QQ PLOT

Yunsi Wang, Tyler Steele, Eva Zhang

Spring 2016

QQ PLOT INTERPRETATION:

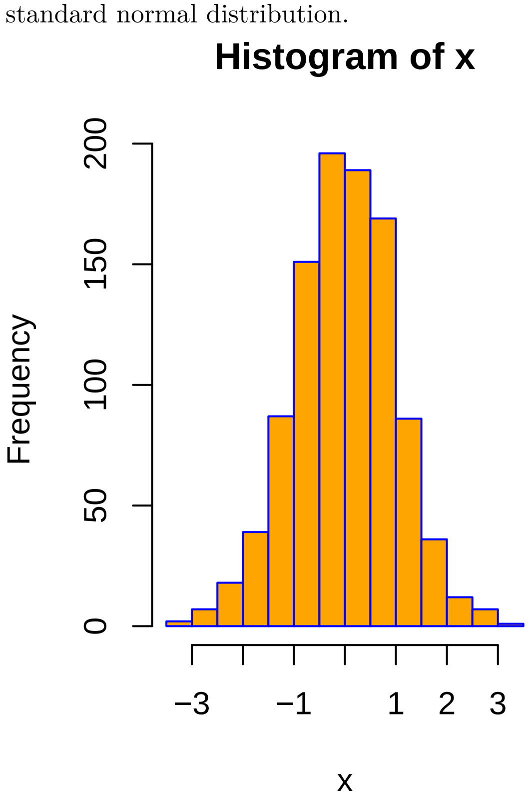
Quantiles:

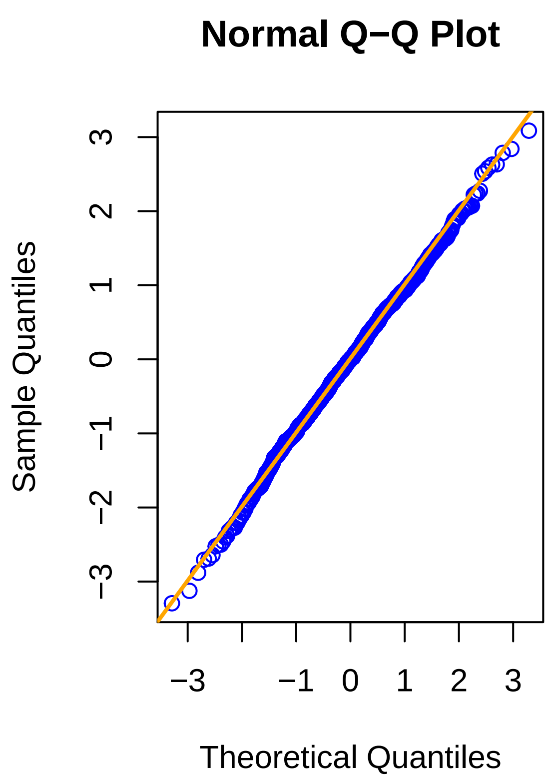
The quantiles are values dividing a probability distribution into equal intervals, with every interval having the same fraction of the total population.

QQ-plot:

The purpose of the quantile-quantile (QQ) plot is to show if two data sets come from the same distribution. Plotting the first data set's quantiles along the x-axis and plotting the second data set's quantiles along the y-axis is how the plot is constructed. In practice, many data sets are compared to the normal distribution. The normal distribution is the base distribution and its quantiles are plotted along the x-axis as the "Theoretical Quantiles" while the sample quantiles are plotted along the y-axis as the "Sample Quantiles". A few examples are presented below.

The first sample is obtained by simulating a standard normal distribution with sample size 1000. We will test this sample against the standard normal distribution to see if the quantiles match. From the histogram we can see that this sample is bell shaped around zero. This leads us to reason that it most likely comes from a standard normal distribution.

