

## Probability vs Likelihood

Probability attaches to possible results; likelihood attaches to hypotheses

Hypotheses, unlike results, are neither mutually exclusive nor exhaustive.

Because we generally do not entertain the full set of alternative hypotheses and because some are nested within others, the likelihoods that we attach to our hypotheses do not have any meaning in and of themselves; only the relative likelihoods — that is, the ratios of two likelihoods — have meaning.

## For a binomial Distribution

To compute probabilities one assumes that the two arguments (number of tries and the probability of success) are *given*. They are the parameters of the distribution. One varies the first argument (the different possible numbers of successes) in order to find the probabilities that attach to those different possible results.

By contrast, in computing a likelihood function, one is *given* the number of successes (7 in our example) and the number of tries (10). Instead of varying the possible results, one varies the probability of success (the third argument, not the first argument) in order to get the binomial likelihood function.

## Maximum Likelihood Function

It helps in Density Estimation which is estimating a joint probability distribution for a dataset.

1. Density Estimation is a process of

1. Choosing the probability distribution function and its parameters that best explains the observed data.

In Maximum Likelihood Estimation, we wish to maximize the probability of observing the data from the joint probability distribution given a specific probability distribution and its parameters,

Maximum Likelihood Estimation framework that is generally used for density estimation can be used to find a supervised learning model and parameters.

In the case of linear regression, the model is constrained to a line and involves finding a set of coefficients for the line that best fits the observed data.

In the case of logistic regression, the model defines a line and involves finding a set of coefficients for the line that best separates the classes. This cannot be solved analytically and is often solved by searching the space of possible coefficient values using an efficient optimization algorithm such as the BFGS algorithm or variants.