4 Conditional Probability and Independence

Example 4.1. Assume that you have a bag with 11 cubes, 7 of which have a fuzzy surface and 4 are smooth. Out of the 7 fuzzy ones, 3 are red and 4 are blue; out of 4 smooth ones, 2 are red and 2 are blue. So, there are 5 red and 6 blue cubes. Other than color and fuzziness, the cubes have no other distinguishing characteristics.

You plan to pick a cube out of the bag at random, but forget to wear gloves. Before you start your experiment, the probability that the selected cube is red is $\frac{5}{11}$. Now, you reach into the bag, grab a cube, and notice it is fuzzy (but you do not take it out or note its color in any other way). Clearly, the probability should now change to $\frac{3}{7}$!

Your experiment clearly has 11 outcomes. Consider the events $R$, $B$, $F$, $S$ that the selected cube is red, blue, fuzzy, and smooth, respectively. We observed that $P(R) = \frac{5}{11}$. For the probability of a red cube, conditioned on it being fuzzy, we do not have notation, so we introduce it here: $P(R|F) = \frac{3}{7}$. Note that this also equals

$$P(R \cap F) = \frac{P(\text{the selected ball is red and fuzzy})}{P(\text{the selected ball is fuzzy})}.$$

This conveys the idea that with additional information the probability must be adjusted. This is common in real life. Say bookies estimate your basketball team’s chances of winning a particular game to be 0.6, 24 hours before the game starts. Two hours before the game starts, however, it becomes known that your team’s star player is out with a sprained ankle. You cannot expect that the bookies’ odds will remain the same and they change, say, to 0.3. Then, the game starts and at half-time your team leads by 25 points. Again, the odds will change, say to 0.8. Finally, when complete information (that is, the outcome of your experiment, the game in this case) is known, all probabilities are trivial, 0 or 1.

For the general definition, take events $A$ and $B$, and assume that $P(B) > 0$. The conditional probability of $A$ given $B$ equals

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

Example 4.2. Here is a question asked on Wall Street job interviews. (This is the original formulation; the macabre tone is not unusual for such interviews.)

“Let’s play a game of Russian roulette. You are tied to your chair. Here’s a gun, a revolver. Here’s the barrel of the gun, six chambers, all empty. Now watch me as I put two bullets into the barrel, into two adjacent chambers. I close the barrel and spin it. I put a gun to your head and pull the trigger. Click. Lucky you! Now I’m going to pull the trigger one more time. Which would you prefer: that I spin the barrel first or that I just pull the trigger?”

Assume that the barrel rotates clockwise after the hammer hits and is pulled back. You are given the choice between an unconditional and a conditional probability of death. The former,
if the barrel is spun again, remains 1/3. The latter, if the trigger is pulled without the extra spin, equals the probability that the hammer clicked on an empty slot, which is next to a bullet in the counterclockwise direction, and equals 1/4.

For a fixed condition $B$, and acting on events $A$, the conditional probability $Q(A) = P(A|B)$ satisfies the three axioms in Chapter 3. (This is routine to check and the reader who is more theoretically inclined might view it as a good exercise.) Thus, $Q$ is another probability assignment and all consequences of the axioms are valid for it.

**Example 4.3.** Toss two fair coins, blindfolded. Somebody tells you that you tossed at least one Heads. What is the probability that both tosses are Heads?

Here $A = \{\text{both H}\}$, $B = \{\text{at least one H}\}$, and

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\text{both H})}{P(\text{at least one H})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

**Example 4.4.** Toss a coin 10 times. If you know (a) that exactly 7 Heads are tossed, (b) that at least 7 Heads are tossed, what is the probability that your first toss is Heads?

For (a),

$$P(\text{first toss H}|\text{exactly 7 H’s}) = \frac{\binom{9}{6}}{\binom{10}{7}} \cdot \frac{1}{2^{10}} = \frac{7}{10}.$$

Why is this not surprising? Conditioned on 7 Heads, they are equally likely to occur on any given 7 tosses. If you choose 7 tosses out of 10 at random, the first toss is included in your choice with probability $\frac{7}{10}$.

For (b), the answer is, after canceling $\frac{1}{2^{10}}$,

$$\frac{\binom{9}{6} + \binom{9}{7} + \binom{9}{8}}{\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}} = \frac{65}{88} \approx 0.7386.$$

Clearly, the answer should be a little larger than before, because this condition is more advantageous for Heads.