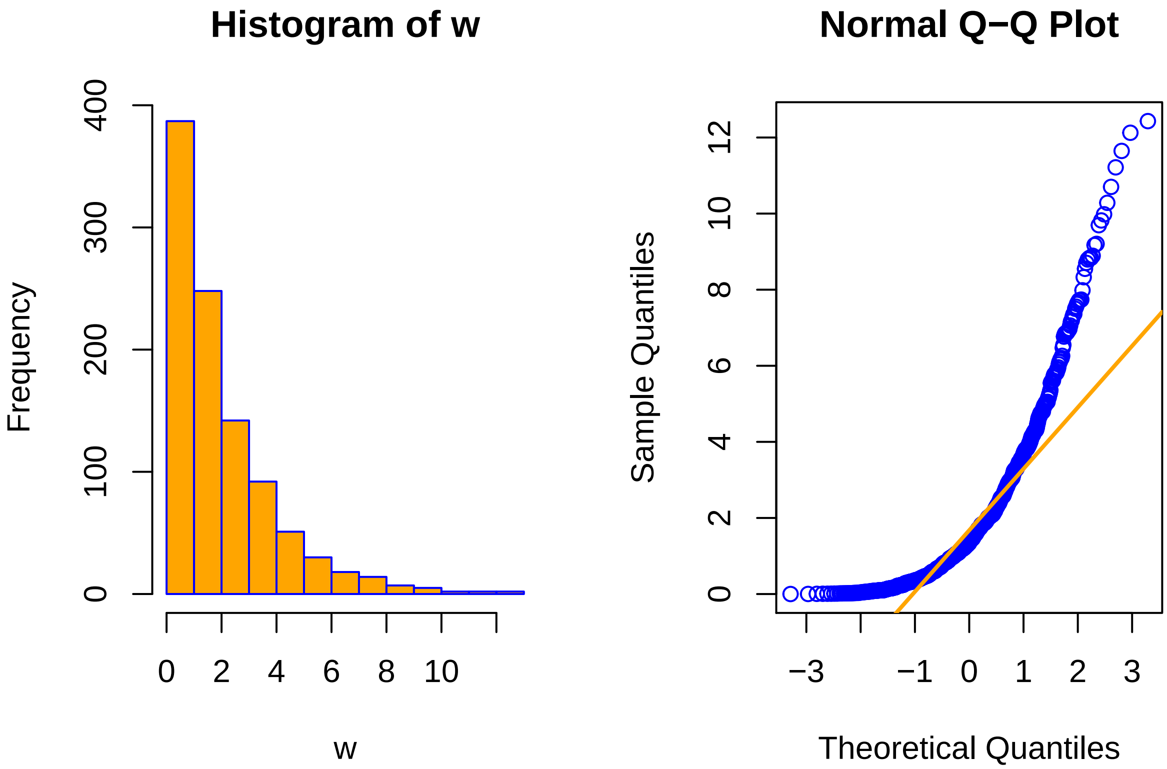




When looking at the QQ plot, we see the points match up along a straight line which shows that the quantiles match. While the line plotted is not a necessary component of the QQ plot, it allows the reader to visualize where the points should line up should the sample match the base distribution.

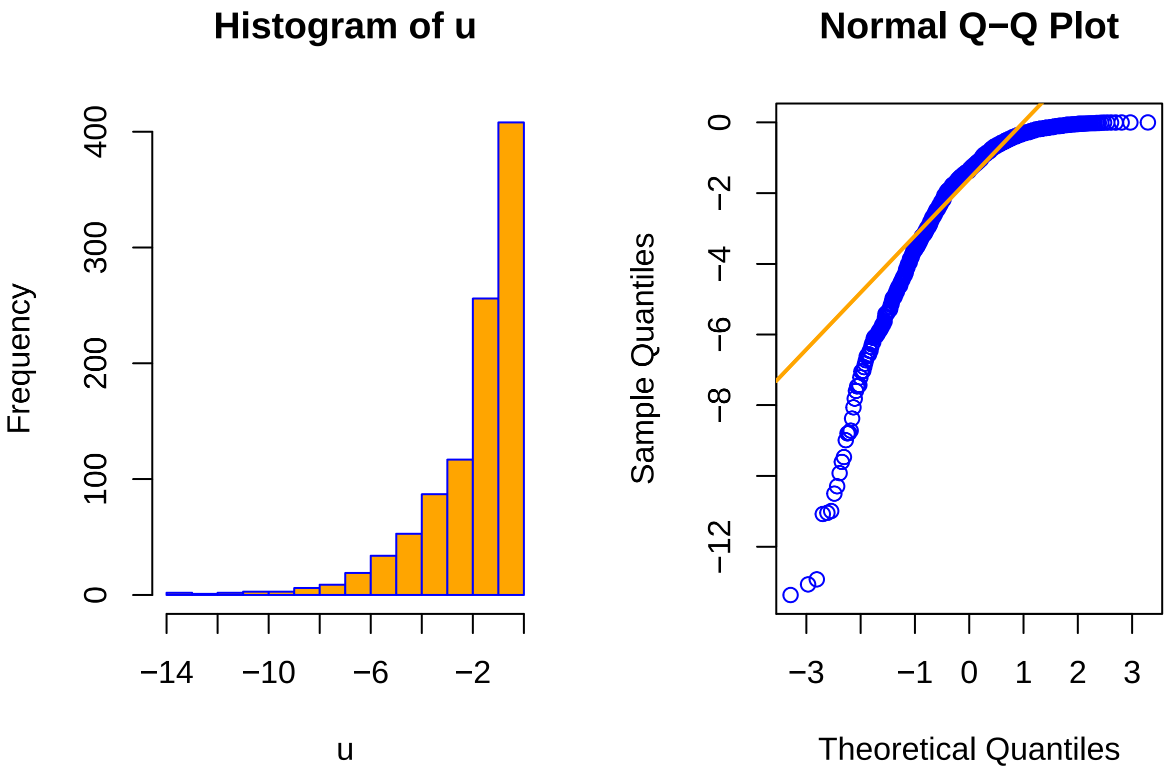
The next examples will show what various QQ plots look like if two data sets do not come from the same distribution.

