EE 178/278A Probabilistic Systems Analysis

Spring 2014 Tse/Hussami Lecture 1

Introduction to Probability[1]

Reading: Chapter 1 from Bertsekas and Tsitsiklis.

Life is full of uncertainty.

Probability is a framework to deal with uncertainty. Probability theory has its origins in gambling — analyzing card games, dice, roulette wheels. Today it is an essential tool in engineering and science. No less so in electrical engineering and computer science, where its use is widespread in algorithms, systems, learning theory, artificial intelligence, communications, control theory and signal processing.

We list some applications that we will cover in the following weeks. There are several forms of uncertainty in each of them, which can be modeled using a probabilistic model.

1. Multiplexing in networks: A typical question that arises when building a network for multiple users is how much hardware is needed to accommodate them. For example, it is important to determine the size of buffers in routers and the capacity of communication links so that packets do not get dropped during congestion. However, the number of users of a network at any given time is uncertain, as well as their usage. Probability can be used to model the behavior of the users.

2. Digital links: Data links are used to transmit and receive digital information. The uncertainty in this application is due to the fact that certain links might fail, and there can be noise in the communication channel. Therefore, some bits might not be communicated reliably. Probability can be used to model the noise and reliability of such system, as well as to predict its performance.

3. Tracking and prediction: An important example is designing self-driving cars. In this application we need to be able to predict where an approaching car is going based on its previous positions. The uncertainty in this setting comes from the decisions taken by the nearby drivers as well as the noise in measuring previous positions.

4. Speech recognition: The difficulty of this task comes from the fact that different speakers have different pronunciations, and the input to the microphone depends on the noise in the environment. Some of the uncertainty can be reduced by training the algorithm on each individual. However, the content of the conversation and the background noise are still random and can be modeled using probability.

Here are some typical statements that you might see concerning probability:

1. The chance of getting a flush in a 5-card poker hand is about 2 in 1000.

2. The chance that a packet is dropped at a router is less than 10^{-3}.

[1] Part of these notes are adapted from the course notes of EECS 70 at Berkeley.

3. The chance that a bit is communicated in error is less than 10^{-7}.

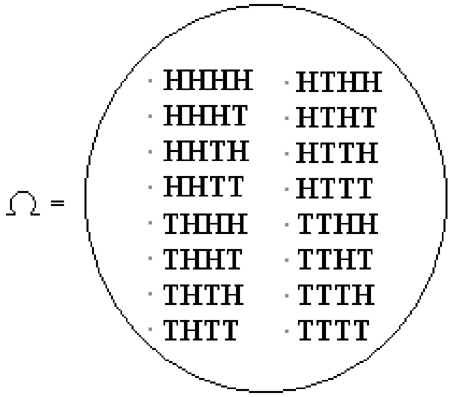
4. The chance that a sentence is recognized wrongly by Siri is less than 10%.

5. There is a 30% chance of a magnitude 8.0 earthquake in Northern California before 2030.

Implicit in all such statements is the notion of an underlying **probability space**. This may be the result of a random experiment that we have ourselves constructed (as in 1, 2, 3 and 4 above), or some model we build of the real world (as in 5 above). None of these statements makes sense unless we specify the probability space we are talking about: for this reason, statements like 5 (which are typically made without this context) are almost content-free.

Let us try to understand all this more clearly. The first important notion here is that of a **random experiment**. An example of such an experiment is tossing a coin 4 times. An outcome of the experiment might be $HTHT$ or it might be $HHHT$. Note that the random experiment results in exactly one outcome. The total number of distinct possible outcomes for this experiment is $2^4 = 16$. If the coin is fair then all these 16 outcomes are equally likely, so the chance that there are exactly two H's is $6/16 = 3/8$.

Now some terminology. . The **sample space**, often denoted by Ω, is the set of all possible outcomes. In our example the sample space has 16 elements:



A **probability space** is a sample space Ω, together with a **probability** $\mathbf{P}(\{\omega\})$ for each outcome ω, such that

- $0 \leq \mathbf{P}(\omega) \leq 1$ for all $\omega \in \Omega$.
- $\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$, i.e., the sum of the probabilities of all outcomes is 1.

The easiest way to assign probabilities to outcomes is to give all of them the same probability (as we saw earlier in the coin tossing example): if $|\Omega| = N$, then $\mathbf{P}(\omega) = \frac{1}{N}\ \forall \omega \in \Omega$. This is known as a uniform distribution. We will see examples of non-uniform probability assignments soon.

