

Categorical Distributions

The Categorical Distribution is a fancy name for random variables which takes on values *other than numbers*. As an example, imagine a random variable for the weather today. A natural representation for the weather is one of a few categories: {sunny, cloudy, rainy, snowy}. Unlike in past examples, these values are not integers or real valued numbers! Are we allowed to continue? Sure! We can represent this random variable as X where X is a categorical random variable.

There isn't much that you need to know about Categorical distributions. They work the way you might expect. To provide the Probability Mass Function (PMF) for a categorical random variable, you just need to provide the probability of each category. For example, if X is the weather today, then the PMF should associate all the values that X could take on, with the probability that X takes on those values. Here is an example PMF for the weather Categorical:

Weather Value	Probability
Sunny	$P(X = \text{Sunny}) = 0.49$
Cloudy	$P(X = \text{Cloudy}) = 0.30$
Rainy	$P(X = \text{Rainy}) = 0.20$
Snowy	$P(X = \text{Snowy}) = 0.01$

Notice that the probabilities must sum to 1.0. This is because (in this version) the weather must be one of the four categories. Since the values are not numeric, this random variable will **not** have an expectation (values are not numbers) variance nor a PMF expressed as a function, as opposed to a table.

Note to your future self: A categorical distribution is a simplified version of a [multinomial distribution](#) (where the number of outcomes is 1)