3 PROBABILITY

3.08 Joint and marginal probabilities

Counts of interesting phenomena from everyday life are often turned into proportions and interpreted as probabilities. By exploiting the power of probability calculus the relations among these phenomena can be understood better or used to make predictions. Joint and marginal probabilities are two important types of probabilities that are encountered in this context. In this video I will explain the meaning of joint and marginal probabilities and show their properties.

Let's start by considering an example. You are at a beach and observe your fellow beach-visitors. You notice three different activities among them, which are mutually exclusive: people resting – they are all sitting or lying on the beach. People playing, these are running around, building sand castles or standing in the water. And finally people swimming. Furthermore, you distinguish by gender. So each person you observe is a case in your data set, while gender and activity are variables. You end up with the following contingency table.

In total, you counted hundred and thirteen people, seventy-nine of whom are resting. Twenty are playing and only fourteen of them are swimming. There are a few more female than male visitors (62 versus 51). These numbers are in fact the row and column totals for each variable separately. These row and column totals are located in the margin of this table … and therefore called marginal values. Please note that this has no bearing with the colloquial meaning of marginal which may be 'unimportant'. In a table like this the marginal values represent the counts for a single variable without regarding any other variable - for example, number of people resting without regarding gender.

Now we turn this frequency table into a table with proportions by dividing the number in every cell by the total number, hundred and thirteen. In this table the central block contains six proportions which together sum to one. Also, the proportions in each column add-up to the values in the marginal column at the right, and the proportions in each row add-up to the values in the marginal row at the bottom. And as a consequence, the marginal column values add-up to one as do the marginal row values.

Your count can be considered as a random sample, measuring the distribution of activity and gender of people at this beach, and you can then think of the proportions as probabilities. In the central block, you have the intersection of activity with gender, for example the probability that a given person is male and swimming. These values are called joint probabilities.

Joint probability is just a shorter name for the probability of the intersection of various events. Each joint probability in our table is a disjoint event from any other joint probability, as each person at the beach is only put in one of the six joint categories, and not counted double. At the same time, the joint probabilities form a jointly exhaustive set of events because there is no other outcome for activity and gender possible in this case. Therefore the joint probabilities sum to one.

In the margins you have, as you might expect, marginal probabilities. These give the probabilities when considering only one variable, for example the probability that a given person is male, regardless of his activities; or that any person would be playing, regardless of gender.
The marginal probabilities result from the union of the joint probabilities – for example the probabilities of resting, playing and swimming per gender. So here the addition rule applies: probabilities are summed.

Consequently, if you'd not have the original counts to your availability but only joint probabilities, you could always calculate marginal probabilities by summation. Conversely, if you were given the marginal probabilities, you would not in every case be able to calculate the joint probabilities.

Let me summarize what I have explained in this video.

- When you have counts on phenomena that originate from a random sample or trial you can turn these into probabilities.
- If you have observations on multiple random variables, you can calculate joint and marginal probabilities for these variables
- Joint probabilities are the probabilities for the intersection of certain outcomes of the variables
- Marginal probabilities are the probabilities for an outcome of each individual variable
- In the special case where you have two discrete random variables, a good way to organize this data is in a contingency-table, where you would have joint probabilities in the center, and marginal probabilities at the boundaries.
- All the joint probabilities sum to 1, and they also add up to the marginal probabilities in both directions.
- You can always reconstruct the marginal probabilities from joint probabilities by summation
- But reconstructing joint probabilities on the basis of marginal probabilities cannot be done without making additional assumptions.