This next example shows a right skewed sample compared to the standard normal distribution. The sample is obtained by simulating a chi-squared distribution with 2 degrees of freedom and a sample size of 1000. From the histogram we can see that the distribution is right skewed since it contains many observations around zero but then rapidly declines in the frequency of values as “w” increases. The QQ plot shows this sample’s quantiles compared to the standard normal. Intuitively, it makes sense that the points should not align along a line since the data sets are not from the same distribution.



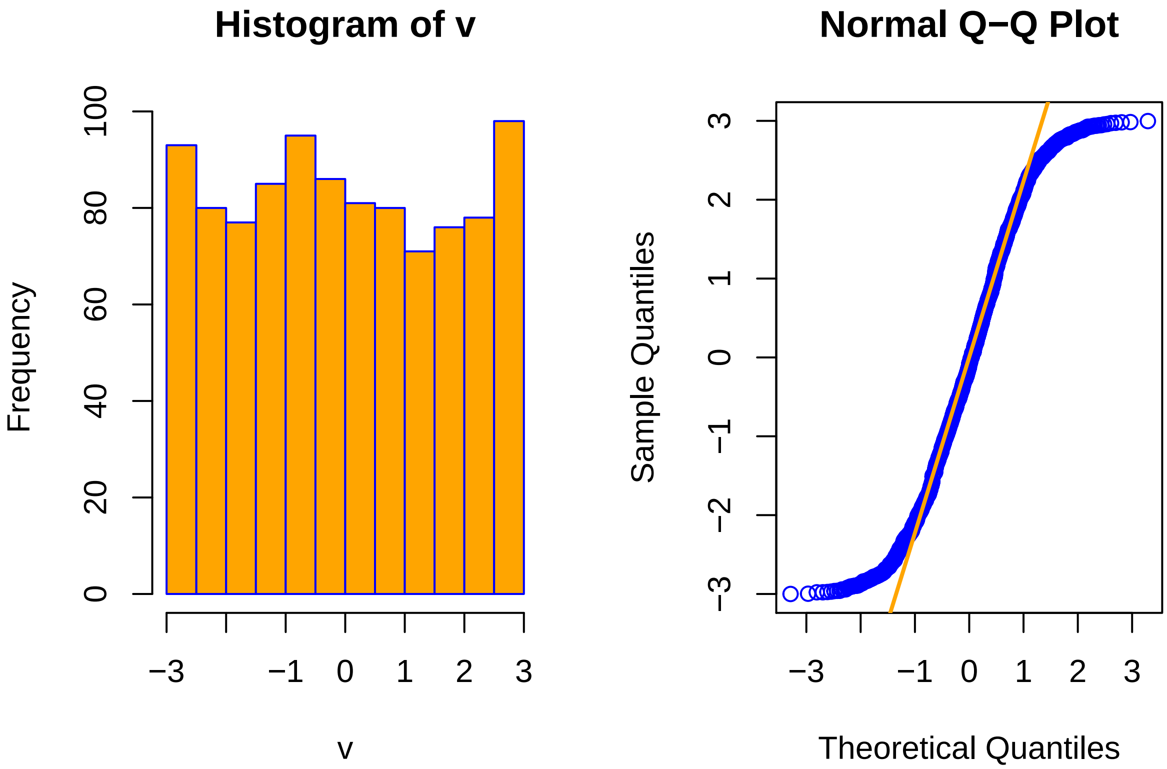
From the QQ plot, we see that the sample has high frequency in values zero to five, therefore its quantiles will increase slower in this region relative to the standard normal quantiles. However, after the sample’s value is above five, the samples quantile will increase faster than the standard normal quantile.

This next example shows a left skewed sample compared to the standard normal distribution. The sample is obtained from a chi-squared distribution with 2 degrees of freedom, however each value is multiplied by (-1) in order to reflect the distribution about the y-axis. The histogram shows that this distribution is in fact left skewed. The QQ plot shows that the points do not align along a line since the data sets come from different distributions.