Conditional probabilities are sometimes *given*, or can be easily determined, especially in sequential random experiments. Then, we can use

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1),$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2),$$

etc.

**Example 4.5.** An urn contains 10 black and 10 white balls. Draw 3 (a) without replacement, and (b) with replacement. What is the probability that all three are white?

We already know how to do part (a):
1. Number of outcomes: \( \binom{20}{3} \).

2. Number of ways to select 3 balls out of 10 white ones: \( \binom{10}{3} \).

Our probability is then \( \frac{\binom{10}{3}}{\binom{20}{3}} \).

To do this problem another way, imagine drawing the balls sequentially. Then, we are computing the probability of the intersection of the three events: \( P(1\text{st ball is white}, 2\text{nd ball is white}, \text{and } 3\text{rd ball is white}) \). The relevant probabilities are:

1. \( P(1\text{st ball is white}) = \frac{1}{2} \).
2. \( P(2\text{nd ball is white}|1\text{st ball is white}) = \frac{9}{19} \).
3. \( P(3\text{rd ball is white}|1\text{st two picked are white}) = \frac{8}{18} \).

Our probability is, then, the product \( \frac{1}{2} \cdot \frac{9}{19} \cdot \frac{8}{18} \), which equals, as it must, what we obtained before.

This approach is particularly easy in case (b), where the previous colors of the selected balls do not affect the probabilities at subsequent stages. The answer, therefore, is \( \left( \frac{1}{2} \right)^3 \).

**Theorem 4.1.** First Bayes’ formula. Assume that \( F_1, \ldots, F_n \) are pairwise disjoint and that \( F_1 \cup \ldots \cup F_n = \Omega \), that is, exactly one of them always happens. Then, for an event \( A \),

\[
P(A) = P(F_1)P(A|F_1) + P(F_2)P(A|F_2) + \ldots + P(F_n)P(A|F_n).
\]

**Proof.**

\[
P(F_1)P(A|F_1) + P(F_2)P(A|F_2) + \ldots + P(F_n)P(A|F_n) = P(A \cap F_1) + \ldots + P(A \cap F_n)
\]

\[
= P(A \cap (F_1 \cup \ldots \cup F_n))
\]

\[
= P(A \cap F_1 \cup \ldots \cup F_n)
\]

\[
= P(A \cap \Omega) = P(A)
\]

We call an instance of using this formula “computing the probability by conditioning on which of the events \( F_i \) happens.” The formula is useful in sequential experiments, when you face different experimental conditions at the second stage depending on what happens at the first stage. Quite often, there are just two events \( F_i \), that is, an event \( F \) and its complement \( F^c \), and we are thus conditioning on whether \( F \) happens or not.

**Example 4.6.** Flip a fair coin. If you toss Heads, roll 1 die. If you toss Tails, roll 2 dice. Compute the probability that you roll exactly one 6.
Here you condition on the outcome of the coin toss, which could be Heads (event $F$) or Tails (event $F^c$). If $A = \{\text{exactly one 6}\}$, then $P(A|F) = \frac{1}{6}$, $P(A|F^c) = \frac{25}{36}$, $P(F) = P(F^c) = \frac{1}{2}$ and so

$$P(A) = P(F)P(A|F) + P(F^c)P(A|F^c) = \frac{2}{9}.$$

**Example 4.7.** Roll a die, then select at random, without replacement, as many cards from the deck as the number shown on the die. What is the probability that you get at least one Ace?

Here $F_i = \{\text{number shown on the die is } i\}$, for $i = 1, \ldots, 6$. Clearly, $P(F_i) = \frac{1}{6}$. If $A$ is the event that you get at least one Ace,

1. $P(A|F_1) = 1 - \frac{48}{52}$.

Therefore, by Bayes’ formula,

$$P(A) = \frac{1}{6} \left( 1 + 1 - \frac{48}{52} + 1 - \frac{48}{52} + 1 - \frac{48}{52} + 1 - \frac{48}{52} \right).$$

**Example 4.8.** *Coupon collector problem*, revisited. As promised, we will develop a computationally much better formula than the one in Example 3.9. This will be another example of conditioning, whereby you (1) reinterpret the problem as a sequential experiment and (2) use Bayes’ formula with “conditions” $F_i$ being relevant events at the first stage of the experiment.

