# Probability

### RANDOMNESS AND PROBABILITY

The word probabilityis often used in everyday language, but it can have quite different meanings. Sometimes people use it to describe how confident they are, as in "*It's probably going to rain today*". Other times it is used to describe the anticipated frequency of something, as in "*The probability of winning the lottery is 1 in a million*". Gratefully, the concept of probabilitythat is used in the type of statistics we are learning is unambiguous and comparatively straightforward. Probabilityis simply the frequency of a particular outcome, or event, if we repeated a random trialover and over and over…

The first step to defining probability is to establish the context. Imagine that we want to know the probability that a newly designed electric scooter would break down. One way to measure the probability is to have people drive the scooters in everyday conditions and then measure the failure rates for every kilometre driven. The act of driving the scooters waiting for failures is the context, which is called a random trial.

More formally, a random trial is any process that has multiple outcomes but where the result of any particular trial is unknown. For example, imagine you flipped a coin in the air, caught it, and put it down on a table without looking at what side faced up. There are two possible outcomes, heads or tails, but you won't know which faces up until you look. The action of flipping a coin is the random trial.

The list, or set, of all possible outcomes is called the sample space. The list is typically shown within curly braces '{...}'. For our example of flipping a coin, the sample space is {heads, tails}.

The final component that we need to define is an event, which is the outcome that you are interested in. The event can be a single element in the sample space. For example, when flipping a single coin we might be interested in the probability of getting a tail, in which case the event (E) would be E={tail}. However, an event can also be any subset of the sample space. For example, consider rolling a single six-sided die. The sample space (S) is S={1, 2, 3, 4, 5, 6}. If you were interested in "the probability of seeing a one", the event is E={1}. But if you were inserted in "the probability of seeing an odd number", the event would be either a 1, 3, or a 5, which is written as E={1,3,5}.

Random trials can have either discrete or continuous variables. The above examples of a random trial considered random variables that had discrete outcomes, either you flip a head or tail of a coin. Likewise, random trials can also be for continuous random variables. For example, commuters on the Toronto TTC subway system can arrive at a station at varying times. If we consider the wait time of a person entering the subway station as a random trial, then it would be a random trial for a continuous outcome because the length of time a person waits is a continuous numerical variable.

The connection between randomness and sampling is the act of selecting a sampling unit and taking a measurement. Recall that measurements are done on an observation unit, and the characteristic being measured is referred to as a measurement variable. It is called a *measurement variable* because the value of any particular measurement is unknown prior to making the observation. As such, the act of measuring an observation on a sampling unit is considered a random trial.

For the purpose of our statistical work, the probabilityof an eventis the proportion of times that the event would occur if a random trialwas repeated many many (many) times.

The key part of the above definition is that the random trial must be repeated many times to estimate probability. How many times? That depends on the nature of the random trial, and how accurate you want the probability to be, but we are talking about tens of thousands of trials. This concept is a mathematical theorem known as the "law of large numbers".

## 

## PROBABILITY DISTRIBUTIONS

A random trial has multiple outcomes, and the probabilities of the different outcomes are often related to each other through the nature of the random trial itself.

The simplest example is flipping a coin. The nature of the random trial is that each side of the coin has an equal chance of landing face-up. If we label one side 'heads' and one side 'tails', then the event 'heads' and the event 'tails' has the same probability because of the nature of the underlying random trial.

Probability distributionsutilize the structure of a random trial to provide a compact way of describing the probability of all events, and a tool for calculating probabilities over a range of events.

The probability of an event is its long-term frequency if the random trial was repeated many times. While looking at the probability of a single event can be useful in some situations, such as when playing a lottery, it is more common to look at probability over a range of events. For example, imagine you are looking at customer satisfaction for patrons of a coffee shop and decide to measure the length of time people wait in line as one criterion. Measuring the length of time a randomly selected person waits in line to buy coffee would be the random trail. It is unlikely that you would not be interested in the probability that someone takes exactly 5 minutes, which is a single event. Rather, it is more likely you would be interested in the probability that someone waits 5 minutes *or longer*, which is a range of events.

Probability distributionsare functions that describe probability over a range of events. Specifically, the probability of observing an outcome within a range of events is the area under the probability distribution function.

Probability distributionshave the following properties:

1. They describe the probability for the entire sample space.
2. The area under the probability distributionalways sums to one.
3. Are used to describe both continuous and discrete random variables.

The first property means that a probability distributioncontains all the necessary information about a random trial because it characterizes all possible outcomes. The second property is basically a restatement of the first in that all possible outcomes are contained within the probability distribution. The final property means that much of what we learn about probability distributionscan be applied to both continuous and discrete numerical variables.

Discrete distributionsare probability distributionsfor discrete random variables. For example, the number of times children ask for ice cream on a hot day is a discrete random variable that would be characterized using a discrete distribution.

A discrete distributionis typically shown as a series of vertical bars with no space between them. Each event gets a separate bar, and the area of that bar represents the probability of the event. The vertical axis of discrete distributionsis called the probability mass.

A continuous distributionis typically shown as a single curve as a function of the continuous event. The area under the curve for a given event range is the probability of observing an outcome in that range. The vertical axis of continuous distributionsis called the probability density.

Probability distributions provide a map between a range of events for a random trial and their probabilities. Importantly, that map goes in both directions. In the first direction, probability distributions are used to calculate the probability for a given range of events. In the second direction, probability distributions are used to calculate a range of events that corresponds to a given probability. Both directions are used in the reporting and interpretation of statistical results.

The lessons introducing probability distributions looked at the first type of question, which is how to calculate the probability of an outcome given an event range. For example, the question "what is the probability that there will be more than a centimetre of snow on any given day during winter" is answered by calculating the probability of observing more than one centimetre of snow from a probability distribution of snowfall.

The second type of question we can answer uses probability distributions to find the event range that is consistent with a given probability. For example, the question "what daily minimum depth of snow do we expect 50% of the time during winter?" is answered by calculating the range of snow depths that correspond to a probability of *p*=0.50 from a probability distribution of snowfall.

## Standard Normal distribution

The standard Normal distributionhas a long history in statistics. It is a Normal distribution, but with some special properties. Specifically:

* The mean of the standard Normal distribution
* is zero (*μ*=0)
* The standard deviation of the standard Normal distributionis one (*σ*=1)
* The x-axis is called the z-score, which is a scale that measures the number of standard deviations from the mean

## Converting to a standard Normal distribution

Any problem that is based on a Normal distribution can be answered by converting to a standard Normal distribution. The conversion to the standard Normal distributionis done as follows

1. Subtract the mean
2. Divide by the standard deviation