

[Become a Top-Notch AI & ML Professional - View Program \(10+ courses\)](https://courses.analyticsvidhya.com/bundles/ai-blackbelt-beginner-to-master/?utm_source=blog&utm_medium=flashstrip&utm_campaign=3_feb_reserve_seat)
(https://courses.analyticsvidhya.com/bundles/ai-blackbelt-beginner-to-master/?utm_source=blog&utm_medium=flashstrip&utm_campaign=3_feb_reserve_seat)

 [LOGIN / REGISTER \(HTTPS://ID.ANALYTICSVIDHYA.COM/AUTH/LOGIN/?\)](https://id.analyticsvidhya.com/auth/login/)

[NEXT=HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/06/BAYESIAN-STATISTICS-BEGINNERS-SIMPLE-ENGLISH/\)](https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/)



(https://courses.analyticsvidhya.com/bundles/ai-blackbelt-beginner-to-master/?utm_source=blog&utm_medium=top_banner_blog&utm_campaign=pre-bookseat)

[BEGINNER \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/BEGINNER/\)](https://www.analyticsvidhya.com/blog/category/beginner/)

[PROBABILITY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/PROBABILITY/\)](https://www.analyticsvidhya.com/blog/category/probability/)

[R \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/R/\)](https://www.analyticsvidhya.com/blog/category/r/)

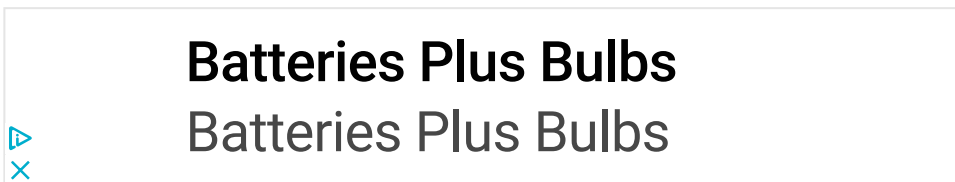


[STATISTICS \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/STATISTICS/\)](https://www.analyticsvidhya.com/blog/category/statistics/)

[TECHNIQUE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/TECHNIQUE/\)](https://www.analyticsvidhya.com/blog/category/technique/)

Bayesian Statistics explained to Beginners in Simple English

[NSS \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/AUTHOR/NSS/\)](https://www.analyticsvidhya.com/blog/author/nss/), JUNE 20, 2016 [LOGIN TO BOOKMARK THIS ARTICLE \(HTTP...](#)



Overview

- The drawbacks of frequentist statistics lead to the need for Bayesian Statistics
- Discover Bayesian Statistics and Bayesian Inference
- There are various methods to test the significance of the model like p-value, confidence interval, etc

Introduction

Bayesian Statistics continues to remain incomprehensible in the ignited minds of many analysts. Being amazed by the incredible power of machine learning (http://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=BayesianStatisticsforBeginnersArticle), a lot of us have become unfaithful to statistics. Our focus has narrowed down to exploring machine learning. Isn't it true?

We fail to understand that machine learning is not the only way to solve real world problems. In several situations, it does not help us solve business problems, even though there is data involved in these problems. To say the least, knowledge of statistics

(http://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=BayesianStatisticsforBeginnersArticle), will allow you to work on complex analytical problems, irrespective of the size of data.

In 1770s, Thomas Bayes introduced 'Bayes Theorem'. Even after centuries later, the importance of 'Bayesian Statistics' hasn't faded away. In fact, today this topic is being taught in great depths in some of the world's leading universities.

With this idea, I've created this beginner's guide on Bayesian Statistics. I've tried to explain the concepts in a simplistic manner with examples. Prior knowledge of basic probability & statistics (http://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=BayesianStatisticsforBeginnersArticle) is desirable. You should check out this course (http://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=BayesianStatisticsforBeginnersArticle) to get a comprehensive low down on statistics and probability.

By the end of this article, you will have a concrete understanding of Bayesian Statistics and its associated concepts.



