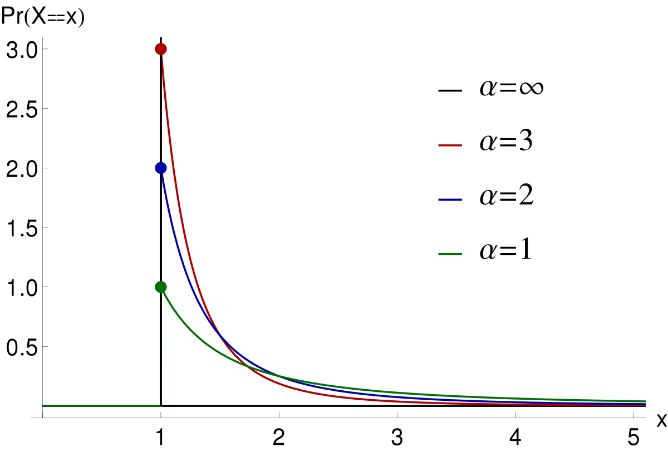
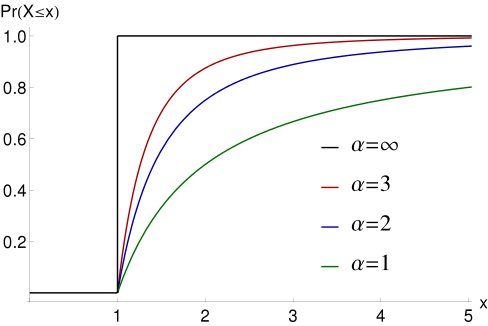
At this point, we have become familiar with multiple distributions such as the normal or right skewed and left skewed. Their histograms can tell us means, modes, and variances, probability mass functions can (PMF) can tell us probabilities of each value occurring, and cumulative density functions (CDF) can tell us percentile ranks. While we can obtain the same summary statistics from all distribution shapes, there are some distributions that have special use cases, such as the Weibull or Pareto distributions. I will be focusing primarily on the Pareto distribution here, explaining its difference from other distributions as well some of its use cases.

The Pareto distribution, named after Italian economist Vilfredo Pareto, is a heavily skewed, power-law probability distribution that resembles an exponential distribution. Unlike other distributions, it illustrates a case of continuous variables where lower numbers have a much higher probability of occurrence than higher numbers, and as such the mode is always the minimum value of the distribution. Pareto had initially used his distribution to describe the “80-20 rule” of wealth inequality (80% of the wealth is owned by 20% of the people). I will refer to it as the “major-minor rule” as trends could be 20-70, 30-90, etc.

Below is an example of three Pareto probability distributions. The x-axis shows our desired data (population in Pareto’s plot) and the y-axis shows the probability as percents (percent wealth in Pareto’s plot). Its probability density function is as follows: (a \* x\_m^a) / x^(a+1), where x is a data point, x\_m is the minimum value of the distribution, and *a* is a parameter describing the shape of the distribution. As *a* increases, so does the percentage of lower numbers in the distribution, and as x gets farther from x\_m, the function approaches zero. In Pareto’s wealth example, increasing *a* means that less people control more of the wealth.



The CDF of the distribution is as follows: CDF = 1 – (x/x\_m) ^ -a, where x, x\_m, and *a* are the same as they were for the PDF. Below is an example of the CDF for different values of *a.* As *a* increases, the CDF increases much more for a given change in x. 

Pareto distributions can be used for more than just visualizing wealth inequality. They can also be used for improving product quality. Whenever problems in products are reported, the producer can usually trace the problems back to a single defect, in accordance with the major-minor rule. A minority of defects can cause a majority of problems in almost all engineering disciplines and programming; this can be modeled by a Pareto distribution with the appropriate *a* and x\_m values. The inverse, that a minority of quality products can be responsible for a majority of revenue, is also true.

Whenever there is a case where a minority is responsible for, or contains, a majority, a Pareto distribution is appropriate to model it. In a hypothetical scenario, let’s say that 10% of customers are responsible for 70% for customer service calls. A Pareto distribution of customers can help determine this 10% so resources can be allocated more towards assisting them. In addition, this distribution can be used to model population density (few large cities and several small ones) and forest fire radius (few large fires and several small ones).

References:

CFI. (2020). *What is a Pareto Distribution?* Retrieved September 27, 2020 from https://corporatefinanceinstitute.com/resources/knowledge/economics/pareto-distribution/#:~:text=The%20Pareto%20distribution%20is%20used,geophysical%20phenomena%20in%20a%20society.&text=Pareto%20observed%20that%2080%25%20of,many%20disciplines%20such%20as%20incomes.

Downey, A. B. (2020). *Think Stats.* O’Reilly Media.

Tam, T. (2007). *AP Statistics Curriculum 2007 Pareto.* UCLA SOCR. Retrieved September 27, 2020 from http://wiki.stat.ucla.edu/socr/index.php/AP\_Statistics\_Curriculum\_2007\_Pareto#:~:text=Definition%3A%20Pareto%20distribution%20is%20a,people%20have%20very%20high%20incomes.