Here is how it works in this example. Let $p_{k,r}$ be the probability that exactly $r$ (out of a total of $n$) birthdays are represented among $k$ people; we call the event $A$. We will fix $n$ and let $k$ and $r$ be variable. Note that $p_{k,n}$ is what we computed by the inclusion-exclusion formula.

At the first stage you have $k - 1$ people; then the $k$'th person arrives on the scene. Let $F_1$ be the event that there are $r$ birthdays represented among the $k - 1$ people and let $F_2$ be the event that there are $r - 1$ birthdays represented among the $k - 1$ people. Let $F_3$ be the event that any other number of birthdays occurs with $k - 1$ people. Clearly, $P(A|F_3) = 0$, as the newcomer contributes either 0 or 1 new birthdays. Moreover, $P(A|F_1) = \frac{r}{n}$, the probability that the newcomer duplicates one of the existing $r$ birthdays, and $P(A|F_2) = \frac{n-r+1}{n}$, the probability that the newcomer does not duplicate any of the existing $r - 1$ birthdays. Therefore,

$$p_{k,r} = P(A) = P(A|F_1)P(F_1) + P(A|F_2)P(F_2) = \frac{r}{n} \cdot p_{k-1,r} + \frac{n-r+1}{n} \cdot p_{k-1,r-1},$$

for $k, r \geq 1$, and this, together with the boundary conditions

$$p_{0,0} = 1,$$

$$p_{k,r} = 0, \text{ for } 0 \leq k < r,$$

$$p_{k,0} = 0, \text{ for } k > 0,$$

makes the computation fast and precise.
Theorem 4.2. Second Bayes’ formula. Let $F_1, \ldots, F_n$ and $A$ be as in Theorem 4.1. Then

$$P(F_j|A) = \frac{P(F_j \cap A)}{P(A)} = \frac{P(A|F_j)P(F_j)}{P(A|F_1)P(F_1) + \ldots + P(A|F_n)P(F_n)}.$$ 

An event $F_j$ is often called a hypothesis, $P(F_j)$ its prior probability, and $P(F_j|A)$ its posterior probability.

Example 4.9. We have a fair coin and an unfair coin, which always comes out Heads. Choose one at random, toss it twice. It comes out Heads both times. What is the probability that the coin is fair?

The relevant events are $F = \{\text{fair coin}\}$, $U = \{\text{unfair coin}\}$, and $B = \{\text{both tosses H}\}$. Then $P(F) = P(U) = \frac{1}{2}$ (as each coin is chosen with equal probability). Moreover, $P(B|F) = \frac{1}{4}$, and $P(B|U) = 1$. Our probability then is

$$\frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot 1} = \frac{1}{5}.$$ 

Example 4.10. A factory has three machines, $M_1$, $M_2$ and $M_3$, that produce items (say, lightbulbs). It is impossible to tell which machine produced a particular item, but some are defective. Here are the known numbers:

<table>
<thead>
<tr>
<th>machine</th>
<th>proportion of items made</th>
<th>prob. any made item is defective</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>0.2</td>
<td>0.001</td>
</tr>
<tr>
<td>$M_2$</td>
<td>0.3</td>
<td>0.002</td>
</tr>
<tr>
<td>$M_3$</td>
<td>0.5</td>
<td>0.003</td>
</tr>
</tbody>
</table>

You pick an item, test it, and find it is defective. What is the probability that it was made by machine $M_2$?

The best way to think about this random experiment is as a two-stage procedure. First you choose a machine with the probabilities given by the proportion. Then, that machine produces an item, which you then proceed to test. (Indeed, this is the same as choosing the item from a large number of them and testing it.)

Let $D$ be the event that an item is defective and let $M_i$ also denote the event that the item was made by machine $i$. Then, $P(D|M_1) = 0.001$, $P(D|M_2) = 0.002$, $P(D|M_3) = 0.003$, $P(M_1) = 0.2$, $P(M_2) = 0.3$, $P(M_3) = 0.5$, and so

$$P(M_2|D) = \frac{0.002 \cdot 0.3}{0.001 \cdot 0.2 + 0.002 \cdot 0.3 + 0.003 \cdot 0.5} \approx 0.26.$$ 

Example 4.11. Assume 10% of people have a certain disease. A test gives the correct diagnosis with probability of 0.8; that is, if the person is sick, the test will be positive with probability 0.8, but if the person is not sick, the test will be positive with probability 0.2. A random person from
the population has tested positive for the disease. What is the probability that he is actually sick? (No, it is not 0.8!)

