CmpE 343 Lecture Notes
2: Probability

Ethem Alpaydın

September 25, 2014

1 Basics

In a *random experiment*, the outcome is not known in advance, but it should be one of a number of outcomes. These possible outcomes together make up the *sample space* that we denote by S. For example, if we toss a coin, we do not know what the outcome will be for any toss, but we know that it will be either heads or tails. Each toss of the coin is an experiment. We can define compound experiments from simpler experiments; for example we can toss the same coin five times.

An *event* is a subset of the sample space. For example, when we toss the coin five times, one event may be having all of them heads, another event may be seeing three or more heads. Events are sets and we can talk about their intersection, union, and complement.

The probability of an event is represented by a number between 0 and 1 that quantifies how likely the event is. If E is an event, probability of E is denoted by $P(E)$. $P(\emptyset) = 0$ (the probability of an impossible event is 0), and $P(S) = 1$ (the probability of an event that is sure to happen is 1).

If we have two events E_1 and E_2 that are mutually exclusive ($E_1 \cap E_2 = \emptyset$), then the probability of either of them occurring is the sum of their probabilities: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$. This can be generalized to more than two events. For example, the probability of seeing three or more heads in five tosses is the sum of the the probability of seeing three heads in five tosses and the probability of seeing four heads in five tosses and the probability of seeing five heads in five tosses.

If E_1 and E_2 are not mutually exclusive, we have: $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

Probability is generally interpreted as the *limiting relative frequency*, that is, the proportion of the occurrence of the event when the experiment is repeated many times. For example, if a coin is fair, and is tossed many times, in approximately half of those experiments, we should see heads. Or if a fair die is tossed many times, in approximately one-sixth of those experiments, we should see two.

For an experiment where all outcomes in the sample space are *equally likely*, we have

$$P(E) = \frac{|E|}{|S|} \tag{1}$$

That is, the probability of E is the number of outcomes that correspond to E divided by the number of all possible outcomes.

Let us say that we toss a fair coin three times and E corresponds to the event where we see two heads and one tail. In this case, $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ where for example “HTT” denotes that the first toss is heads, the second toss is tails and the third toss is tails. This is a compound experiment whose sample space is calculated from the product of the sample spaces of the basic experiment: The first toss has two possible outcomes and the second toss is independent and has two outcomes and the third toss has two outcomes so there are $2^3 = 8$ outcomes for the three tosses.

If the coin is fair, all eight outcomes are equally likely and so the probability of each is $1/8$. We have $E = \{HHT, HTH, THH\}$—the outcome of tails can be seen in the third toss, the second toss, or the first toss. So we have $P(E) = 3/8$. Note that here the event is defined as “two heads and one tails” without specifying where the heads and tails should be. So we need to count all $\binom{3}{2} = 3$ possible ways of having two heads and one tails in three tosses. If we were asked the probability that the first two tosses are heads and the third tails (a specific order) then E would contain only “HHT” and the probability would be $1/8$.

Note that equation (1) is only valid if all the outcomes in the sample space are equally likely, where each have "weight" $1/|S|$. Otherwise, we have

$$P(E) = \frac{\sum_{e_i \in E} w(e_i)}{\sum_{s_j \in S} w(s_j)} \tag{2}$$

where e_i are the elements of E, s_j are the elements of S, and $w(e)$ is the "weight" of any element. For example, let us say that we have a coin where heads is twice as likely as tails. Then each heads has a weight of two and tails has a weight of one, and the probability of heads at any toss is $2/3$.

2 Conditional Probability

Sometimes after a random experiment two observations may be related and in such a case, knowing one may have an effect on the other. Let us say that in a bag, there are two red balls and three black balls, I draw a ball at random from the bag and let us say R_1 corresponds to the event that the ball is a red one. We have $P(R_1) = 2/5$. Then without returning the ball back to the bag I draw a second ball and R_2 corresponds to the event that this second ball is also red. Probability of R_2 depends on whether the first ball is red or not. If the first ball is red, of the remaining four balls there is only one red left; this we write as the *conditional probability* $P(R_2|R_1) = 1/4$. If the first ball is not red but black, there are two red balls left: $P(R_2|B_1) = 2/4$. If the first ball is thrown back into the bag before drawing the second ball, then whether it is red or black has no effect and we say that the draws are *independent*: $P(R_2|R_1) = P(R_2)$.

