Mathematical Foundations of Computer Science
Lecture Outline
September 26, 2024

Introduction to Probability

Probability theory has many applications in engineering, medicine, etc. It has also found many useful applications in computer science, such as cryptography, networking, game theory etc. Many algorithms are randomized and we need probability theory to analyze them. In this course, our goal is to understand how to describe uncertainty using probabilistic arguments. To do this we first have to define a probabilistic model.

A probabilistic model is a mathematical description of a random process or an experiment. In a random process exactly one outcome from a set of outcomes is sure to occur but no outcome can be predicted with certainty. For example, tossing a coin is an experiment. Below are definitions of entities associated with the probabilistic model.

- The *sample space* of a random process or experiment is the set of all possible outcomes. The sample space is often denoted by Ω. Since we are going to study discrete probability Ω will be finite or countably infinite (such as integers and not real numbers).
- The *probability space* is a sample space together with a *probability distribution* in which a probability is assigned to each outcome $\omega \in \Omega$, such that
  - $0 \leq \Pr[\omega] \leq 1$
  - $\sum_{\omega \in \Omega} \Pr[\omega] = 1$

In an experiment we are usually interested in the probability with an event occurs. For example, when tossing a coin we may be interested in knowing the probability that the result is heads. Below we define formally what an event is and what does it mean to calculate the probability of an event.

- A subset of the sample space is called an *event*.
- For any event, $A \subseteq \Omega$, the probability of A is defined as

$$\Pr[A] = \sum_{\omega \in A} \Pr[\omega]$$

We are now ready to work through some problems. Before we proceed, keep in mind that probability is a slippery topic; it is very easy to make mistakes. Solving the problem systematically is the key to avoid mistakes. The following four-step process that is described in the notes by Lehman and Leighton is a way to systematically solve the problem at hand.

(a) Define the sample space, Ω, of the experiment, i.e., find the set of all possible outcomes of the experiment.

(b) Define the probability distribution.

(c) Find the event of interest, A, i.e., find the subset of outcomes, $A \subseteq \Omega$ that are of interest.

(d) Compute the probability of A by adding up the probabilities of the outcomes in A.

Example. On flipping a fair coin what is the probability that the result is heads?

Solution. $\Omega = \{H, T\}, \Pr[H] = \Pr[T] = 1/2, A = \{H\}, \Pr[A] = 1/2.$

Example. Consider a biased coin in which the probability of heads is 1/3. Suppose we flip the coin twice. What is the probability that we obtain one tails and one heads?

Solution. $\Omega = \{HH, HT, TH, TT\}$. The probability distribution is given by

$$\begin{aligned} \Pr[HH] &= \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} \\ \Pr[HT] &= \frac{1}{3} \times \frac{2}{3} = \frac{2}{9} \\ \Pr[TH] &= \frac{2}{3} \times \frac{1}{3} = \frac{2}{9} \\ \Pr[TT] &= \frac{2}{3} \times \frac{2}{3} = \frac{4}{9} \end{aligned}$$

Note that the assigned probabilities form a valid probability distribution. Event $A = \{HT, TH\}$. The probability of the event A is given by

$$\Pr[A] = \Pr[HT] + \Pr[TH] = \frac{4}{9}$$

Example. We roll two dice. Compute the probability that the two numbers are equal when (i) two dice are distinct, (ii) the dice are indistinguishable.

Solution. (a) Each outcome of the experiment can be denoted by an ordered pair $(\omega_1, \omega_2), 1 \leq \omega_1, \omega_2 \leq 6$, where ω_1 and ω_2 are the numbers on dice 1 and dice 2 respectively. Note that $|\Omega| = 36$ and each outcome is equally likely. The event that the two numbers are equal is given by $A = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$. The probability that A occurs is given by

$$\Pr[A] = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

(b) When the die are indistinguishable, the order of the two numbers is not important, hence each outcome of the experiment can be denoted by a 2-set $\{\omega_1, \omega_2\}, 1 \leq \omega_1, \omega_2 \leq 6$, where ω_1 and ω_2 are the numbers on the two die. Note that $|\Omega| = 21$. Each outcome of

the form $\{\omega_1, \omega_2\}, \omega_1 \neq \omega_2$ occurs with a probability of $\frac{2}{36} = \frac{1}{18}$ and outcomes of the form $\{\omega, \omega\}$ occur with the probability of $\frac{1}{36}$. The event that the two numbers are equal is given by $A = \{\{1,1\},\{2,2\},\{3,3\},\{4,4\},\{5,5\},\{6,6\}\}$. The probability that A occurs is given by

$$\Pr[A] = 6 \times \frac{1}{36} = \frac{1}{6}$$

Example. Suppose we throw m distinct balls into n distinct bins. Assume that there is no bound on the number of balls that a bin contains. What is the probability that a particular bin, say bin 1, contains all the m balls?