Let us define the three relevant events: $S = \{\text{sick}\}$, $H = \{\text{healthy}\}$, $T = \{\text{tested positive}\}$.

Now, $P(H) = 0.9$, $P(S) = 0.1$, $P(T|H) = 0.2$ and $P(T|S) = 0.8$. We are interested in

$$P(S|T) = \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|H)P(H)} = \frac{8}{26} \approx 31\%.$$ 

Note that the prior probability $P(S)$ is very important! Without a very good idea about what it is, a positive test result is difficult to evaluate: a positive test for HIV would mean something very different for a random person as opposed to somebody who gets tested because of risky behavior.

**Example 4.12. O. J. Simpson’s first trial, 1995.** The famous sports star and media personality O. J. Simpson was on trial in Los Angeles for murder of his wife and her boyfriend. One of the many issues was whether Simpson’s history of spousal abuse could be presented by prosecution at the trial; that is, whether this history was “probative,” i.e., it had some evidentiary value, or whether it was merely “prejudicial” and should be excluded. Alan Dershowitz, a famous professor of law at Harvard and a consultant for the defense, was claiming the latter, citing the statistics that $< 0.1\%$ of men who abuse their wives end up killing them. As J. F. Merz and J. C. Caulkins pointed out in the journal Chance (Vol. 8, 1995, pg. 14), this was the wrong probability to look at!

We need to start with the fact that a woman is murdered. These numbered 4,936 in 1992, out of which 1,430 were killed by partners. In other words, if we let

$$A = \{\text{the (murdered) woman was abused by the partner}\},$$

$$M = \{\text{the woman was murdered by the partner}\},$$

then we estimate the prior probabilities $P(M) = 0.29$, $P(M^c) = 0.71$, and what we are interested in is the posterior probability $P(M|A)$. It was also commonly estimated at the time that about 5\% of the women had been physically abused by their husbands. Thus, we can say that $P(A|M^c) = 0.05$, as there is no reason to assume that a woman murdered by somebody else was more or less likely to be abused by her partner. The final number we need is $P(A|M)$. Dershowitz states that “a considerable number” of wife murderers had previously assaulted them, although “some” did not. So, we will (conservatively) say that $P(A|M) = 0.5$. (The two-stage experiment then is: choose a murdered woman at random; at the first stage, she is murdered by her partner, or not, with stated probabilities; at the second stage, she is among the abused women, or not, with probabilities depending on the outcome of the first stage.) By Bayes’ formula,

$$P(M|A) = \frac{P(M)P(A|M)}{P(M)P(A|M) + P(M^c)P(A|M^c)} = \frac{29}{36.1} \approx 0.8.$$ 

The law literature studiously avoids quantifying concepts such as probative value and reasonable doubt. Nevertheless, we can probably say that 80\% is considerably too high, compared to the prior probability of 29\%, to use as a sole argument that the evidence is not probative.
Independence

Events $A$ and $B$ are independent if $P(A \cap B) = P(A)P(B)$ and dependent (or correlated) otherwise.

Assuming that $P(B) > 0$, one could rewrite the condition for independence,

$$P(A|B) = P(A),$$

so the probability of $A$ is unaffected by knowledge that $B$ occurred. Also, if $A$ and $B$ are independent,

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c),$$

so $A$ and $B^c$ are also independent — knowing that $B$ has not occurred also has no influence on the probability of $A$. Another fact to notice immediately is that disjoint events with nonzero probability cannot be independent: given that one of them happens, the other cannot happen and thus its probability drops to zero.

Quite often, independence is an assumption and it is the most important concept in probability.

**Example 4.13.** Pick a random card from a full deck. Let $A = \{\text{card is an Ace}\}$ and $R = \{\text{card is red}\}$. Are $A$ and $R$ independent?