So if we draw *with replacement* from a bag or deck of cards and so on, the two events are independent; if the draw is *without replacement* there is dependence and we need to use conditional probabilities. If we are tossing a coin or rolling a die so on where the outcome of a previous event has no effect on the outcome of a later one, successive events are independent. Or if we are drawing from a (very large) bag containing two million red balls and three million black balls, whether we draw with or without replacement is negligible and independence can be assumed.

A *contingency table* contains the frequency of occurrences of event combinations in a large sample and is a concept widely used in statistics. For example, let us say in a course, we have male and female students from two departments whose numbers are given as below:

	Gender		
Department	Male	Female	Total
CmpE	60	20	80
IE	30	10	40
Total	90	30	120

We get probabilities by dividing the numbers with the total. Let us look at the different type of probabilities: We have the *joint probability*; for example, $P(\text{CmpE} \wedge \text{Male}) = 60/120$. The *conditional probability*; for example $P(\text{CmpE—Male}) = 60/90$—if we have Male given, the sample space reduces from the whole 120 students to the 90 male students, of which 60 are from CmpE. We also have the *marginal probability* (because they are the sums in the margins); $P(\text{CmpE}) = 80/120$ and $P(\text{Male}) = 90/120$.

We can define the conditional probability in terms of the joint and marginal probabilities:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \text{ if } P(B) > 0 \tag{3}$$

If two events are independent, $P(A|B) = P(A)$ and $P(A \wedge B) = P(A)P(B)$—the joint probability is the product of the marginal probabilities.

Let us go back to our example of the bag with two red and three black balls. If I draw two balls in succession without replacement, what is the probability that both are red ?

We are asked the joint probability

$$P(R_1 \wedge R_2) = P(R_1)P(R_2|R_1) = \frac{2}{5}\frac{1}{4} = \frac{1}{10}$$

Now let us talk about *marginalization*. Let us say we have the same setting and that we are asked the probability that the second ball is red, namely, $P(R_2)$. We are not specified the color of the first ball so we should sum (or *marginalize*) over the possible results of the first ball. Note that in the contingency table too, we have $P(Male) = P(\text{CmpE} \wedge \text{Male}) + P(\text{IE} \wedge \text{Male})$. Similarly here, we have

$$P(R_2) = P(R_1 \wedge R_2) + P(B_1 \wedge R_2) = P(R_1)P(R_2|R_1) + P(B_1)P(R_2|B_1) = \frac{2}{5}\frac{1}{4} + \frac{3}{5}\frac{2}{4} = \frac{8}{20} = \frac{2}{5}$$

In this case, the unknown event has two possible outcomes, red and black and hence we have summation over two joints. In the general case, we can have more, and even infinite, if we have a continuous outcome where we integrate instead of the discrete sum—as we will see in later lectures.

3 Bayes' Rule

From equation 3, we can write

$$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A) \tag{4}$$

and in turn

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5}$$

which is known as *Bayes' rule* after the English philosopher Thomas Bayes (1701–1761).

The nice thing about Bayes' rule is that if you have a conditional probability in one direction, you can calculate the conditional probability in the other direction.

For example, let us say that where I live the probability it rains on any day, $P(R)$, is 0.4. Let us say that when it rains the probability that my grass gets wet, $P(W|R)$, is 0.9; in 10 per cent of the time, it does not rain or strong enough for me to consider my grass watered enough. Let us say that the probability my grass gets wet when it does not rain (for example, due to my neighbor's sprinkler), $P(W| \sim R)$, is 0.2. Let us say that I leave for travel early in the morning and when I return in the evening I find my grass wet. What is the probability that it rained that day ?

$P(W|R)$ is the *causal direction* because rain is the cause of wet grass. Bayes' rule allows us to do a diagnostic in the opposite direction, that is, to go from the symptom back to the possible cause:

$$\begin{aligned} P(R|W) &= \frac{P(W|R)P(R)}{P(W)} \\ &= \frac{P(W|R)P(R)}{P(W|R)P(R) + P(W| \sim R)P(\sim R)} \\ &= \frac{0.9 \cdot 0.4}{0.9 \cdot 0.4 + 0.2 \cdot 0.6} = 0.75 \end{aligned}$$

Note how we use marginalization to calculate $P(W)$ in the denominator. Bayes' rule allows us to update our belief: In the absence of any other knowledge, $P(R)$ is 0.4, whereas after observation, the newly calculated $P(R|W)$ is much higher; knowing that the grass is wet greatly increased our belief in rain.