

# Bayesian vs. Frequentist Statistics

## Introduction

- 1. [Ben Lambert - A Student’s Guide to Bayesian Statistics](#)
- 2. [Bayesian Evaluation of Informative Hypotheses](#)

## Likelihood

Likelihood is a type of probability except that this describes the probability of the data that has already been observed given a certain hypothesis parameter. Probability, in general, refers to the possibility of occurrence of future events. It is worthwhile to draw the distinction between probability and likelihood here. Probability refers to varying data given a fixed hypothesis parameter(s). Likelihood refers to variable hypothesis for the fixed observed data. If ‘x’ represents our data and  $\theta$  represents our hypothesis parameter, the Likelihood is written as

$$Likelihood(\theta|data) = P(x|\theta)$$

It is worth pointing out likelihood does not form a valid probability since the Probability Density Function formed by varying theta with the observed data does not integrate to 1. It is for this reason that the term likelihood exists, i.e. to remind us that it is not a valid probability density.

## Inference

Inference refers to the process of identifying the distribution for the parameters that represent our hypothesis. However, this is a distinctly Bayesian definition of Inference, since a Frequentist believes in the existence of true parameter. We will spend some time understanding these differences in the next section. The posterior distribution is computed during the inference step and can be denoted as

$$P(\theta|x)$$

## The Bayesian and the Frequentist

In frequentist statistics, our probabilities are only based on what we have observed. In Bayesian statistics, we also use our prior understanding of the environment or the system that generated the observations to infer the possibility of future events. As Ben Lambert says in his book ‘A Students Guide to Bayesian Statistics’, ‘for Bayesians it is unnecessary for events to be repeatable in order to define a probability. They are merely abstractions to help express our uncertainty’. The fundamental difference between a Bayesian and a Frequentist point of view is determined by how one treats parameters and data. For a Frequentist, the uncertainty associated with a probability assigned to an event comes from randomness (aleatoric uncertainty), they do not consider the uncertainty arising from a lack of knowledge (epistemic uncertainty).

## The Bayesian

Data is fixed and in light of new data we update our beliefs. The belief represented by some parameter is a random variable that we update. They are comfortable with the idea that our knowledge of these parameters evolve over time. In the case where there is a true parameter value that can be estimated, Bayes theorem allows one to specify an uncertainty over this parameter.

## The Frequentist

Frequentist - The data is the result of sampling from a random process. They see the data as varying and the parameter of this random process that generates the data as being fixed. They view this parameter as being the average of an infinite number of experiments. This approach is particularly problematic for events that cannot be repeated.

☰ Contents

- [Introduction](#)
- [Likelihood](#)
- [Inference](#)
- [The Bayesian and the Frequentist](#)
- [The Bayesian](#)
- [The Frequentist](#)
- [The Differences Explained](#)
- [Inference](#)
- [Features of Bayesian Inference](#)
- [Features of Frequentist Inference](#)
- [Hypothesis Testing](#)
- [GRADED QUESTIONS \(15 mins\)](#)

# The Differences Explained

Ben Lambert’s example of the US elections helps to illustrate this subtle difference.

**Bayesian** - The probability of the Democrats winning the election is 0.75.

**Frequentist** - Since there is only one sample that can be obtained for a particular election, this is not a repeatable exercise and we cannot sample from a population of outcomes for this election.

For example, in a coin flip that is repeated 5 times where 4 heads show up, the probability of heads for a frequentist would be 4/5. The distribution that represents the generative process of getting a heads or a tails from a coin toss can be represented by the parameter  $\theta$ . When we want to determine the number of heads from ‘n’ coin tosses, this can be modeled using a Binomial distribution. If ‘O’ represents the observed outcome and ‘H’ represents the outcome being heads, then the likelihood can be written as

$$Likelihood = P(O = H / \theta) = 0.8$$

Frequentists use this as the probability of the parameter  $\theta$ .

However, in a Bayesian approach to estimate the posterior probability of the parameter  $\theta$  when a heads shows up, given that the prior probability  $P(\theta) = 0.5$

$$Posterior \propto Likelihood \times Prior \longrightarrow P(\theta|O = H) \propto P(O = H|\theta) \times P(\theta)$$

$$Posterior \propto 0.8 * 0.5$$

**To summarize: In Inference, we want to measure the probability of a hypothesis given certain data. Frequentists choose a hypothesis and determine the probability of the data given this hypothesis. They then use this as evidence of the hypothesis. Bayesians go one step further to invert this to get the probability of the hypothesis given the data.**

