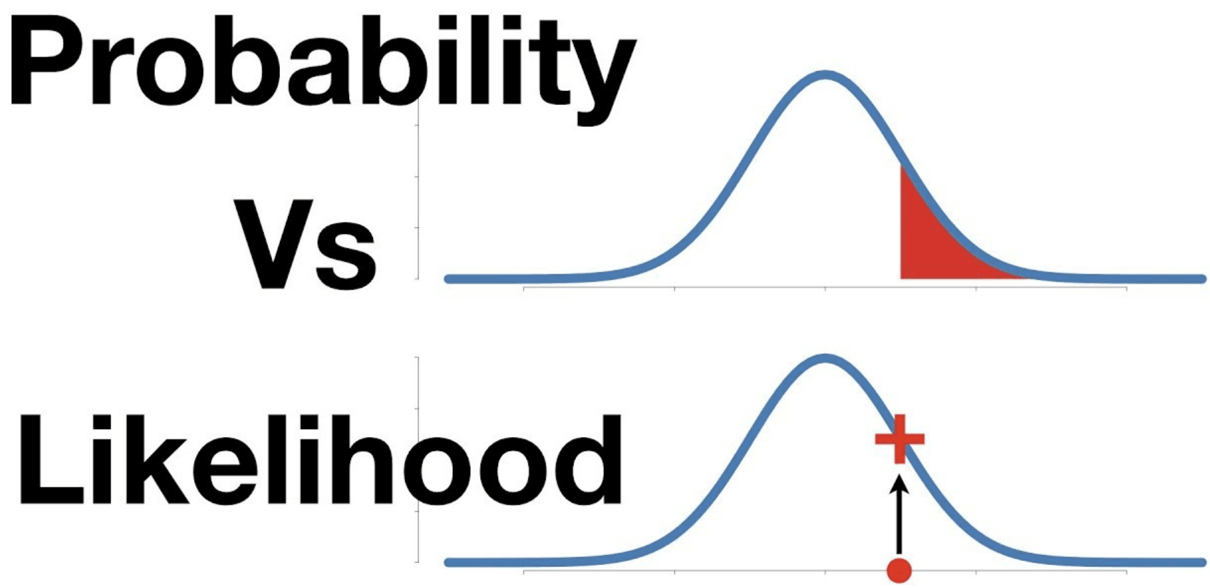
Understanding the Distinct Concepts of Likelihood and Probability in Statistics

medium.com/aimonks/understanding-the-distinct-concepts-of-likelihood-and-probability-in-statistics-44cae9e1b56a

Everton Gomede, PhD January 2, 2024

Introduction

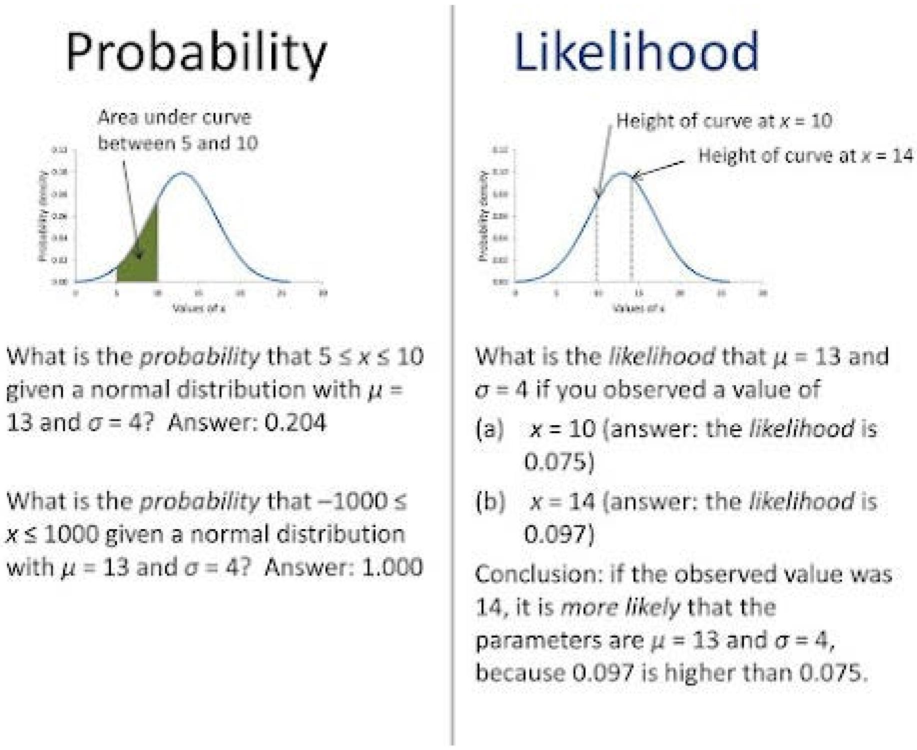
The realms of statistics and probability theory are rife with concepts that are both foundational and complex. Two such concepts are “likelihood” and “probability,” which, despite their frequent interchangeability in casual discourse, embody distinct meanings and applications within the statistical framework. This essay endeavors to delineate these concepts, elucidating their definitions, contextual applications, and the nuances that set them apart. By exploring these terms in detail, we aim to provide clarity and deepen the understanding of these pivotal statistical concepts.