As we saw in the coin tossing example above, after performing an experiment we are often interested only in knowing whether a certain event occurred. Thus we considered the event that there were exactly two H's in the four tosses of the coin. Here are some more examples of events we might be interested in:

- The sum of the rolls of 2 dice is ≥ 10.
- The poker hand I dealt to you is a flush (i.e., all 5 cards have the same suit).

- In n coin tosses, at least $\frac{n}{3}$ of the tosses come up tails.

Let us now formalize the notion of an event. Formally, an **event** A is just a subset of the sample space, $A \subseteq \Omega$. As we saw above, the event "exactly 2 H's in four tosses of the coin" is the subset $\{HHTT, HTHT, HTTH, THHT, THTH, TTHH\} \subseteq \Omega$.

How should we define the probability of an event A? Naturally, we should just *add up* the probabilities of the outcomes in A.

For any event $A \subseteq \Omega$, we define the **probability of** A to be

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

Thus the probability of getting exactly two H's in four coin tosses can be calculated using this definition as follows. The event A consists of all sequences that have exactly two H's, there are 6 such sequences. There are $|\Omega| = 2^4 = 16$ possible outcomes for flipping four coins. Thus, each outcome $\omega \in A$ has probability $\frac{1}{16}$; and, as we saw above, there are six outcomes in A, giving us $\mathbf{P}(A) = 6 \cdot \frac{1}{16} = \frac{3}{8}$.

We will now look at examples of probability spaces and typical events that may occur in such experiments.

1. Flip a fair coin. Here $\Omega = \{H, T\}$, and $\mathbf{P}(H) = \mathbf{P}(T) = \frac{1}{2}$.

2. Flip a fair coin three times. Here $\Omega = \{(t_1, t_2, t_3) : t_i \in \{H, T\}\}$, where t_i gives the outcome of the ith toss. Thus Ω consists of $2^3 = 8$ points, each with equal probability $\frac{1}{8}$. More generally, if we flip the coin n times, we get a sample space of size 2^n (corresponding to all sequences of length n over the alphabet $\{H, T\}$), each outcome having probability $\frac{1}{2^n}$. We can look, for example, at the event A that all three coin tosses are the same. Then $A = \{HHH, TTT\}$, with each outcome having probability $\frac{1}{8}$. Thus, $\mathbf{P}(A) = \mathbf{P}(HHH) + \mathbf{P}(TTT) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

3. Flip a biased coin once. Suppose the bias is two-to-one in favor of Heads, i.e., it comes up Heads with probability $\frac{2}{3}$ and Tails with probability $\frac{1}{3}$. The sample space here is exactly the same as in the first example, but the probabilities are different: $\mathbf{P}(H) = \frac{2}{3}, \mathbf{P}(T) = \frac{1}{3}$. This is the first example in which the outcomes have non-uniform probabilities.

4. Flip the biased coin in the previous example three times. The sample space is exactly the same as in the second example, but it is not immediately obvious how to assign probabilities to the outcomes. This is because the bias of the coin only tells us how to assign probabilities to the outcome of *one* flip, not the outcome of *multiple* flips. It is, however, clear that the probabilities of the outcomes should not be uniform. We will return to this example after learning the important notion of *independence* in the next Lecture.

5. Roll two fair dice. Then $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$. Each of the 36 outcomes has equal probability, $\frac{1}{36}$. We can look at the event A that the sum of the dice is at least 10 (which is satisfied for the pairs $(5,5), (6,6), (6,4), (4,6), (5,6)$ and $(6,5)$), and the event B that there is at least one 6. In this example (and in 1 and 2 above), our probability space is **uniform**, i.e., all the outcomes have the *same* probability (which must be $\frac{1}{|\Omega|}$, where $|\Omega|$ denotes the size of Ω). In such circumstances, the probability of any event A is clearly just
$$\mathbf{P}(A) = \frac{\text{\# of outcomes in } A}{\text{\# of outcomes in } \Omega} = \frac{|A|}{|\Omega|}.$$
So for uniform spaces, computing probabilities reduces to *counting* outcomes! Using this observation, it is now easy to compute the probabilities of the two events A and B above: $\mathbf{P}(A) = \frac{6}{36} = \frac{1}{6}$, and $\mathbf{P}(B) = \frac{11}{36}$.