Bayesian Statistics for Data Science

Ankit
Rathi

Bayesian Statistics for Data Science

This is the 5th post of blog post series 'Probability & Statistics for Data Science', this post covers these topics related to Bayesian statistics and their significance in data science.

- *Frequentist Vs Bayesian Statistics*
- *Bayesian Inference*
- *Test for Significance*
- *Significance in Data Science*

Frequentist Vs Bayesian Statistics

Frequentist Statistics tests whether an event (hypothesis) occurs or not. It calculates the probability of an event in the long run of the experiment. A very common flaw found in

frequentist approach i.e. dependence of the result of an experiment on the number of times the experiment is repeated.

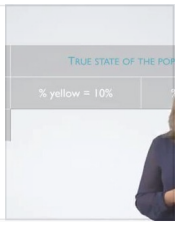
Frequentist statistics suffered some great flaws in its design and interpretation which posed a serious concern in all real life problems:

1. p-value & Confidence Interval (C.I) depend heavily on the sample size.
2. Confidence Intervals (C.I) are not probability distributions

Bayesian statistics is a mathematical procedure that applies probabilities to statistical problems. It provides people the tools to update their beliefs in the evidence of new data.

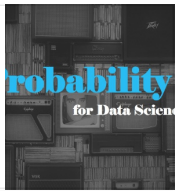
Frequentist Statistics	Bayesian Statistics
Parameters fixed	Parameters vary
Data varies	Data fixed
Probability P(D/Ø)	Likelihood P(Ø/D)
Confidence Interval	Credible Interval
No Prior	Strength of Prior

Frequentist Vs Bayesian Statistics



Bayesian Inference

To understand *Bayesian Inference*, you need to understand *Conditional Probability* & *Bayes Theorem*, if you want to review these concepts, please refer my earlier post in this series.

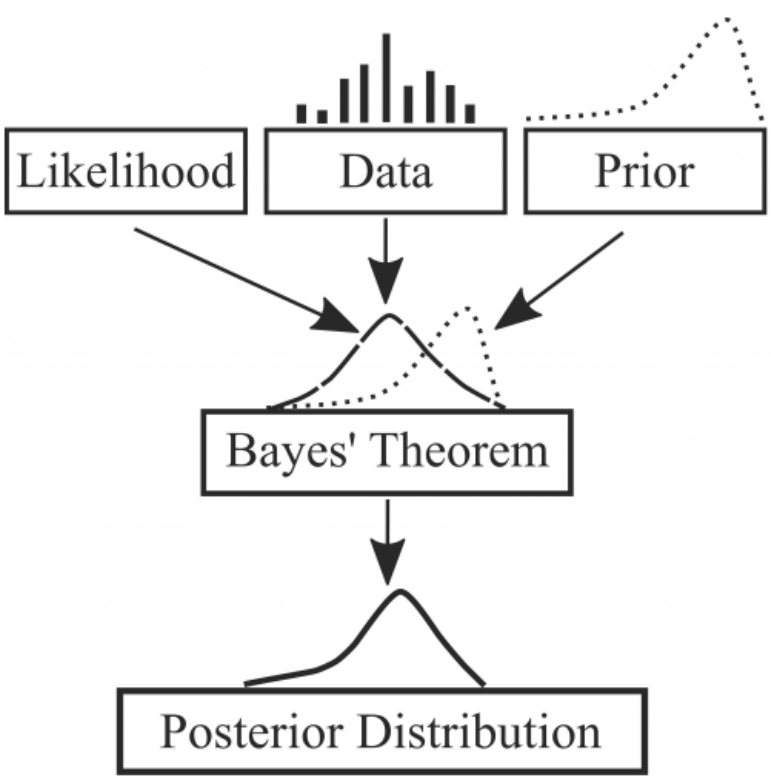


Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available

available.

An important part of *Bayesian Inference* is the establishment of *parameters* and *models.* Models are the mathematical formulation of the observed events. Parameters are the factors in the models affecting the observed data. To define our model correctly , we need two mathematical models before hand. One to represent the *likelihood function* and the other for representing the distribution of *prior beliefs* **.** The product of these two gives the *posterior belief* distribution.