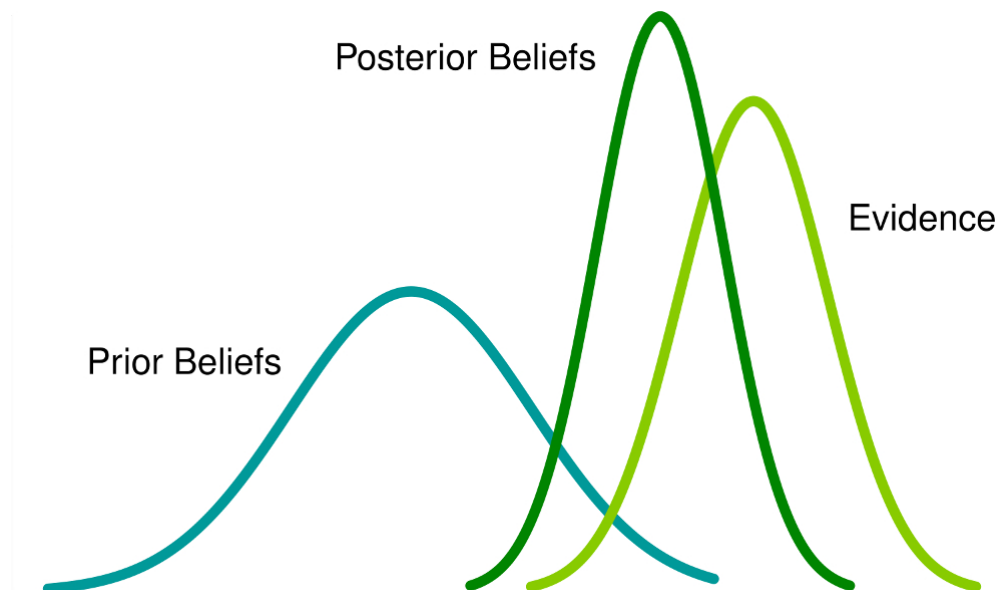


Table of Contents

1. Frequentist Statistics
2. The Inherent Flaws in Frequentist Statistics
3. Bayesian Statistics
 - Conditional Probability
 - Bayes Theorem
4. Bayesian Inference
 - Bernoulli likelihood function
 - Prior Belief Distribution
 - Posterior belief Distribution



5. Test for Significance – Frequentist vs Bayesian

- p-value
- Confidence Intervals
- Bayes Factor
- High Density Interval (HDI)

Before we actually delve in Bayesian Statistics, let us spend a few minutes understanding *Frequentist Statistics*, the more popular version of statistics most of us come across and the inherent problems in that.

1. Frequentist Statistics

The debate between *frequentist* and *bayesian* have haunted beginners for centuries. Therefore, it is important to understand the difference between the two and how does there exists a thin line of demarcation!

It is the most widely used inferential technique in the statistical world. Infact, generally it is the first school of thought that a person entering into the statistics world comes across.

Frequentist Statistics tests whether an event (hypothesis) occurs or not. It calculates the probability of an event in the long run of the experiment (i.e the experiment is repeated under the same conditions to obtain the outcome).

Here, the sampling distributions of **fixed size** are taken. Then, the experiment is theoretically repeated **infinite number of times** but practically done with a stopping intention. For example, I perform an experiment with a stopping intention in mind that I will stop the experiment when it is repeated 1000 times or I see minimum 300 heads in a coin toss.

Let's go deeper now.

Now, we'll understand *frequentist statistics* using an example of coin toss. The objective is to estimate the fairness of the coin. Below is a table representing the frequency of heads:

no. of tosses	no. of heads	difference
10	4	-1
50	25	0
100	44	-6
500	255	5
1000	502	2
5000	2533	33
10000	5067	67

We know that probability of getting a head on tossing a fair coin is 0.5. No. of heads represents the actual number of heads obtained. Difference is the difference between $0.5 * (\text{No. of tosses}) - \text{no. of heads}$.

An important thing is to note that, though the difference between the actual number of heads and expected number of heads(50% of number of tosses) increases as the number of tosses are increased, the proportion of number of heads to total number of tosses approaches 0.5 (for a fair coin).

This experiment presents us with 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.*

To know more about frequentist statistical methods, you can head to this [excellent course](https://classroom.udacity.com/courses/ud201/lessons/1306898579/concepts/1611758530923) (<https://classroom.udacity.com/courses/ud201/lessons/1306898579/concepts/1611758530923>) on inferential statistics.