We have $P(A) = \frac{1}{13}$, $P(R) = \frac{1}{2}$, and, as there are two red Aces, $P(A \cap R) = \frac{2}{52} = \frac{1}{26}$. The two events are independent — the proportion of aces among red cards is the same as the proportion among all cards.

Now, pick two cards out of the deck sequentially without replacement. Are $F = \{\text{first card is an Ace}\}$ and $S = \{\text{second card is an Ace}\}$ independent?

Now $P(F) = P(S) = \frac{1}{13}$ and $P(S|F) = \frac{3}{51}$, so they are not independent.

**Example 4.14.** Toss 2 fair coins and let $F = \{\text{Heads on 1st toss}\}$, $S = \{\text{Heads on 2nd toss}\}$. These are independent. You will notice that here the independence is in fact an assumption.

How do we define independence of more than two events? We say that events $A_1, A_2, \ldots, A_n$ are independent if

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}),$$

where $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ are arbitrary indices. The occurrence of any combination of events does not influence the probability of others. Again, it can be shown that, in such a collection of independent events, we can replace an $A_i$ by $A_i^c$ and the events remain independent.

**Example 4.15.** Roll a four sided fair die, that is, choose one of the numbers 1, 2, 3, 4 at random. Let $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. Check that these are pairwise independent (each pair is independent), but not independent.
Indeed, \( P(A) = P(B) = P(C) = \frac{1}{2} \) and \( P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} \) and pairwise independence follows. However,

\[
P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8}.
\]

The simple reason for lack of independence is

\[ A \cap B \subseteq C, \]

so we have complete information on the occurrence of \( C \) as soon as we know that \( A \) and \( B \) both happen.

**Example 4.16.** You roll a die, your friend tosses a coin.

- If you roll 6, you win outright.
- If you do not roll 6 and your friend tosses Heads, you lose outright.
- If neither, the game is repeated until decided.

What is the probability that you win?

One way to solve this problem certainly is this:

\[
P(\text{win}) = P(\text{win on 1st round}) + P(\text{win on 2nd round}) + P(\text{win on 3rd round}) + \ldots
\]

\[
= \frac{1}{6} + \left( \frac{5}{6} \cdot \frac{1}{2} \right) \frac{1}{6} + \left( \frac{5}{6} \cdot \frac{1}{2} \right)^2 \frac{1}{6} + \ldots,
\]

and then we sum the geometric series. **Important note:** we have implicitly assumed independence between the coin and the die, as well as between different tosses and rolls. This is very common in problems such as this!

You can avoid the nuisance, however, by the following trick. Let

\[
D = \{ \text{game is decided on 1st round} \},
\]

\[
W = \{ \text{you win} \}.
\]

The events \( D \) and \( W \) are independent, which one can certainly check by computation, but, in fact, there is a very good reason to conclude so immediately. The crucial observation is that, provided that the game is not decided in the 1st round, you are thereafter facing the same game with the same winning probability; thus

\[
P(W|D^c) = P(W).
\]

In other words, \( D^c \) and \( W \) are independent and then so are \( D \) and \( W \), and therefore

\[
P(W) = P(W|D).
\]
This means that one can solve this problem by computing the relevant probabilities for the 1st round:

\[ P(W|D) = \frac{P(W \cap D)}{P(D)} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{51}{62}} = \frac{2}{7}, \]

which is our answer.

**Example 4.17. Craps.** Many casinos allow you to bet even money on the following game. Two dice are rolled and the sum \( S \) is observed.

- If \( S \in \{7, 11\} \), you win immediately.
- If \( S \in \{2, 3, 12\} \), you lose immediately.
- If \( S \in \{4, 5, 6, 8, 9, 10\} \), the pair of dice is rolled repeatedly until one of the following happens:
  - \( S \) repeats, in which case you win.
  - 7 appears, in which case you lose.

What is the winning probability?

Let us look at all possible ways to win:

1. You win on the first roll with probability \( \frac{8}{36} \).
2. Otherwise,
   - you roll a 4 (probability \( \frac{3}{36} \)), then win with probability \( \frac{3}{3 + 6} = \frac{1}{3} \);
   - you roll a 5 (probability \( \frac{4}{36} \)), then win with probability \( \frac{4}{4 + 6} = \frac{2}{5} \);
   - you roll a 6 (probability \( \frac{5}{36} \)), then win with probability \( \frac{5}{5 + 6} = \frac{5}{11} \);
   - you roll a 8 (probability \( \frac{5}{36} \)), then win with probability \( \frac{5}{5 + 6} = \frac{5}{11} \);
   - you roll a 9 (probability \( \frac{4}{36} \)), then win with probability \( \frac{4}{4 + 6} = \frac{2}{5} \);
   - you roll a 10 (probability \( \frac{3}{36} \)), then win with probability \( \frac{3}{3 + 6} = \frac{1}{3} \).