Solution. Each outcome can be represented by a m-tuple $(\omega_1, \omega_2, \ldots, \omega_m)$, where ω_i denotes the bin that contains the ith ball. Note that $|\Omega| = n^m$ and each outcome is equally likely. Since there is only one way in which all balls can be in bin 1, the probability of this event is $\frac{1}{n^m}$.

Example. What is the probability of rolling a six-sided die six times and having all the numbers 1 through 6 result (in any order)?

Solution. Each element in Ω can be represented by $(\omega_1, \omega_2, \ldots, \omega_6)$, where ω_i is the number that results on the ith roll of the die. Using the multiplication rule we get $|\Omega| = 6^6$. Let $A \subseteq \Omega$ be the set of outcomes in which the numbers of the rolls are different. By multiplication rule $|A| = 6!$. Since each outcome is equally likely, the desired probability is given by

$$\frac{|A|}{|\Omega|} = \frac{6!}{6^6} = \frac{5}{324}$$

Example. On "Let's Make a Deal" show, there are three doors. There is a prize behind one of the doors and goats behind the other two. The contestant chooses a door. Then the host opens a different door behind which there is a goat. The contestant is then given a choice to either switch doors or to stay put. The contestant wins the prize if and only if the contestant chooses the door with the prize behind it. Is it to the contestant's benefit to switch doors?

Solution. Each outcome of the sample space can be denoted by a 3-tuple $(\omega_1, \omega_2, \omega_3)$, where ω_1 denotes the door hiding the prize, ω_2 denotes the door chosen by the contestant initially, and ω_3 is the door chosen by the host. Now, lets assign probabilities to each of the outcomes[1]. There are two types of outcomes, those in which $\omega_1 = \omega_2$ and those in which $\omega_1 \neq \omega_2$. It is easy to verify that there are 6 outcomes of each type. Each outcome of the first type occurs with a probability of $\frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}$. If the outcome is of the second type then there is only one choice for ω_3, i.e., there is only one choice of door for the host. Each outcome of the second type occurs with a probability $\frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}$. The event in which

[1]We are making the following assumptions:(i) the prize is equally likely to be behind any of the doors, (ii) the contestant is equally likely to choose any of the three doors, (iii) the host opens any of the possible doors with equal probability

the contestant switches doors and wins is the set of all outcomes in which $\omega_1 \neq \omega_2$. Since the size of this set is 6 and each outcome occurs with a probability of $\frac{1}{9}$ the probability of the contestant winning the prize by switching is $\frac{6}{9} = \frac{2}{3}$. Thus, it is to contestant's benefit to switch.

Inclusion-Exclusion Formula

For two events A and B we have

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

For three events A, B, and C, we have

$$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[B \cap C] - \Pr[A \cap C] + \Pr[A \cap B \cap C].$$

For events $A_1, A_2, \ldots, A_n$ in some probability space, let $S_1 = \{(i_1) | 1 \leq i_1 \leq n\}$, $S_2 = \{(i_1, i_2) | 1 \leq i_1 < i_2 \leq n\}$, and more generally let $S_p = \{(i_1, i_2, \ldots, i_p) | 1 \leq i_1 < i_2 < \ldots \leq i_p \leq n\}$. Then we have

$$\Pr[\cup_{i=1}^{n} A_i] = \sum_{i \in S_1} \Pr[A_i] - \sum_{(i_1,i_2) \in S_2} \Pr[A_{i_1} \cap A_{i_2}] + \sum_{(i_1,i_2,i_3) \in S_3} \Pr[A_{i_1} \cap A_{i_2} \cap A_{i_3}] - \cdots + (-1)^{n-1} \Pr[\cap_{x=1}^{n} A_x]$$

Note that there are $2^n - 1$ non-empty subsets of a set of n events. To compute the probability of the intersection of every such subset is not possible when n is large. In such cases we have to approximate the probability of a union of n events. The successive terms of the above formula actually give an overestimate and underestimate respectively of the actual probability. In many situations the upper-bound given by the first term itself is quite useful. It is called the *union-bound* and is given by

$$\Pr[\cup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]$$

Note that when the events are pairwise disjoint, the inequality in the above expression becomes an equality.

Example. Consider three flips of a fair coin. What is the probability that result is heads on the first flip or the third flip?

Solution. Let H_1 and H_2 denote the events that the first flip results in heads and the third flip results in heads respectively. By the inclusion-exclusion formula, we have

$$\begin{aligned} \Pr[H_1 \cup H_2] &= \Pr[H_1] + \Pr[H_2] - \Pr[H_1 \cap H_2] \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

Example. When three dice are rolled what is the probability that one of the dice results in 4?