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$



Courtesy: http://jason-doll.com/wordpress/?page_id=127

Likelihood Function

A *likelihood function* is a function of the parameters of a statistical model, given specific observed data. *Probability* describes the plausibility of a random outcome, without reference to any observed data while *Likelihood* describes the plausibility of a model parameter value, given specific observed data.

$$L(\emptyset/x) = cg(x/\emptyset)$$

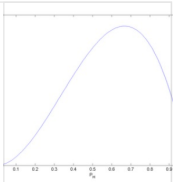
L = Likelihood function
Ø = Parameters in probability model
x = Data
g = Probability density function
c = Constant

Likelihood function

Likelihood function - Wikipedia

In Bayesian inference, although one can speak about the likelihood of any proposition or random variable given another…

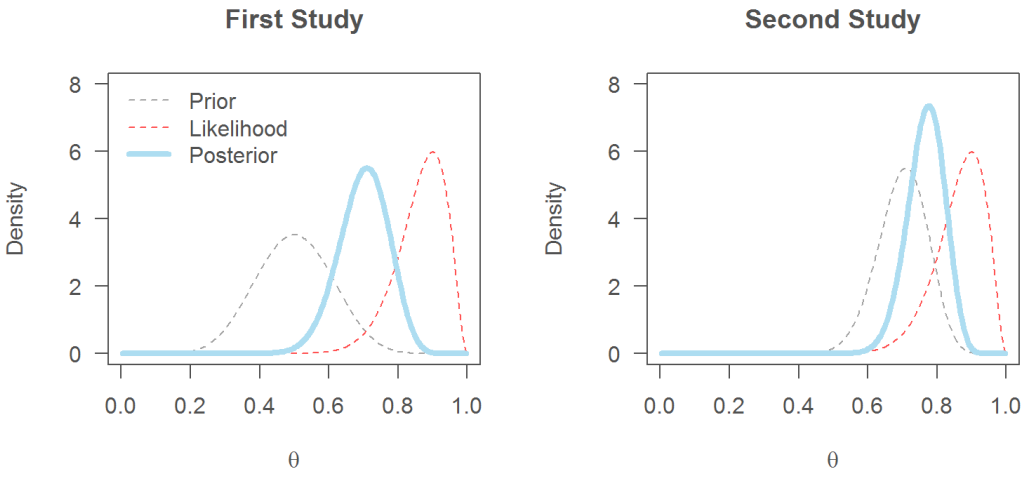
en.wikipedia.org



Prior & Posterior Belief distribution

Prior Belief distribution is used to represent our strengths on beliefs about the parameters based on the previous experience. *Posterior Belief distribution* is derived from multiplication of *likelihood function & Prior Belief distribution.*

As we collect more data, our posterior belief move towards prior belief from likelihood:

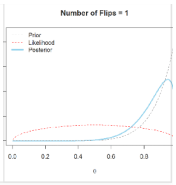


Courtesy: https://jimgrange.wordpress.com/2016/01/18/pesky-priors/

(Pesky?) Priors

When I tell people I am learning Bayesian statistics, I tend to get one of two responses: either people look at me…

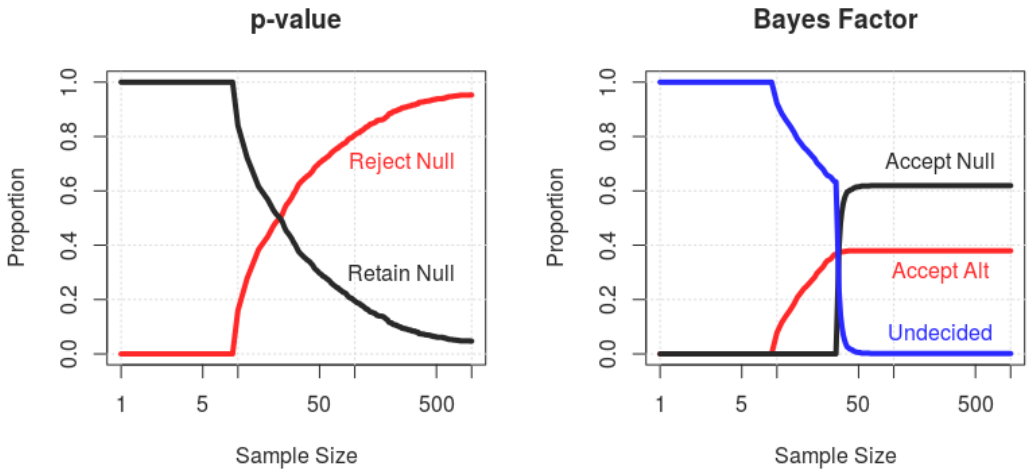
jimgrange.wordpress.com



Test for Significance

Bayes factor

Bayes factor is the equivalent of *p-value* in the *Bayesian* framework. The *null hypothesis* in Bayesian framework assumes ∞ probability distribution only at a particular value of a parameter (say θ=0.5) and a zero probability else where. The *alternative hypothesis* is that all values of θ are possible, hence a flat curve representing the distribution.



