The Probability Function:

This function is measuring the probability that two hashes, when compared, are found to be the same.

$$P(f) = (1 - (fq)^n)(1 - A)$$

Where f is the fraction of hash variables changed, q and n are unknown constants, and A is the variable used to account for user error. One possible goal with this function is to create a python program that data, $P(f), f$, can be fed into and through analysis the variables A, q, and n could be determined. Another python program would need to be created to check the accuracy of the first one, by using known values for all the variables and comparing the data output to the data input for the first program.

The function will heavily resemble the function: $(1 - x^n)$ where n is the some unknown exponent in both cases. The $(1 - A)$ part of the probability function will be responsible for putting an upper bound on the $P(f)$ data range, with there being an upper bound of 1 when A is zero, and an upper bound of zero when A is 1. By measuring the difference in the upper bound from positive 1 at zero on the x axis, the variable A can be determined .

For the variables q and n, the lower bound of the data range must be examined. The values of q must fall between 0 and 1, otherwise negative data will be produced, which is unreasonable for these purposes. The values of n however can range from 0 to very large numbers. The lower bound of the data range will naturally fall above zero, but by measuring the difference from zero at 1 on the x axis, an estimate of $(q^n)(1 - A)$ can be determined. Since A can be independently deterined, the $(1 - A)$ becomes a constant and the value of q^n can be determined.

By analyzing the shape of the $P(f)$ curve, the value of n can be determined roughly. Since the function resembles $(1-x^n)$, the value of n can be found by comparing the $P(f)$ curve to the known curves of functions such as $(1 - x), (1 - x^2), (1 - x^3), (1 - x^4)....$

Bayesian Statistics:

Bayesian statistics, as opposed to frequentist statistics, use certain evidence regarding the event being measured to create a more complete probability of the event occurring.

$$p(A|B) = \frac{p(B|A)*p(A)}{p(B)}$$

where $p(A|B)$ is the probability of event A given evidence B, $p(B|A)$ is the probability of evidence B happening given event A happens, $p(A)$ is the probability of event A happening, and $p(B)$ is the probability of evidence B happening. Bayesian statistics gives a more realistic picture of the probability of an event occurring by using the probability of concurrent evidence to shape the statistics. In this case, the prior evidence will be probabilities of prior $p(f)$ results. Thus the probability of any particular set of data given a, q, and n will be the product of every individual $p(f)$, where each f value is distinct.