> Probability tells the tale of what might happen in a world of chance, while likelihood reveals the story of our observations, explaining the world we see.

Understanding Probability

At its core, probability is a fundamental concept in statistics, reflecting the measure of the likelihood that a particular event will occur. It is rooted in the theory of probability, which is a branch of mathematics dealing with the analysis and interpretation of random events. The value of probability is quantitatively expressed on a scale from 0 to 1, where 0 denotes an impossible event, and 1 indicates an event that is certain to occur.



The practical applications of probability are vast and varied, spanning from simple scenarios like calculating the odds of a dice roll, to complex analyses such as predicting the likelihood of certain outcomes in financial markets. Probability theory is indispensable in fields such as actuarial science, weather forecasting, and risk assessment, where predicting future events based on a set of known parameters is crucial.

The terms “likelihood” and “probability” are often used in statistics, but they have distinct meanings and are used in different contexts.

1. refers to the measure of the chance that a certain event will occur. It’s a fundamental concept in the field of statistics and is used in a variety of contexts. Probabilities are always between 0 and 1, where 0 indicates impossibility and 1 indicates certainty. For example, the probability of flipping a coin and it landing on heads is 0.5, assuming the coin is fair.
2. , on the other hand, is a concept used in the context of statistical models. It is a measure of how probable a particular set of observations is, given a set of model parameters. Unlike probability, likelihood is not confined to a range between 0 and 1. It’s used in statistical inference, especially in methods like maximum likelihood estimation, where we try to find the parameter values that maximize the likelihood of observing the data we have.

In summary, while probability is a measure of the chance of an event occurring given certain conditions, likelihood is a measure of how well a set of data supports a particular model or parameter values. Probability is used more generally to describe the likelihood of events, while likelihood is a specific term used in the context of statistical models and inference.

Distinguishing Likelihood

Likelihood, while closely related to probability, occupies a distinct niche in statistical analysis. It refers to a measure used in the context of statistical models, primarily focusing on how probable a particular set of observed data is, given a set of model parameters. Unlike probability, likelihood is not bound by the 0 to 1 range and can take any non-negative value.

The concept of likelihood is pivotal in statistical inference, particularly in methodologies like maximum likelihood estimation (MLE). MLE is a method used to estimate the parameters of a statistical model, where the chosen parameters are those that maximize the likelihood of the observed data. In simpler terms, likelihood helps in determining which parameter values in a statistical model are most plausible, given the observed data.

Comparative Analysis

The distinction between likelihood and probability becomes more evident when we consider their contextual applications. Probability is used to predict the occurrence of future events based on known conditions or parameters. For instance, the probability of drawing a red card from a standard deck of playing cards can be calculated because the conditions (the composition of the deck) are known.

Conversely, likelihood deals with how well a set of observed data supports a specific model or set of parameters. It is retrospective, not predictive. Likelihood assesses the plausibility of different parameter values after the data has been observed, as seen in fields like econometrics and biostatistics.

Code

To illustrate the concepts of likelihood and probability using Python, we can create a synthetic dataset and analyze it through both lenses. For this example, let's consider a simple binary outcome scenario, such as a coin flip, and analyze it using both probability and likelihood concepts.

1. Creating a Synthetic Dataset: We will simulate coin flips.
2. Probability Analysis: We will calculate the probability of observing a certain number of heads in our dataset.
3. Likelihood Analysis: We will use a likelihood approach to estimate the probability of heads (parameter estimation) and visualize how likely different probabilities of heads are given our dataset.