Solution. Let $F_i, i \in \{1,2,3\}$ be the event that the ith dice results in a 4. We are interested in $\Pr[F_1 \cup F_2 \cup F_3]$. By inclusion-exclusion formula we have

$$\Pr[F_1 \cup F_2 \cup F_3] = \Pr[F_1]+\Pr[F_2]+\Pr[F_3]-\Pr[F_1 \cap F_2]-\Pr[F1 \cap F_3]-\Pr[F_2 \cap F_3]+\Pr[F_1 \cap F_2 \cap F_3]$$

Since the events F_1, F_2, F_3 are mutually independent we can rewrite the above expression as

$$\begin{aligned}
\Pr[F \cup F_2 \cup F_3] &= \Pr[F_1] + \Pr[F_2] + \Pr[F_3] - \Pr[F_1]\Pr[F_2] - \Pr[F1]\Pr[F_3] - \Pr[F_2]\Pr[F_3] \\
&\quad + \Pr[F_1]\Pr[F_2]\Pr[F_3] \\
&= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6} \times \frac{1}{6}\right) - \left(\frac{1}{6} \times \frac{1}{6}\right) - \left(\frac{1}{6} \times \frac{1}{6}\right) + \left(\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}\right) \\
&= \frac{91}{216}
\end{aligned}$$

An easier way to solve this is as follows. Let $\overline{F_i}$ be the complement of event F_i, $i = 1, 2, 3$.

$$\Pr[F_1 \cup F_2 \cup F_3] = 1 - \Pr[\overline{F_1} \cap \overline{F_2} \cap \overline{F_3}] = 1 - (5/6)^3 = \frac{91}{216}$$

Example. A coin is tossed 10 times. What is the probability that eight or more heads turn up?

Solution. Let E_i denote the event that exactly i heads turn up. We are interested in $\Pr[E_8 \cup E_9 \cup E_{10}]$. Since the events E_i are disjoint, we have

$$\Pr[E_8 \cup E_9 \cup E_{10}] = \Pr[E_8] + \Pr[E_9] + \Pr[E_{10}]$$

Note that $\Pr[E_i] = \binom{10}{i}/2^{10}$. Hence, we have

$$\Pr[E_8 \cup E_9 \cup E_{10}] = \frac{1}{2^{10}}\left(\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right) = \frac{56}{2^{10}}$$

Example. (Birthday Paradox) Suppose there are k people in a room and n days in a year. We are interested in the probability that there are at least two people in the room with the same birthday. What is the smallest value of k for which this probability is at least 1/2? Assume that it is equally likely for a person to be born on any of the n days of the year.

Solution. Let B be the event that at least two people in the room have the same birthday. We are interested in $\Pr[B]$.

$$\begin{aligned}
\Pr[B] &= 1 - \Pr[\overline{B}] \\
&= 1 - \frac{P(n,k)}{n^k}
\end{aligned}$$

For $n = 365$, the smallest value of k for which the RHS is at least $1/2$ is 23. If $k = 40$ then $\Pr[B] = 0.89$, and if $k = 60$ then $\Pr[B] = 0.994$. This means that if there are 60 people then it is almost certain that there exists two among them sharing the same birthday. To illustrate how good our model is, consider the set of presidents of the United States of America. Through Bill Clinton, 41 people belong to this set. The chances of two of them sharing the same birthday is at least 89%. Indeed, James Polk (11th president) and Warren Harding (29th president) are both born on Nov. 2.

Conditional Probability

We now introduce a very important concept of conditional probability. Conditional probability allows us to calculate the probability of an event when some partial information about the result of an experiment is known. As we shall see conditional probability is often a convenient way to calculate probabilities even when no information about the result of an experiment is available.

Suppose we want to calculate the probability of event A given that event B has already occured. We denote this by $\Pr[A|B]$ (read as "the probability of A given B"). Since we know that event B has occured our sample space reduces to the outcomes in B. Is this a valid probability space? No, because the sum of probabilities of the outcomes in B is less than 1. How do we change the probabilities so that this is a valid probability distribution while making sure that the relative probabilities of outcomes in B do not change? We do this by scaling the probability of all sample points in B by $\frac{1}{\Pr[B]}$. Thus for each sample point $\omega \in B$,

$$\Pr[\omega|B] = \frac{\Pr[\omega]}{\Pr[B]}$$

To calculate $\Pr[A|B]$ we just sum up the probabilities of sample points in $A \cap B$. Thus we get

$$\Pr[A|B] = \sum_{\omega \in A \cap B} \Pr[\omega|B] = \sum_{\omega \in A \cap B} \frac{\Pr[\omega]}{\Pr[B]} = \frac{\Pr[A \cap B]}{\Pr[B]}$$

In order to avoid division by 0, we only define $\Pr[A|B]$ when $\Pr[B] > 0$. Conditional probabilities can sometimes get tricky. To avoid pitfalls, it is best to use the above mathematical definition of conditional probability. Note that the R.H.S. of the above equation are unconditional probabilities.