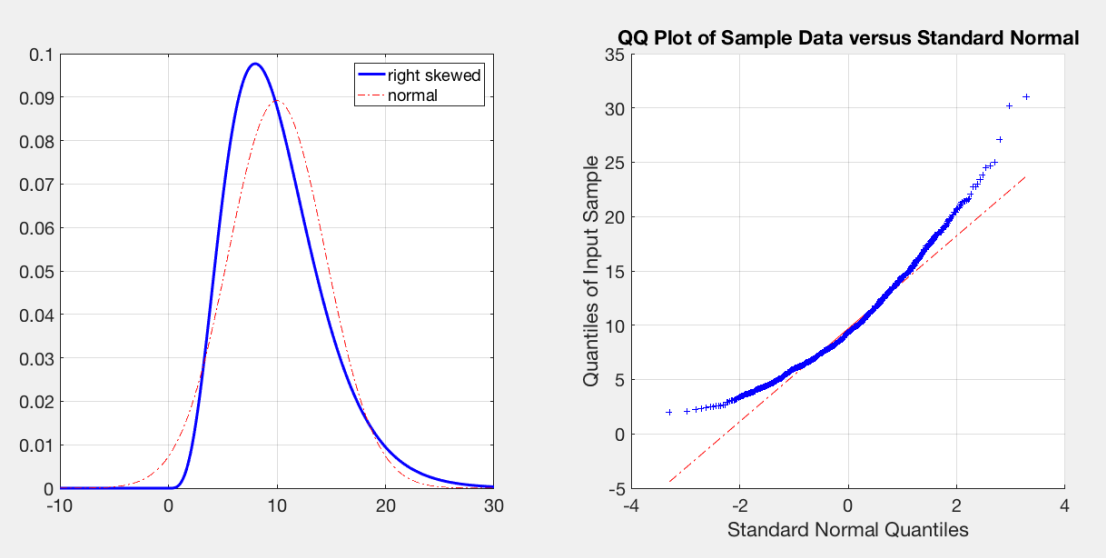
Probability and Statistics Fall 2020
Normal Quantile-Quantile Plot (12/25/2020)

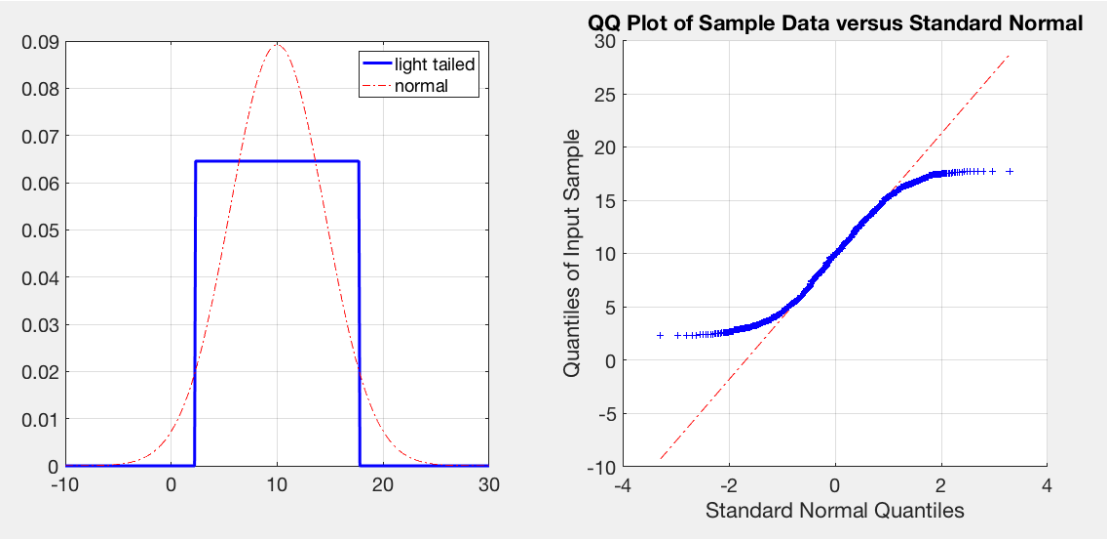
In data analysis, a normal quantile-quantile plot is usually used for checking whether data (i.e., actual sample) present normality. For example, if X is a random variable following normal distribution, then sample points of X are likely to be normal (or close to normal). Therefore, by comparing X versus Z (i.e., standard normal distribution) at the SAME fraction (i.e., *f* in the textbook), the normal quantile-quantile plot should be linear (or close to linear). A normal quantile-quantile plot is a very practical tool for checking normality of data before you even start to do any analysis. You may find clues of population distribution (e.g., skewness). In addition, for double check, you may also plot a normal quantile-quantile plot again after you do variable transformation. The following are examples to help you understand normal quantile-quantile plots.

1. Right skewed distribution



The left figure at above shows pdf of a right skewed distribution X (blue) together with a reference normal random variable N (dotted red) for reference. A normal quantile-quantile plot is shown on the right from a random sample of X simulated in Matlab. Notice at the left tail (i.e., when fraction *f* is small), X is larger than N at the SAME fraction *f*. Therefore, the normal quantile-quantile plot shifts upward from a straight line when X and Z are small. Similarly, at larger fraction *f* (i.e., right tail), X is also larger than N. Therefore, the normal quantile-quantile plot also shifts upward from a straight line when X and Z are large.

2. Light tailed distribution



The left figure shows pdf of a light tailed uniform distribution X (blue) together with a reference normal random variable N (dotted red) for reference. A normal quantile-quantile plot is shown on the right from a random sample of X simulated in Matlab. Notice at the left tail (i.e., when fraction *f* is small), X is larger than N at the SAME fraction *f*. Therefore, the normal quantile-quantile plot shifts upward from a straight line when X and Z are small. However, at larger fraction *f* (i.e., right tail), X is lower than N. Therefore, the normal quantile-quantile plot shifts downward from a straight line when X and Z are large.

From the above, it is shown that the shape of a normal quantile-quantile plot is determined by how far the cumulative distribution of X deviates from normal. Therefore, normal quantile-quantile plot is an useful tool for quickly checking normality in data (i.e., sample).