2. The Inherent Flaws in Frequentist Statistics

Till here, we've seen just one flaw in *frequentist statistics*. Well, it's just the beginning.

20th century saw a massive upsurge in the *frequentist statistics* being applied to numerical models to check whether one sample is different from the other, a parameter is important enough to be kept in the model and various other manifestations of hypothesis testing. But *frequentist statistics* suffered some great flaws in its design and interpretation which posed a serious concern in all real life problems. For example:

1. *p-values* measured against a sample (fixed size) statistic with some stopping intention changes with change in intention and sample size. i.e If two persons work on the same data and have different stopping intention, they may get two different *p-values* for the same data, which is undesirable.

For example: Person A may choose to stop tossing a coin when the total count reaches 100 while B stops at 1000. For different sample sizes, we get different t-scores and different *p-values*. Similarly, intention to stop may change from fixed number of flips to total duration of flipping. In this case too, we are bound to get different *p-values*.

2- Confidence Interval (C.I) like *p-value* depends heavily on the sample size. This makes the stopping potential absolutely absurd since no matter how many persons perform the tests on the same data, the results should be consistent.

3- Confidence Intervals (C.I) are not probability distributions therefore they do not provide the most probable value for a parameter and the most probable values.

These three reasons are enough to get you going into thinking about the drawbacks of the *frequentist approach* and why is there a need for *bayesian approach*. Let's find it out.



From here, we'll first understand the basics of Bayesian Statistics.

3. Bayesian Statistics

"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."

You got that? Let me explain it with an example:

Suppose, out of all the 4 championship races (F1) between Niki Lauda (https://en.wikipedia.org/wiki/Niki_Lauda) and James hunt (https://en.wikipedia.org/wiki/James_Hunt), Niki won 3 times while James managed only 1.

So, if you were to bet on the winner of next race, who would he be ?

I bet you would say Niki Lauda.

Here's the twist. What if you are told that it rained once when James won and once when Niki won and it is definite that it will rain on the next date. So, who would you bet your money on now ?

By intuition, it is easy to see that chances of winning for James have increased drastically. But the question is: how much ?

To understand the problem at hand, we need to become familiar with some concepts, first of which is conditional probability (explained below).

In addition, there are certain pre-requisites:

Pre-Requisites:



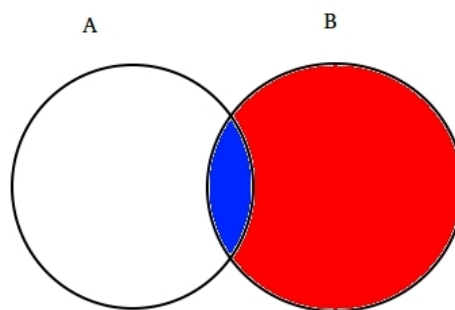
1. Linear Algebra : To refresh your basics, you can check out Khan's Academy Algebra (<https://www.khanacademy.org/math/linear-algebra>).
2. Probability and Basic Statistics : To refresh your basics, you can check out another course (<https://www.khanacademy.org/math/probability>), by Khan Academy.

3.1 Conditional Probability

It is defined as the: Probability of an event A given B equals the probability of B and A happening together divided by the probability of B."

For example: Assume two partially intersecting sets A and B as shown below.

Set A represents one set of events and Set B represents another. We wish to calculate the probability of A given B has already happened. Lets represent the happening of event B by shading it with red.



Now since B has happened, the part which now matters for A is the part shaded in blue which is interestingly $A \cap B$. So, the probability of A given B turns out to be:

$$\frac{BlueArea}{RedArea + BlueArea}$$



Therefore, we can write the formula for event B given A has already occurred by:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

or

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Now, the second equation can be rewritten as :

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

This is known as **Conditional Probability**.

Let's try to answer a betting problem with this technique.

Suppose, B be the *event of winning of James Hunt*. A be the *event of raining*. Therefore,

1. $P(A) = 1/2$, since it rained twice out of four days.
2. $P(B)$ is $1/4$, since James won only one race out of four.
3. $P(A|B) = 1$, since it rained every time when James won.

Substituting the values in the conditional probability formula, we get the probability to be around 50%, which is almost the double of 25% when rain was not taken into account (Solve it at your end).