Suppose that you are measuring the heights of citizens in your state and there is a **hypothesis** that their heights are normally distributed with mean 5.75ft and a standard deviation of 0.4ft. We observe a person with a height of 6.5ft. Frequentists would have to compute the probability of this observation happening as

$$P(h = 6.5|N(\mu, \sigma))$$

and use a threshold to determine if our original hypothesis was valid. In Bayesian statistics, the posterior is a probability distribution of the parameters of our hypothesis given our observation of seeing a sample with height 6.5ft

$$P(N(\mu, \sigma)|h = 6.5)$$

However, in this case we use a subjective prior to compute this posterior. We trade an arbitrary threshold (usually 1% or 5% depending on the problem being solved) in frequentist statistics for a subjective prior in Bayesian statistics. One could argue that both of these are different ways to incorporate domain-knowledge and subjective expertise into the decision making process. Bayesians, however, do not need to accept or reject a hypothesis since they have the full distribution available as a result of the posterior and can therefore quantify the uncertainty with the hypothesis. More information about hypothesis testing is in the section below.

# Inference

[Reference - MIT Course Notes, Jeremy Orloff and Jonathan Bloom](#)

[Reference - Null Hypothesis Significance Testing \(NHST\).](#)

# Features of Bayesian Inference

1. Assign a probability to both the hypothesis (Posterior) and the data (Likelihood)
2. Utilizes expert knowledge through the formulation of ‘subjective’ priors. The use of priors has been a source of debate, however when this is clearly stated it allows everyone to understand and challenge the assumptions behind the results possibly allowing for refinement of the priors.

3. Can be computationally expensive to compute the posterior due to the need to integrate over several parameters. A lot of the times, we have to resort to approximate techniques since the integrals associated with the posterior calculation in Bayesian statistics cannot be computed analytically. A number of approximation techniques are employed.
  1. Laplacian approximation
  2. Variational approximation
  3. Monte Carlo techniques (the subject of this Specialization)
  4. Message passing algorithms

***The primary advantage to using a Bayesian approach is that it allows us to quantify the uncertainty associated with the parameters of interest.***

## Features of Frequentist Inference

1. No probability for the hypothesis (no posterior)
2. Based on NHST
3. No use of a prior, but subjectivity still exists through the use of p-values
4. Less computationally intensive
5. One disadvantage of the Maximum Likelihood approach used in Frequentist methods is that the model is prone to overfitting.

## Hypothesis Testing

Hypothesis testing using significance testing is performed by formulating a null hypothesis, and testing the probability of an event having occurred under this null hypothesis. If this probability is less than a predefined threshold (usually 5% but is very much context dependent), we reject the null hypothesis otherwise we say that the event has a reasonable chance of occurring under the null hypothesis. The general notion behind the null hypothesis is that this is a hypothesis we want to prove wrong, the opposite of the null hypothesis is the alternate hypothesis. P-values and confidence intervals are routinely used in research to convey the credibility of experiments. However, these p-values depend on the [experimental setup](#).

Suppose we have a group of measurements for temperatures in Virginia for the month of December in 2020. It turns out that the mean temperature is 45F. The mean temperature across all the years from 1980 to 2020 was 50F. We want to test if this was an anomaly or not using NHST with p-values. Our null hypothesis here is that this is not an anomaly and the alternate hypothesis is that this was an unusually colder December. We have to select a threshold to accept or reject our hypothesis, as mentioned above this threshold is usually selected to be 5%. In order to confirm our suspicion we must reject the null hypothesis. We start by obtaining the probability of the data occurring under the null hypothesis, i.e. the probability of the mean temperature in 2020 given the distribution of mean temperatures from 1980 to 2020. If this probability is less than 5%, we reject the null hypothesis saying that it is unlikely for the data in 2020 to have come from the null hypothesis distribution. It is therefore a colder winter than Virginia is accustomed to seeing. If the probability is greater than 5%, we fail to reject the null hypothesis since there is some reason to believe (based on our subjective threshold) that this data could have been produced by the null hypothesis distribution. It is therefore no colder than expected and hence not an anomaly.

## GRADED QUESTIONS (15 mins)

1. In the Bayesian approach, it is unnecessary for events to be repeatable in order to define a probability
  - a. True
  - b. False
2. Bayesians use a prior to incorporate previous knowledge to make inferences while Frequentists do not
  - a. True
  - b. False
3. One way Frequentists incorporate their domain knowledge in through the use of p-values to reject a hypothesis
  - a. True
  - b. False
4. Frequentists express a probability over a hypothesis while Bayesians do not
  - a. True
  - b. False

5. Bayesians do not need to accept or reject a hypothesis since they have the full distribution of the hypothesis allowing them to quantify the uncertainty of that hypothesis
- a. True
  - b. False

---

By Srijith Rajamohan, Ph.D.

© Copyright 2021.