4 PROBABILITY DISTRIBUTIONS

4.02 Cumulative probability distributions

After you have come to terms with the basic probability rules and got acquainted with the probability distribution, it is time to meet its alter-ego: the cumulative probability distribution. In this video I will explain how the two distributions relate to each other and especially how you can make probability statements or find important values of a random variable by using the cumulative probability distribution.

Let's quickly jump into it. Consider this simple discrete distribution. Can you answer the question: what the probability would be that X takes a value of either two or three? [...]

The answer is obtained by adding up the probabilities that X is two and X is three, a union of the probabilities, so that is 0.7. All the probabilities listed in the table (or along the x-axis of the probability mass function) are disjoint, so any union of probabilities is simply the sum of these probability values. Similarly, the probability that X is greater than one is equal to one minus the probability of X is one, which is 0.9, by the complement rule.

Now let's step-up to a next idea. Based on a probability distribution we can easily calculate probabilities for values that are less than or equal to a given value. For example, the probability that X is less than or equal to one is 0.1, the probability that X is less than or equal to two is 0.1 plus 0.3, which is 0.4, etcetera.

The resulting probabilities are called cumulative probabilities, and the list of all cumulative probabilities is called a cumulative probability distribution or cumulative distribution function. The probability histogram for this cumulative probability distribution can be made as well and looks as follows.

Also probability density functions – the probability distributions for continuous random variables – have a corresponding cumulative distribution. Consider for example this probability distribution - the corresponding cumulative distribution is given here. …
An interesting aspect of this step is that the y-variable changes from a probability density to a probability, because in the cumulative distribution it is the area from the smallest value of x up to the value of interest in the probability density function that is put on the y-axis.

As you see, cumulative probability distributions have continuously increasing values, starting at zero and incrementing to a maximum of one for the largest value that the random variable can take. The cumulative distribution, especially its graphical form, is very convenient because it can answer two questions. You can select a certain value of the random variable at the x-axis and then find which fraction of the observations will be lower than or equal to this value at the y-axis. Or, reversely, you can select the fraction at the y-axis and then find the corresponding threshold-value at the x-axis. There is in fact a shorter way of saying that a fraction of the values fall below a threshold-value, using the term quantile. For example, a threshold-value below which 0.1 of the values are found, is called the 0.1 quantile. So the cumulative probability distribution is in fact showing the quantiles for a random variable. You find for example for the cumulative-probability of 0.5 the median value of the random variable, and for 0.25 the lower quartile. It is noteworthy that for symmetric probability distributions the median coincides with the mean, so for symmetric distributions also the mean is found at a cumulative-probability of 0.5.

Let me summarize what I hope you understood from this video:

- A cumulative probability of a random variable is the probability of obtaining a value lower than or equal to a threshold-value.
- Considered in the other direction, a cumulative probability specifies a quantile of the random variable, for example at the cumulative probability of oh point five the median of the random variable is found.
- Just like a probability distribution, the cumulative probability distribution can exist in the form of a table, a graph or an equation.
- It can be obtained by calculating a cumulative sum of the probabilities from the smallest up to the largest value of the random variable.
- And it is continuously increasing with an increasing value for the random variable, starting at zero and incrementing to a probability of one.