Using Bayes’ formula,

\[ P(\text{win}) = \frac{8}{36} + 2 \left( \frac{3}{36} \cdot \frac{1}{3} + \frac{4}{36} \cdot \frac{2}{5} + \frac{5}{36} \cdot \frac{5}{11} \right) \approx 0.4929, \]

a decent game by casino standards.
Bernoulli trials

Assume \( n \) independent experiments, each of which is a success with probability \( p \) and, thus, failure with probability \( 1 - p \).

In a sequence of \( n \) Bernoulli trials, \( P(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k (1 - p)^{n-k} \).

This is because the successes can occur on any subset \( S \) of \( k \) trials out of \( n \), i.e., on any \( S \subset \{1, \ldots, n\} \) with cardinality \( k \). These possibilities are disjoint, as exactly \( k \) successes cannot occur on two different such sets. There are \( \binom{n}{k} \) such subsets; if we fix such an \( S \), then successes must occur on \( k \) trials in \( S \) and failures on all \( n - k \) trials not in \( S \); the probability that this happens, by the assumed independence, is \( p^k (1 - p)^{n-k} \).

Example 4.18. A machine produces items which are independently defective with probability \( p \). Let us compute a few probabilities:

1. \( P(\text{exactly two items among the first 6 are defective}) = \binom{6}{2} p^2 (1 - p)^4 \).
2. \( P(\text{at least one item among the first 6 is defective}) = 1 - P(\text{no defects}) = 1 - (1 - p)^6 \)
3. \( P(\text{at least 2 items among the first 6 are defective}) = 1 - (1 - p)^6 - 6p(1 - p)^5 \)
4. \( P(\text{exactly 100 items are made before 6 defective are found}) \) equals

\[
P(100\text{th item defective, exactly 5 items among 1st 99 defective}) = p \cdot \binom{99}{5} p^5 (1 - p)^{94}.
\]

Example 4.19. Problem of Points. This involves finding the probability of \( n \) successes before \( m \) failures in a sequence of Bernoulli trials. Let us call this probability \( p_{n,m} \).

\[
p_{n,m} = P(\text{in the first } m + n - 1 \text{ trials, the number of successes is } \geq n) = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1 - p)^{n+m-1-k}.
\]

The problem is solved, but it needs to be pointed out that computationally this is not the best formula. It is much more efficient to use the recursive formula obtained by conditioning on the outcome of the first trial. Assume \( m, n \geq 1 \). Then,

\[
p_{n,m} = P(\text{first trial is a success}) \cdot P(n - 1 \text{ successes before } m \text{ failures}) + P(\text{first trial is a failure}) \cdot P(n \text{ successes before } m - 1 \text{ failures})
\]

\[
= p \cdot p_{n-1,m} + (1 - p) \cdot p_{n,m-1}.
\]

Together with boundary conditions, valid for \( m, n \geq 1 \),

\[
p_{n,0} = 0, \quad p_{0,m} = 1,
\]
which allows for very speedy and precise computations for large \( m \) and \( n \).

**Example 4.20.** *Best of 7.* Assume that two equally matched teams, \( A \) and \( B \), play a series of games and that the first team that wins four games is the overall winner of the series. As it happens, team \( A \) lost the first game. What is the probability it will win the series? Assume that the games are Bernoulli trials with success probability \( \frac{1}{2} \).

We have

\[
P(A \text{ wins the series}) = P(4 \text{ successes before 3 failures})
= \sum_{k=4}^{6} \binom{6}{k} \left( \frac{1}{2} \right)^6 = \frac{15 + 6 + 1}{2^6} \approx 0.3438.
\]

**Example 4.21.** *Banach Matchbox Problem.* A mathematician carries two matchboxes, each originally containing \( n \) matches. Each time he needs a match, he is equally likely to take it from either box. What is the probability that, upon reaching for a box and finding it empty, there are exactly \( k \) matches still in the other box? Here, \( 0 \leq k \leq n \).

