# BASICS OF PROBABILITY

## RANDOM VARIABLES

Random variable **X** converts the outcomes of experiments to measurable values.

A random variable is a function that maps the outcomes of a random experiment to real numbers.

It assigns a numerical value to each outcome in the sample space of the experiment, allowing us to quantify and analyze uncertainty.

**For example**, let’s say as a data analyst at a bank, you are trying to find out which of the customers will default on their loan, i.e., stop paying their loans. Based on some data, you have been able to make the following predictions:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Customer Number | Yearly Income | Amount of Loan Due | Number of Dependants | Default Prediction (Yes/No) |
| 1 | 10 lakhs | 75 lakhs | 3 | Yes |
| 2 | 15 lakhs | 50 lakhs | 2 | No |
| 3 | 20 lakhs | 40 lakhs | 1 | No |

Now, instead of processing the yes/no response, it will be much easier if you define a random variable **X** to indicate whether the customer is predicted to default or not. The values will be assigned according to the following rule:

X = 1, if the customer defaults.

X = 0, if the customer does not default.

Now, the data changes to the following:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer Number | Yearly Income | Amount of Loan Due | Number of Dependants | Default Prediction (Yes/No) | X (Random Variable) |
| 1 | 10 lakhs | 75 lakhs | 3 | Yes | 1 |
| 2 | 15 lakhs | 50 lakhs | 2 | No | 0 |
| 3 | 20 lakhs | 40 lakhs | 1 | No | 0 |

Now, in this form, the table is entirely quantified, i.e., converted to numbers. Now that the data is entirely in quantitative terms, it becomes possible to perform a number of different kinds of statistical analyses on it

## PROBABILITY DISTRIBUTIONS

A probability distribution is a form of representation that tells the probability for all the possible values of **X.**