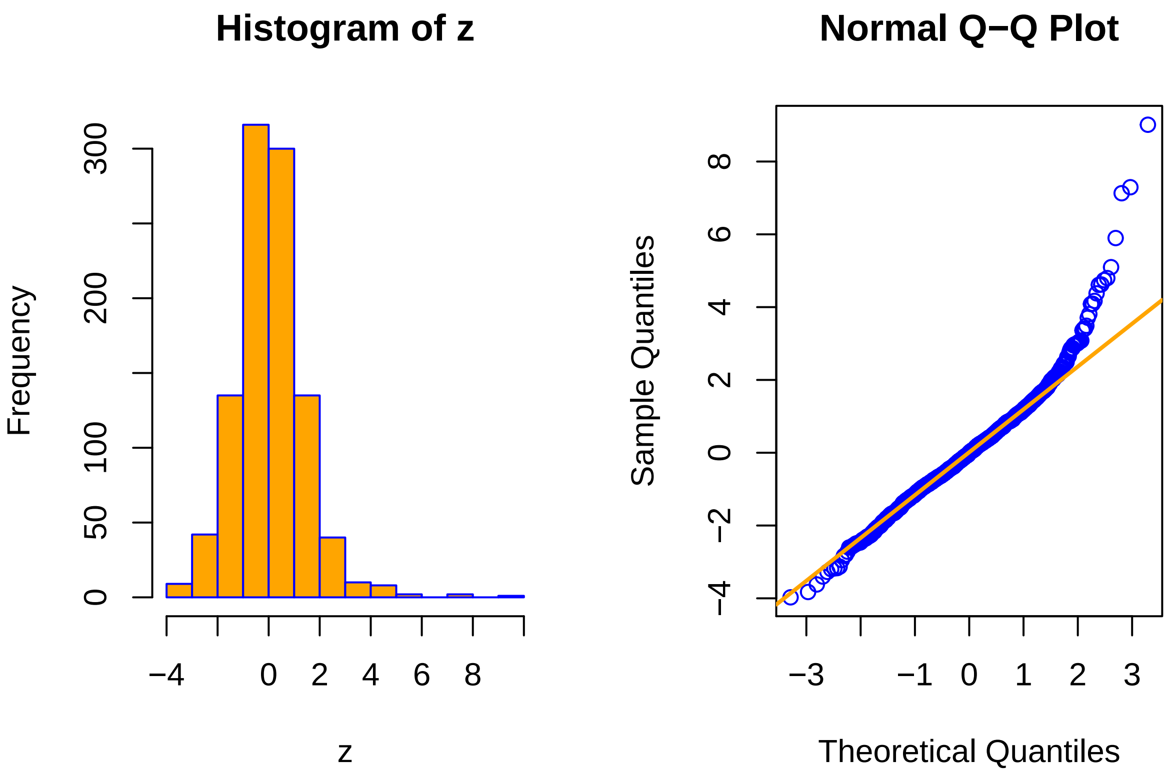
From the QQ plot, we see that the sample has low frequency in values -15 to -5; therefore its quantiles will increase rapidly in this region. However, after the sample's value is above -5, the samples quantile will increase slowly and tail off at 0, since that is the highest value in the sample.

The next two examples show samples that come from heavy and light tail distributions. The first example shows a sample taken from a uniform distribution (-3, 3) compared to the standard normal distribution. Although the comparison of this sample to the standard normal is not truly fair since the sample is strictly bounded between (-3,3), the results of the test are worth mentioning. Looking at the histogram, the sample has no tails beyond -3 or 3. This presents an interesting looking qq plot that is depicted below. The light tailed distributions yield an s shape depicted in the qq plot.



Approximately from the values (-3, -1.5), the sample grows slower than the standard normal distribution; therefore it takes longer for the sample quantiles to increase. This is shown by the concave up portion of the graph. From the values (-1.5,1.5), the sample seems to grow at approximately the same pace as the standard normal distribution; therefore their quantiles match in this region. Lastly, from the values (1.5, 3), the sample grows faster than the standard normal distribution; therefore the sample reaches its highest quantile before the standard normal distribution. This is why the sample quantile looks flat at the top; the sample has reached its highest quantile, but the standard normal has not and still needs to increase a little to reach it.

The last example depicts a sample with heavy tails relative to the standard normal distribution. This sample is obtained by simulating a random sample of a student's t distribution with 5 degrees of freedom. The histogram shows that the sample looks bell shaped, however when looking at the QQ plot we see an inverted s shape.