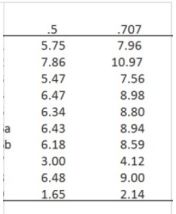
Courtesy: http://areshenk-research-notes.com/bayes-factors-and-stopping-rules/

Using *Bayes Factor* instead of *p-values* is more beneficial in many cases since they are independent of intentions and sample size.

Replacing p-values with Bayes-Factors: A Miracle Cure for the Replicability Crisis in Psychological…

How Science Should Work Lay people, undergraduate students, and textbook authors have a simple model of science…

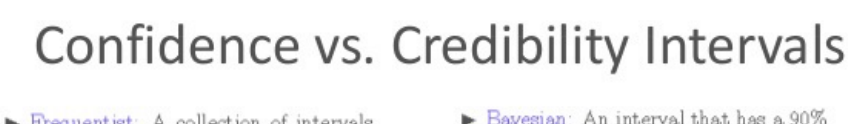
replicationindex.wordpress.com

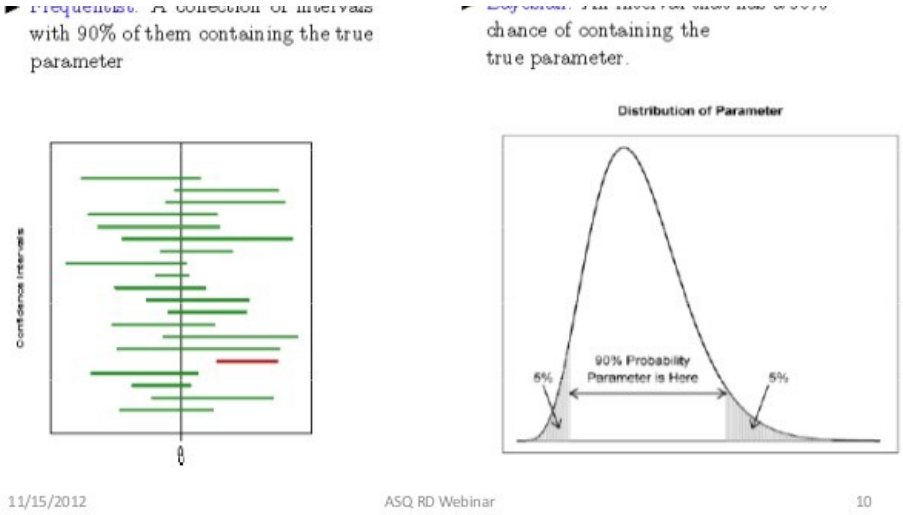


	.5	.707
	5.75	7.96
	7.86	10.97
	5.47	7.56
	6.47	8.98
	6.34	8.80
a	6.43	8.94
b	6.18	8.59
	3.00	4.12
	6.48	9.00
	1.65	2.14

High Density Interval (HDI)

High Density Interval (HDI) or *Credibility Interval* is equivalent to *Confidence Interval* (CI) in *Bayesian* framework. HDI is formed from the posterior distribution after observing the new data.





Courtesy: https://www.slideshare.net/ASQwebinars/bayesian-methods-in-reliability-engineering-15204318

Using *High Density Interval* (HDI) instead of *Confidence Interval* (CI) is more beneficial since they are independent of intentions and sample size.

Confidence vs. Credibility Intervals

Tomorrow, for the final lecture of the Mathematical Statistics course, I will try to illustrate - using Monte Carlo…

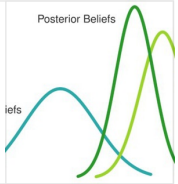
freakonometrics.hypotheses.org

Moreover, there is a nice article published on AnalyticsVidhya on this which elaborate on these concepts with examples:

Bayesian Statistics explained to Beginners in Simple English

Introduction Bayesian Statistics continues to remain incomprehensible in the ignited minds of many analysts. Being…

www.analyticsvidhya.com



Significance in Data Science

Bayesian statistics encompasses a specific class of models that could be used for Data Science. Typically, one draws on Bayesian models for one or more of a variety of reasons, such as:

- having relatively few data points
- having strong prior intuitions
- having high levels of uncertainty