

4.1: The Distribution Function as a Summary of Experimental Results

In [Section 2.10](#), we derive Boyle's law from Newton's laws using the assumption that all gas molecules move at the same speed at a given temperature. This is a poor assumption. Individual gas molecules actually have a wide range of velocities. In Chapter 4, we derive the Maxwell–Boltzmann distribution law for the distribution of molecular velocities. This law gives the fraction of gas molecules having velocities in any range of velocities. Before developing the Maxwell–Boltzmann distribution law, we need to develop some ideas about distribution functions. Most of these ideas are mathematical. We discuss them in a non-rigorous way, focusing on understanding what they mean rather than on proving them.

The overriding idea is that we have a real-world source of data. We call this source of data the **distribution**. We can collect data from this source to whatever extent we please. The datum that we collect is called the distribution's **random variable**. We call each possible value of the random variable an **outcome**. The process of gathering a set of particular values of the random variable from a distribution is often called **sampling** or **drawing a sample**. The set of values that is collected is called **the sample**. The set of values that comprise the sample is often called “the data.” In scientific applications, the random variable is usually a number that results from making a measurement on a physical system. Calling this process “drawing a sample” can be inappropriate. Often we call the process of getting a value for the random variable “doing an experiment”, “doing a test”, or “making a trial”.

As we collect increasing amounts of data, the accumulation quickly becomes unwieldy unless we can reduce it to a mathematical model. We call the mathematical model we develop a **distribution function**, because it is a function that expresses what we are able to learn about the data source—the distribution. A distribution function is an equation that summarizes the results of many measurements; it is a mathematical model for a real-world source of data. Specifically, it models the **frequency** of an event with which we obtain a particular outcome. We usually believe that we can make our mathematical model behave as much like the real-world data source as we want if we use enough experimental data in developing it.

Often we talk about **statistics**. By a statistic, we mean any mathematical entity that we can calculate from data. Broadly speaking a distribution function is a statistic, because it is obtained by fitting a mathematical function to data that we collect. Two other statistics are often used to characterize experimental data: the **mean** and the **variance**. The mean and variance are defined for any distribution. We want to see how to estimate the mean and variance from a set of experimental data collected from a particular distribution.

We distinguish between discrete and continuous distributions. A **discrete distribution** is a real-world source of data that can produce only particular data values. A coin toss is a good example. It can produce only two outcomes—heads or tails. A **continuous distribution** is a real-world source of data that can produce data values in a continuous range. The speed of an automobile is a good example. An automobile can have any speed within a rather wide range of speeds. For this distribution, the random variable is automobile speed. Of course we can generate a discrete distribution by aggregating the results of sampling a continuous distribution; if we lump all automobile speeds between 20 mph and 30 mph together, we lose the detailed information about the speed of each automobile and retain only the total number of automobiles with speeds in this interval.

This page titled [4.1: The Distribution Function as a Summary of Experimental Results](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Paul Ellgen](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [3.1: The Distribution Function as a Summary of Experimental Results](#) by [Paul Ellgen](#) is licensed [CC BY-SA 4.0](#). Original source: <https://www.amazon.com/Thermodynamics-Chemical-Equilibrium-Paul-Ellgen/dp/1492114278>.