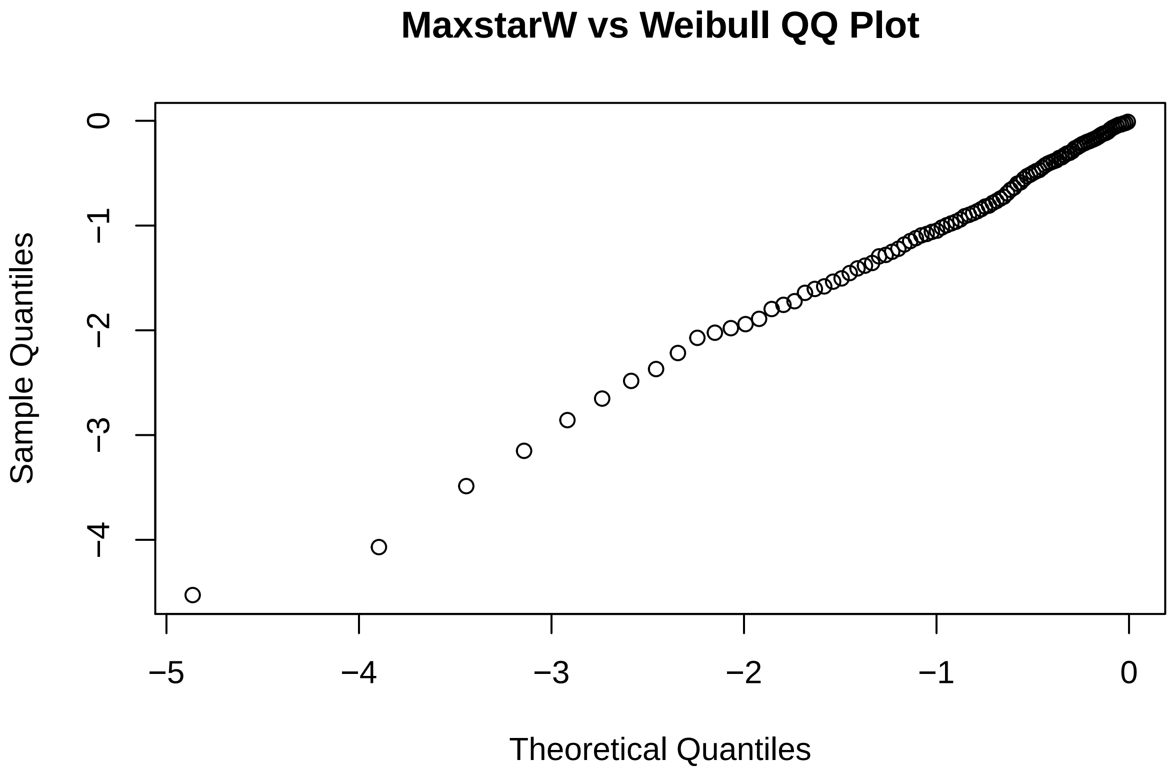
Approximately from the values (-3, -1.5), the sample grows faster than the standard normal distribution; therefore it takes a shorter time for the sample quantiles to increase. From the values (-1.5,1.5), the sample seems to grow at approximately the same pace as the standard normal distribution; therefore their quantiles match in this region. Lastly, from the values (1.5, 3), the sample grows slower than the standard normal distribution; therefore the sample reaches its highest quantile before the standard normal distribution. This is why the sample quantile looks vertical at the top; the standard normal distribution has reached its highest quantile, but the sample has not and still needs to increase to reach it.

These different types of plots help us distinguish how the sample compares to the base distribution. For example, if we have a sample and would like to see how it compares to the standard normal, we construct a QQ plot. If the QQ plot yields an inverted s shape, then we would reason that the sample probably does not come from the normal distribution. In addition, from our analysis of the different QQ plots, we would reason that the sample has heavy tails. Therefore, we have the option of comparing our sample to a heavy tailed distribution such as a two parameter Pareto, or a Weibull distribution. If we now construct a QQ plot of our sample against one of these heavy tailed distributions and the QQ plot yields a straight line, then we have reason to believe that our sample has a high probability of coming from the distribution that we tested.

QQ PLOT APPLICATION:

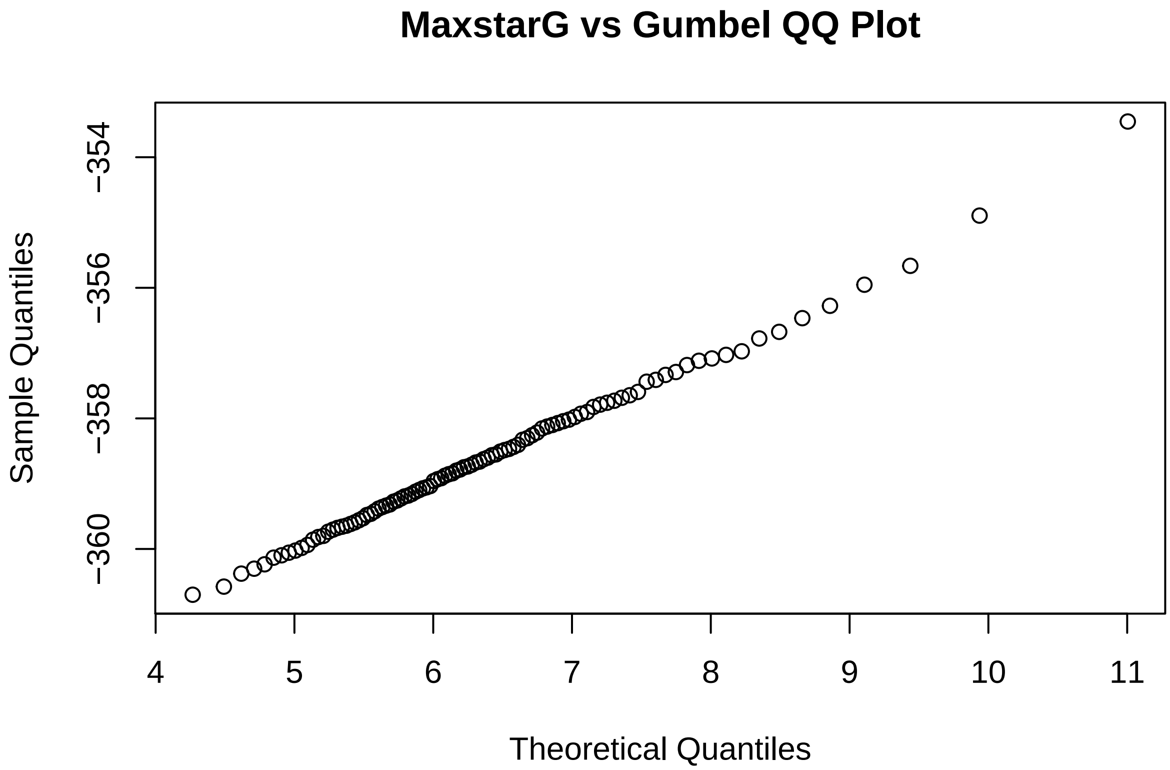
Part one of this document discusses an analysis of the extreme valuation theorem. Maximum Likelihood estimates are calculated from simulating different random variables.

In this first case, we will look at the sample taken from the Uniform simulation. According to the extreme valuation theorem (explained in greater detail in Part One), this sample should converge to a Weibull distribution as the sample size increases.



The QQ plot is constructed by plotting the sample generated from part one (we will name it MaxstarW) compared to the Weibull distribution. The parameters of the Weibull distribution are the maximum likelihood estimates found in part one. As one can see, the plot shows a straight line which shows that the quantiles match. Therefore, we have reason to believe that extreme valuation theorem does hold in this case.

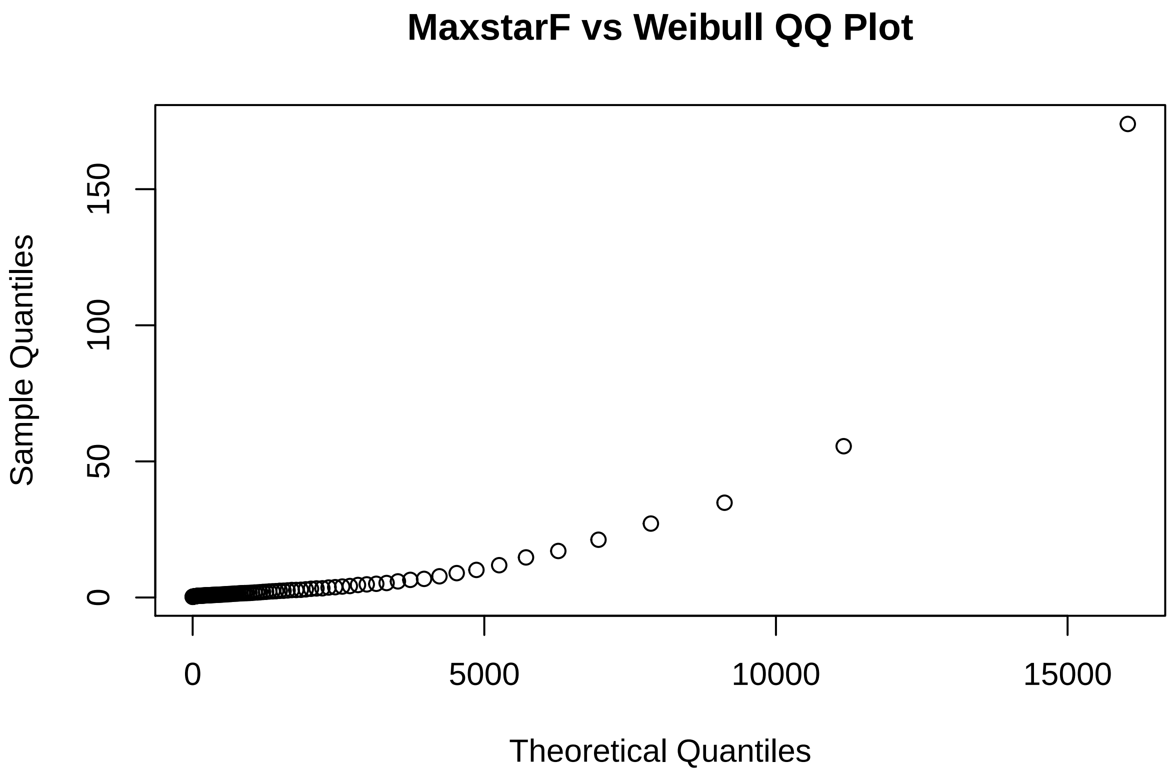
In the next case, we will look at the sample taken from the Exponential simulation. According to the extreme valuation theorem (explained in greater detail in Part One), this sample should converge to a Gumbel distribution as the sample size increases.



