal{E} \), then
\[ P\left[ \bigcap_{j=1}^{n} A_j \right] = \prod_{j=1}^{n} P\left[ A_j \mid \bigcap_{i=1}^{j-1} A_i \right], \]
where we define \( P[A_1 \mid \emptyset] := P[A_1] \).

Proof. By induction, \( P\left[ \bigcap_{j=1}^{n+1} A_j \right] = P[A_{n+1} \cap A_{[n]} \mid \emptyset] = P[A_{n+1} \mid A_{[n]}] P\left[ \bigcap_{j=1}^{n} A_j \right] \), \( A_{[n]} := \bigcap_{j=1}^{n} A_j \).

Example 4.11 (Guessing game). Consider a deck of four cards labeled \( A, K, Q, T \). Cards are shuffled and laid face down. They will be turned over one by one. Before each card is turned over, you predict
\[ \bullet \text{ next card is } A. \text{ If it is, you win, otherwise you lose; or} \]
\[ \bullet \text{ next card is not } A. \text{ If it is, you lose, otherwise continue to next card.} \]

General analysis of the guessing game: Suppose there are \( n \) cards with one special card marked \( \ast \). You win if you predict when the card comes up. What strategy gives you the best chance to win?

Claim: Regardless of your strategy, \( P[\text{Win}] = 1/n \).

Proof by induction. Let \( W \) denote the event that you win. Clearly, for \( n = 1 \), \( P[W] = 1 \) and the claim is true. Now, we assume our claim is true for an \( n - 1 \) card deck and use this to prove it for an \( n \) card deck. Let \( A_1 \) be the event that you assert that the first card is special and let \( S_1 \) be the event that the first card is special.
1st assertion outcome/probability/win or lose

\[
\begin{align*}
A_1 & \quad \{ S_1 \quad \frac{1}{n} \quad \text{win} \\
A_1^c & \quad \{ S_1^c \quad \frac{(n-1)}{n} \quad \text{lose} \\
A_1^c & \quad \{ S_1^c \quad \frac{(n-1)}{n} \quad \text{lose} \\
\end{align*}
\]

Therefore,

\[
P[W_n \mid A_1^c] = P[S_1^c] \times P[W_{n-1} \mid S_1^c] = \frac{n-1}{n} \times \frac{1}{n-1} = \frac{1}{n}
\]

and also

\[
P[W_n \mid A_1] = \frac{1}{n}.
\]

Hence, your strategy on the first card does not matter. After the first card, the remaining game is the same, but with a deck of \( n - 1 \) cards. We already know that the strategy for the first card does not matter in that game. Therefore, the strategy for the second card (assuming we get that far) does not matter, and so on.

4.3. Odds.

**Definition 4.12.** The (chance) odds of an event \( A \in \mathcal{E} \) happening are

\[
\text{odds}(A) := \frac{P[A]}{1 - P[A]}.
\]

Conversely, if \( \text{odds}(A) = x \), then \( P[A] = x/(1 + x) \).

For a bet on \( A \), the payoff odds are defined by

\[
\begin{align*}
\text{net gain if win} & \quad \text{net gain if lose} \\
\end{align*}
\]

**Example 4.13 (Diagnosis).** A doctor is initially 60% certain that a patient has some disease, so she orders a test. The following statistics are known about the accuracy of the test.

- If the patient has the disease, the test almost always gives a positive result.
- For a diabetic patient without the disease, there are 30% false positives.

Our patient is diabetic and obtained a positive result. What is the probability the patient has the disease?

Let \( D := \{ \text{has disease} \} \) and \(+ \) (resp. \(- \)) denote the event that the test is positive (resp. negative). Then

\[
P[D \mid +] = \frac{P[+ \mid D]P[D]}{P[+ \mid D]P[D] + P[+ \mid D^c]P[D^c]} = \frac{0.6}{0.6 + 0.3(0.4)} = \frac{0.6}{0.72} = \frac{5}{6}.
\]
Alternatively, we can compute the posterior odds:

\[
\text{posterior odds on } D \text{ given } + = \frac{P[D | +]}{P[D^c | +]} = \frac{P[D \cap +] / P[+]}{P[D^c \cap +] / P[+]} = \frac{P[D] P[+ | D]}{P[D^c] P[+ | D^c]} = \left[ \frac{\text{Prior odds on } D}{\text{Likelihood of } + \text{ given } D} \right] \times \left[ \frac{\text{Likelihood of } + \text{ given } D^c}{\text{Likelihood of } + \text{ given } D} \right].
\]