Let's start by creating the synthetic dataset:

```
import numpy as np
import matplotlib.pyplot as plt

# Seed for reproducibility
np.random.seed(42)

true_p_heads = flips = np.random.binomial(, true_p_heads, )
```

Next, we will perform probability analysis:

```
# Count the number of heads
n_heads = np.sum(flips)

# Total number of flips
n_flips = len(flips)

probability = n_heads / n_flips()
```

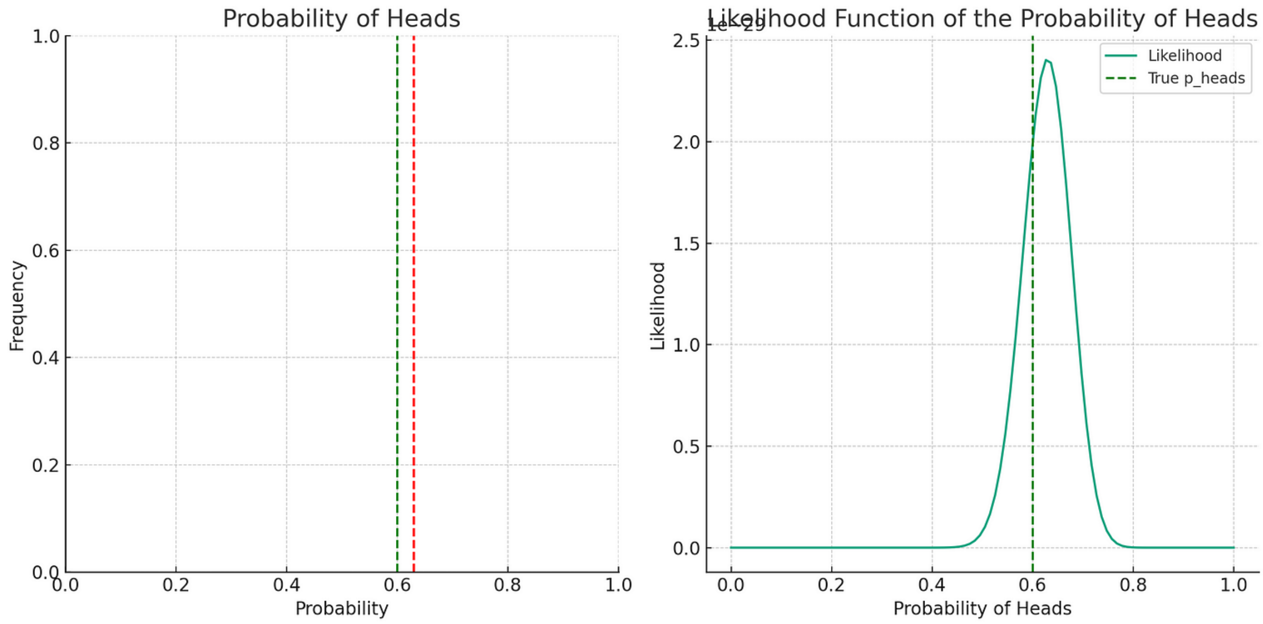
Finally, we will perform a likelihood analysis:

```
# Define a range of possible values for the probability of heads
p_heads_range = np.linspace(0, 1, 100)

# Calculate the likelihood for each value of p_heads
likelihoods = [(p ** n_heads) * ((1 - p) ** (n_flips - n_heads)) for p in
p_heads_range]

plt.plot(p_heads_range, likelihoods)plt.xlabel()plt.ylabel()plt.title()plt.show()
```

In this code, the probability part calculates the observed frequency of heads in the dataset, while the likelihood part calculates and plots how likely different probabilities of heads are, given the observed data. Let’s execute this code to see the results.



The analysis has produced two plots:

1. : This plot shows a red dashed line at the observed probability of heads (0.63), which is calculated from our synthetic dataset. The green dashed line represents the true probability of heads (0.6) that we used to generate the data. This plot illustrates the concept of probability as the observed frequency of an event (heads in this case) in the dataset.

2. : This plot displays the likelihood function for different probabilities of heads. The likelihood values are calculated for a range of probabilities (from 0 to 1). The green dashed line at 0.6 marks the true probability of heads. The peak of the curve indicates the most likely value for the probability of heads given the observed data, aligning closely with the true probability.

From the results, we can see that the observed probability of heads in the dataset is 0.63, with 63 heads observed out of 100 flips. The likelihood plot shows that the probability values around the true probability (0.6) are more likely given the observed data. This example illustrates how probability and likelihood are used differently: probability as a measure of the frequency of an event, and likelihood as a measure of how well different probabilities explain the observed data.

Conclusion

In conclusion, while the terms “likelihood” and “probability” are often used interchangeably, their meanings and applications in the realm of statistics are markedly different. Probability quantifies the chance of an event occurring under known conditions, while likelihood assesses how well a set of data supports a particular model or parameter values. Understanding these differences is crucial for anyone delving into statistical analysis, as it enables a more nuanced and accurate interpretation of statistical models and their outcomes. The distinction not only enriches one’s statistical vocabulary but also enhances the precision and depth of analytical reasoning in the field of statistics.