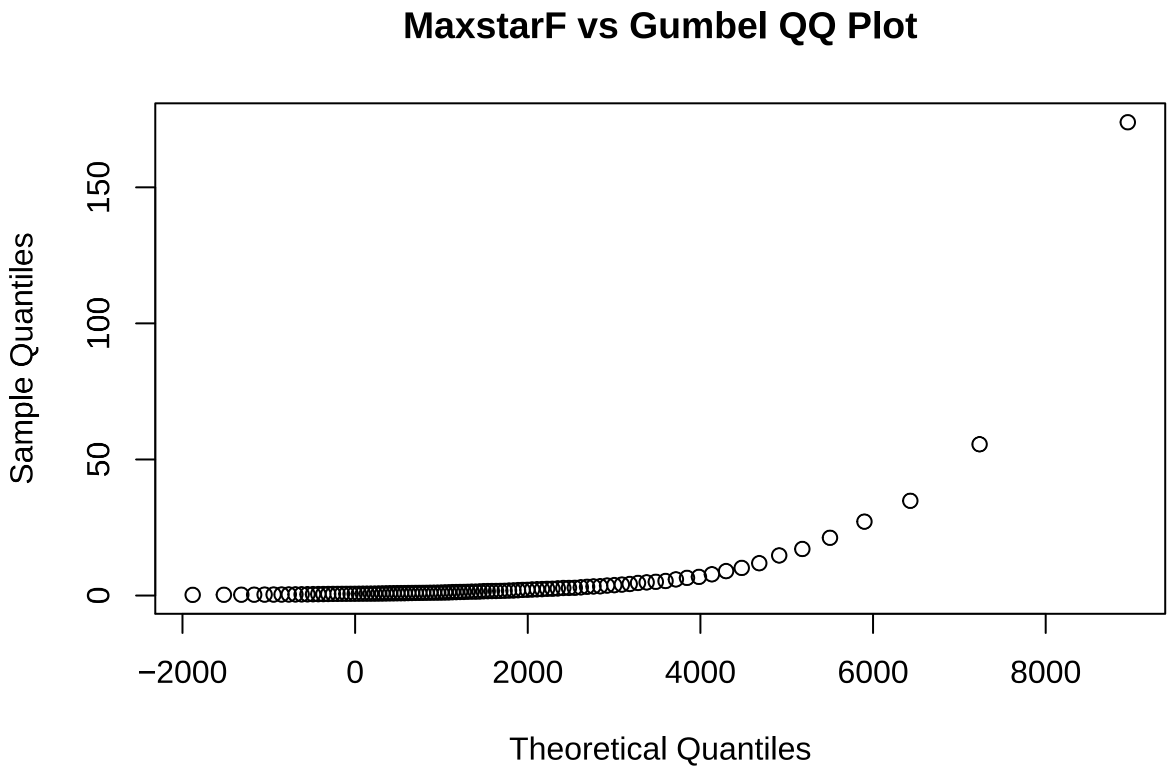
The QQ plot is constructed by plotting the sample generated from an Exponential simulation (we will name it MaxstarG) compared to the Gumbel distribution. The parameters of the Gumbel distribution are the maximum likelihood estimates found in part one. As one can see, the plot shows a straight line which shows that the quantiles match. Therefore, we have reason to believe that extreme valuation theorem does hold in this case also.

In the final case, we will look at the sample taken from the Frechet simulation. According to the extreme valuation theorem (explained in greater detail in Part One), this sample should converge to a Frechet distribution as the sample size increases.

