Lecture 7: Conditional Probability, Bayes' rule and Random Variables

Anup Rao

January 19, 2018

We discuss conditional probability. We learn about the famous Bayes' rule, and start talking about Random Variables.

Conditional Probability

SOMETHING VERY NICE HAPPENS TO PROBABILITY spaces when you *zoom in* to a particular event. For example, consider the two events A, B in the sample space Ω shown below. As usual, let p denote the distribution of the outcomes in Ω.

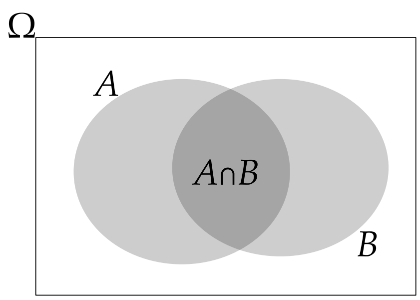

Figure 1: Two events and their intersection in a probability space.

Let us think about the event $A \cap B$. This event corresponds to a subset of the sample space Ω, but it is also a subset of A. In a sense, we could think of A itself as a sample space, and $A \cap B$ as an event in that sample space.

This view is particularly useful to modify our view of the probability space when some partial information has been revealed to us. If we have a probability space as above, and we know that the event A has happened, then the probability that B also happens, *given* that A has happened is

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

In particular, this definition gives:

$$p(B|A) \cdot p(A) = p(A \cap B) = p(A|B) \cdot p(B),$$

which implies that

$$p(B|A) = \frac{p(A|B) \cdot p(B)}{p(A)}.$$

This last equation is called *Bayes' rule*.

Example: Two Dice

Suppose you roll two dice at the same time. There are $6 \times 6 = 36$ possible outcomes of these rolls, and all are equally likely. What is the probability that the dice add up to 8?

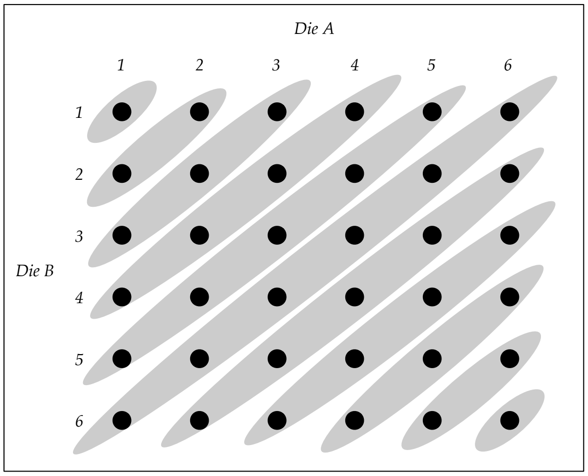

Figure 2: The sample space when two dice are rolled.

There are 11 possible values for the sum of the two dice, from $2 - 12$. Figure 4 shows the 11 corresponding events where the sum of the dice is fixed to a value. So, for example, the probability that the sum of the dice is 2 is only $1/36$, but the probability that the sum is 7 is the largest: $6/36 = 1/6$. If F denotes the event that the sum of the dice is 8, we have

$$p(F) = 5/36.$$

Now, let us consider a different kind of question. What is the probability that the sum of the dice is 8, given that the first die gives a value that is ≤ 4?

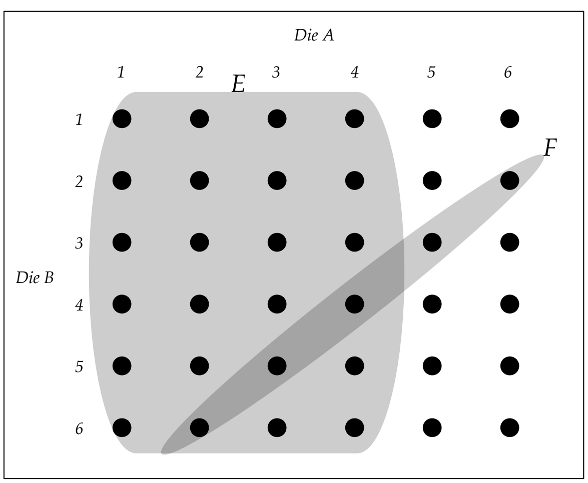

Figure 3: The events corresponding to the sum of the dice being 8 and the first die being ≤ 4 are shown.

If E denotes the event that the first die gives a value ≤ 4 and F denotes the event that the sum of the dice is 8, then we see that

$$p(E) = 4/6 = 2/3,$$

$$p(F) = 5/36,$$

and

$$p(E \cap F) = 3/36 = 1/12.$$

So,

$$p(F|E) = \frac{p(E \cap F)}{p(E)} = \frac{1/12}{4/6} = 1/8.$$

You can also calculate $p(F|E)$ directly from the picture: it corresponds to picking 3 out of 4×6 outcomes, so the probability is $3/24 = 1/8$. From the picture, we see that

$$p(E|F) = \frac{3}{5},$$

and we can verify Bayes' rule:

$$p(F|E) = \frac{1}{8} = \frac{(3/5) \cdot (5/36)}{2/3} = \frac{p(E|F) \cdot p(F)}{p(E)}.$$

If H is the event that the sum of the dice is 6, we see that $p(H|E) = 4/24 = 1/6$. So, even though the probablity of the sum being 6 is the same as the probability that the sum is 8, once we know that the first die roll is at most 4, the lower sum values get a boost: the probablity that the sum is 5 conditioned on E is larger than the probability that the sum is 8 conditioned on E.

The Conditional Probabilities give a Probability Space

It is easy to check that the probabilities $p(x|E) = \frac{p(x \cap E)}{p(E)}$ satisfy all the axioms that a probability space is supposed to satisfy. It is clear that $p(x|E) \geq 0$. Moreover, we have:

$$\sum_{x \in E} p(x|E) = \sum_{x \in E} \frac{p(x \cap E)}{p(E)} = \frac{p(E)}{p(E)} = 1.$$

So we can think of a new distribution $q(x)$ given by $q(x) = p(x|E)$. This distribution gives 0 weight to the points in the sample space outside of E, and is a valid distribution supported on Ω. It can also be viewed as a distribution supported on E, since it does not assign any weight to the points outside E.

Random Variables

It can be quite cumbersome to talk about events in a probability space, because there are so many of them. One piece of notation that really helps is the concept of a *random variable*. A random variable is just a function $X : \Omega \to S$ that maps the points in the sample space to some other set.

Random variables can be thought of as a partition of the entire sample space into disjoint events, namely those sets where the random variables is constant.

For example, in the case that we are throwing two dice, we can define the random variable X to be the value of the first die, and Y to be the value of the second die.

Then the event E is the same as the event that $X \leq 4$, and the event F is the same as the event that $X + Y = 8$.

The distribution p on the probability space induces a distribution $p(X)$ on the values in $[6]$ taken by the random variable X, a distribution $p(Y)$ on the values in $[6]$ taken by the random variable Y and a *joint distribution* $p(X, Y)$ on the values in $[6] \times [6]$ taken by both values. Often, if $p(X, Y)$ is the distribution of X and Y, then $p(X)$ is referred to as the *marginal distribution* on X.

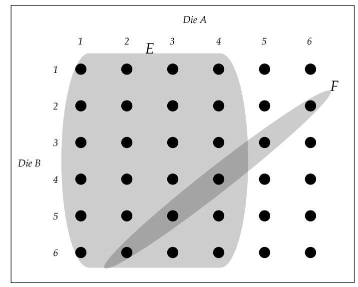

Figure 4: The events corresponding to the sum of the dice being 8 and the first die being ≤ 4 are shown.