Let \( A_1 \) be the event that matchbox 1 is the one discovered empty and that, at that instant, matchbox 2 contains \( k \) matches. Before this point, he has used \( n + n - k \) matches, \( n \) from matchbox 1 and \( n - k \) from matchbox 2. This means that he has reached for box 1 exactly \( n \) times in \( (n + n - k) \) trials and for the last time at the \( (n + 1 + n - k) \)th trial. Therefore, our probability is

\[
2 \cdot P(A_1) = 2 \cdot \frac{1}{2} \binom{2n - k}{n} \frac{1}{2^{2n-k}} = \binom{2n - k}{n} \frac{1}{2^{2n-k}}.
\]

**Example 4.22.** Each day, you independently decide, with probability \( p \), to flip a fair coin. Otherwise, you do nothing. (a) What is the probability of getting exactly 10 Heads in the first 20 days? (b) What is the probability of getting 10 Heads before 5 Tails?

For (a), the probability of getting Heads is \( p/2 \) independently each day, so the answer is

\[
\binom{20}{10} \left( \frac{p}{2} \right)^{10} \left( 1 - \frac{p}{2} \right)^{10}.
\]

For (b), you can disregard days at which you do not flip to get

\[
\sum_{k=10}^{14} \binom{14}{k} \frac{1}{2^{14}}.
\]

**Example 4.23.** You roll a die and your score is the number shown on the die. Your friend rolls five dice and his score is the number of 6’s shown. Compute (a) the probability of event \( A \) that the two scores are equal and (b) the probability of event \( B \) that your friend’s score is strictly larger than yours.
In both cases we will condition on your friend’s score — this works a little better in case (b) than conditioning on your score. Let $F_i$, $i = 0, \ldots, 5$, be the event that your friend’s score is $i$. Then, $P(A|F_i) = \frac{1}{6}$ if $i \geq 1$ and $P(A|F_0) = 0$. Then, by the first Bayes’ formula, we get

$$P(A) = \sum_{i=1}^{5} P(F_i) \cdot \frac{1}{6} = \frac{1}{6} (1 - P(F_0)) = \frac{1}{6} - \frac{5}{6^6} \approx 0.0997.$$ 

Moreover, $P(B|F_i) = \frac{i-1}{6}$ if $i \geq 2$ and 0 otherwise, and so

$$P(B) = \sum_{i=1}^{5} P(F_i) \cdot \frac{i - 1}{6}$$

$$= \frac{1}{6} \sum_{i=1}^{5} i \cdot P(F_i) - \frac{1}{6} \sum_{i=1}^{5} P(F_i)$$

$$= \frac{1}{6} \sum_{i=1}^{5} i \cdot P(F_i) - \frac{1}{6} + \frac{5}{6^6}$$

$$= \frac{1}{6} \sum_{i=1}^{5} i \cdot \binom{5}{i} \left( \frac{1}{6} \right)^i \left( \frac{5}{6} \right)^{5-i} - \frac{1}{6} + \frac{5}{6^6}$$

$$= \frac{1}{6} \cdot \frac{5}{6} - \frac{1}{6} + \frac{5}{6^6} \approx 0.0392.$$ 

The last equality can be obtained by computation, but we will soon learn why the sum has to equal $\frac{5}{6}$.

**Problems**

1. Consider the following game. Pick one card at random from a full deck of 52 cards. If you pull an Ace, you win outright. If not, then you look at the value of the card (K, Q, and J count as 10). If the number is 7 or less, you lose outright. Otherwise, you select (at random, without replacement) that number of additional cards from the deck. (For example, if you picked a 9 the first time, you select 9 more cards.) If you get at least one Ace, you win. What are your chances of winning this game?

2. An item is defective (independently of other items) with probability 0.3. You have a method of testing whether the item is defective, but it does not always give you correct answer. If the tested item is defective, the method detects the defect with probability 0.9 (and says it is good with probability 0.1). If the tested item is good, then the method says it is defective with probability 0.2 (and gives the right answer with probability 0.8).

A box contains 3 items. You have tested all of them and the tests detect no defects. What is the probability that none of the 3 items is defective?