3 PROBABILITY

3.11 More conditional probability, decision trees and Bayes' law

In this video I will explain the relation between conditional probability, decision trees and an equation that relates different conditional probabilities, Bayes' law.

Let's look at conditional probability as a way by which a sample space is reduced. You may decide that if event B occurs before event A, the sample space is effectively reduced to a new, smaller, sample space.

When calculating conditional probabilities, you are considering the probabilities for event A within a reduced sample space rather than the complete sample space. The important point is that all the properties that you know to be applicable for events in the complete sample space apply to this reduced sample space as well. For example, if A in the reduced sample space, that is A given B, could take 3 values, the union of these three probabilities should be 1.

Let's apply this to a concrete example. You have made a count of different activities by the people on your beach, which you distinguished by gender. And you have turned it into this table with probabilities. The conditional probabilities for the three activities, given that a person were male would be given by dividing the respective joint probabilities with 0.45; and for females, the conditional probabilities for the activities would be given by dividing with 0.55. The resulting conditional probabilities for the activities, given gender, are shown here. And the same can be done for the conditional probabilities of gender, given activity. You can see that the rule, stating that the sum of all conditional probabilities should be one, does apply.

Sometimes you deal with conditional probabilities without really noticing, you may have done so already while making calculations with a tree diagram. Let's turn the table with joint and marginal probabilities into a tree diagram. You can imagine that at the beach you would first look at all the people resting and subsequently count the number of male and female persons in that group, and the same for the other two activities. So your tree would have a first split over activity and a second split over gender, within each 'activity branch'.

Now you'd find these probabilities along each branch. The marginal probabilities for the activities would be placed at the first node. In the second step, the split is by gender. Here you would deal with conditional probabilities: gender, given that an activity is known. You would calculate these conditional probabilities by dividing the counts for male and female per activity by the total people in that activity. For example, you would divide the number of male persons that are resting by the total number of persons resting.

The joint probabilities at the end would result from multiplying the probability per activity with the conditional probabilities of gender given activity.
So in this tree diagram the probabilities that you encounter at the first node are marginal probabilities, for the activities in this case. The probabilities at the second nodes are conditional probabilities for gender, given the outcome of a certain activity. And the probabilities at the end of a tree are the joint probabilities. To get the marginal probabilities for gender, you would have to add the respective joint probabilities.

If you would interchange gender and activity in the tree diagram, the marginal and conditional probabilities would change, but the joint probabilities would remain the same.
This leads to an interesting insight:

The joint probability of A and B equals the multiplication of conditional probability of event A given B times probability of B, but it also equals the multiplication of conditional probability of event B given A times probability of A. Therefore, we can just as well say that the conditional probability of A given B is related to B given A. This equation is known as Bayes' law, it is in fact the shorthand version, because often the marginal probability of B is not known directly but has to be computed from the sum of the different conditional probabilities for B, given an outcome for A.

If we return to our example once more: Bayes' law allows to calculate the probability that for example a person would be swimming, when you know the person is female; while we have only the reversed conditional probability to our availability: a person is female given that this person is swimming. The conditional probability that someone is female given the person is swimming is multiplied with the probability of swimming and this is divided by the sum of the probabilities for the three activities times their conditional probabilities.

In simple discrete probability calculations, Bayes' law is a convenient function that follows from the axioms of probability calculus. However, it is often interpreted in a more abstract way, where the left hand side of the equation is understood as the degree of believe in a hypothesis A after observing B; and the right hand side represents the believe in A, prior to knowing B, times the support that B provides for A.

And in this wider context, the probability of A is called prior probability, because it is your knowledge about A prior to observing B. And the conditional probability of A, given B is called posterior probability.

There are no simple frequency counts to inform you about the belief in A, the prior probability, so it needs to be assessed qualitatively. As you can imagine, considering probability as a belief rather than a frequency based quantity is quite a different viewpoint. As a consequence, it has led to a many philosophical debates about the meaning of probability and the utility of Bayesian methods. But in practical situations the different perspectives appear not to be far apart: in any realistic analysis there is prior knowledge involved, as well as assumptions and subjective choices; while wherever possible one would use observations and frequency-based methods to estimate probabilities.

Let me summarize what I have explained in this video.
- The Probability of A conditional on B can be considered as the probability of A in the reduced sample space where B occurred, to which all rules of a sample space apply.
- A tree diagram contains different probabilities. At the first node it has marginal probabilities and for any node further on, it has conditional probabilities, the joint probabilities at the end result from multiplying marginal and conditional probabilities along a branch.
- The fact that joint probability can be calculated based on the conditional probability of A given B as well as B given A, leads to Bayes' law which relates these two conditional probabilities
- Bayes' law is also used express how a prior belief in a hypothesis A can be updated by information B; in that context the probability of A is called prior probability and the conditional probability of A given B is called posterior probability.
- Considering probability as a belief rather than a frequency based entity, is a different theoretical starting point. However, in practice elements of the two are often combined.