Quantile

To understand the mechanism of a quantile-quantile plot (Q-Q plot), we need to start with the definition of a quantile. From **Wikipedia**:

A kth q-quantile for a random variable is a value x such that the probability that the random variable will be less than x is at most $\frac{k}{q}$ *and the probability that the random variable will be greater than x is at most* $\frac{q-k}{q} = 1 - \frac{k}{q}$*. There are q*1 *of the q-quantiles, one for each integer k satisfying* $0 < k < q$*. In some cases the value of a quantile may not be uniquely determined, as can be the case for the median of a uniform probability distribution on a set of even size.*

Some q-quantiles have special names:
The only 2*-quantile is called the median*
The 3*-quantiles are called tertiles or terciles*
The 4*-quantiles are called quartiles*

Quantile-quantile Plot

Quantile-quantile plot is a graphical tool to compare distributions. It is basically plotting the quantiles of one distribution against the same quantiles of another distribution. To be specific, let x_i, $i = 1, \ldots, q$, be the q-quantiles of random variable x and let y_i, $i = 1, \ldots, q$, be the q-quantiles of random variable y. Then the Q-Q plot is basically the scatter plot of pairs (x_i, y_i), $i = 1, \ldots, q$.

And why scatter-plot of the paired quantiles of two distributions would help compare them? Intuitively, if these two distributions are exactly the same and if you are using their analytical quantiles then $x_i = y_i$, for $i = 1, \ldots, q$. In other words, (x_i, y_i)'s will be strictly on a straight 45° line.

And in most cases you would not have the analytical quantiles from the distribution (you might barely know the distribution). All you have are the independent realizations (data) from that distribution. Fortunately, easy and fast numerical way exists to calculate the "sample quantiles". `quantile` function in `R` does that! Because of the randomness of the data, even though the two distributions are exactly the same, their paired quantiles would not be strictly aligned on a straight 45° line. However, they should be quite centered around this line.

Optional Practice: Create a Q-Q plot yourself!

Study the `quantile` function in `R`. Simulate `y1` from standard normal distribution with sample size 1000, simulate `y2` from standard Cauchy distribution. Compute the 50-quantiles of both `y1` and `y2`, and scatter plot these quantiles.