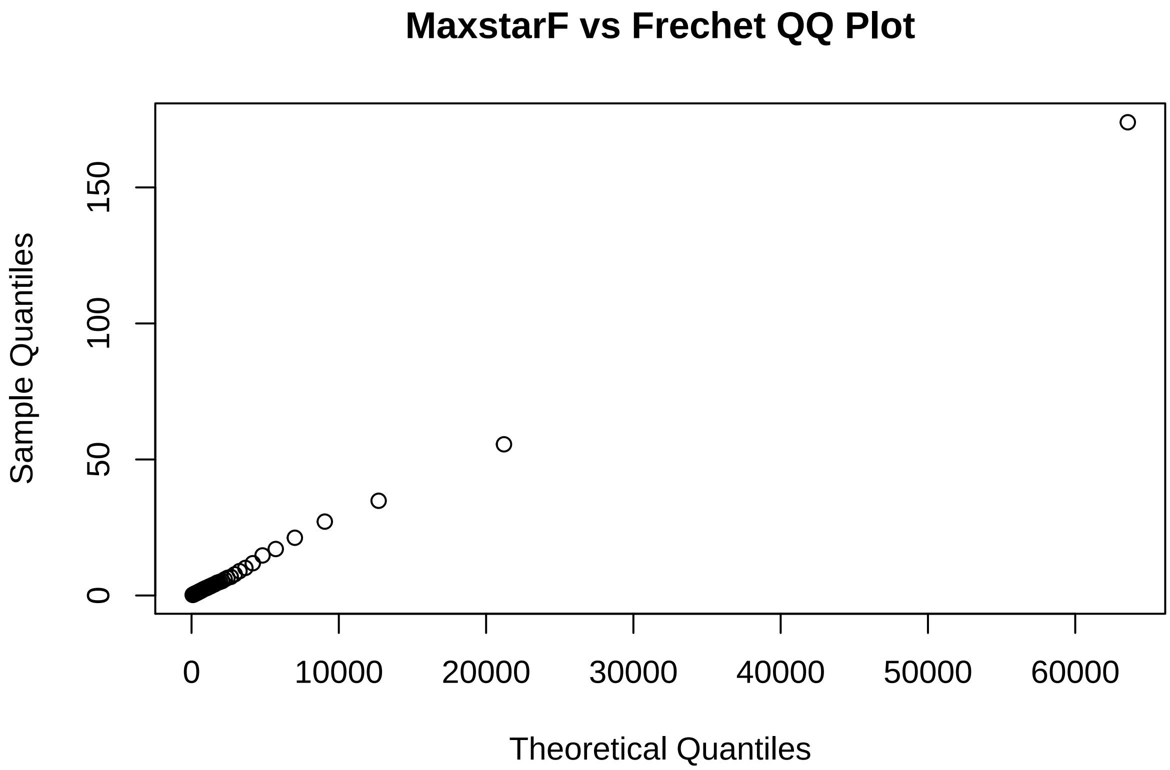
Using QQ plots, we will show that the Frechet distribution is the best distribution of the three to use for the Frechet simulation.



This QQ plot is constructed by plotting the sample generated from Frechet simulation (we will name it MaxstarF) compared to the Weibull distribution. The parameters of the Weibull distribution are found using the maximum likelihood of the Weibull distribution with this sample. As one can see from the plot, the quantiles do not match, and according to our QQ plot interpretation, the sample seems to be skewed compared to the Weibull distribution. Therefore, we do not have reason to believe that this sample tends to a Weibull distribution.



This QQ plot is constructed by plotting the sample generated from Frechet simulation (MaxstarF) compared to the Gumbel distribution. The parameters of the Gumbel distribution are found using the maximum likelihood of the Gumbel distribution with this sample. As one can see from the plot, the quantiles do not match, and according to our QQ plot interpretation, the sample seems to be skewed compared to the Gumbel distribution. Therefore, we do not have reason to believe that this sample tends to a Gumbel distribution.



The final QQ plot is constructed by plotting the sample generated from Frechet simulation (MaxstarF) compared to the Frechet distribution. The parameters of the Frechet distribution are found using the maximum likelihood of the Frechet distribution with this sample. As one can see from the plot, the quantiles do match which leads us to believe that this sample does tend toward a Frechet distribution. This is what the extreme valuation theorem predicts, and therefore we reason that the theorem holds true for all three cases.

References:

Engineering Statistics Handbook “Quantile-Quantile Plot” (2016) http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm

“Skews and Tails” (2016) http://www.google.com/search?q=heavy+tailed+qq+plot&client=safari&rls=en&prmd=ivns&ei=nhc-V5SwF8THmQG166Ug&start=10&sa=N

University of Virginia Library Research Data Services “Understanding Q-Q Plots” (2016) http://data.library.virginia.edu/understanding-q-q-plots/