This further strengthened our belief of James winning in the light of new *evidence* i.e rain. You must be wondering that this formula bears close resemblance to something you might have heard a lot about. Think!

Probably, you guessed it right. It looks like **Bayes Theorem**.

Bayes theorem is built on top of conditional probability and lies in the heart of Bayesian Inference. Let's understand it in detail now.

3.2 Bayes Theorem

Bayes Theorem comes into effect when multiple events A_i form an exhaustive set with another event B. This could be understood with the help of the below diagram.

A1	B	
A2		
A3		

Now, B can be written as

$$B = \sum_{i=1}^n B \cap A_i$$

So, probability of B can be written as,

$$P(B) = \sum_{i=1}^n P(B \cap A_i)$$



But

$$P(B \cap A_i) = P(B|A_i) \times P(A_i)$$

So, replacing P(B) in the equation of conditional probability we get

$$P(A_i|B) = (P(B|A_i) \times P(A_i)) / \left(\sum_{i=1}^n (P(B|A_i) \times P(A_i)) \right)$$

This is the equation of **Bayes Theorem**.

4. Bayesian Inference

There is no point in diving into the theoretical aspect of it. So, we'll learn how it works! Let's take an example of coin tossing to understand the idea behind *bayesian inference*.

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. For example, in tossing a coin, **fairness of coin** may be defined as the parameter of coin denoted by θ . The outcome of the events may be denoted by D.

Answer this now. What is the probability of 4 heads out of 9 tosses(D) given the fairness of coin (θ). i.e $P(D|\theta)$



Wait, did I ask the right question? No.

We should be more interested in knowing : Given an outcome (D) what is the probability of coin being fair ($\theta=0.5$)

Lets represent it using Bayes Theorem:

$$P(\theta|D) = (P(D|\theta) \times P(\theta)) / P(D)$$

Here, $P(\theta)$ is the **prior** i.e the strength of our belief in the fairness of coin before the toss. It is perfectly okay to believe that coin can have any degree of fairness between 0 and 1.

$P(D|\theta)$ is the likelihood of observing our result given our distribution for θ . If we knew that coin was fair, this gives the probability of observing the number of heads in a particular number of flips.

$P(D)$ is the evidence. This is the probability of data as determined by summing (or integrating) across all possible values of θ , weighted by how strongly we believe in those particular values of θ .

If we had multiple views of what the fairness of the coin is (but didn't know for sure), then this tells us the probability of seeing a certain sequence of flips for all possibilities of our belief in the coin's fairness.

$P(\theta|D)$ is the posterior belief of our parameters after observing the evidence i.e the number of heads .

From here, we'll dive deeper into mathematical implications of this concept. Don't worry. Once you understand them, getting to its *mathematics* is pretty easy.

To define our model correctly , we need two mathematical models before hand. One to represent the **likelihood function $P(D|\theta)$** and the other for representing the distribution of **prior beliefs** . The product of these two gives the **posterior belief $P(\theta|D)$** distribution.

Since prior and posterior are both beliefs about the distribution of fairness of coin, intuition tells us that both should have the same mathematical form. Keep this in mind. We will come back to it again.



So, there are several functions which support the existence of bayes theorem. Knowing them is important, hence I have explained them in detail.

4.1. Bernoulli likelihood function

Lets recap what we learned about the likelihood function. So, we learned that:

It is the probability of observing a particular number of heads in a particular number of flips for a given fairness of coin. This means our probability of observing heads/tails depends upon the fairness of coin (θ).

$P(y=1|\theta) = \theta^y$ [If coin is fair $\theta=0.5$, probability of observing heads ($y=1$) is 0.5]

$P(y=0|\theta) = (1 - \theta)^{1-y}$ [If coin is fair $\theta=0.5$, probability of observing tails($y=0$) is 0.5]

It is worth noticing that representing 1 as heads and 0 as tails is just a mathematical notation to formulate a model. We can combine the above mathematical definitions into a single definition to represent the probability of both the outcomes.

$$P(y|\theta) = \theta^y \cdot (1 - \theta)^{1-y}$$

This is called the **Bernoulli Likelihood Function** and the task of coin flipping is called Bernoulli's trials.

$$y = \{0, 1\}, \theta = (0, 1)$$

And, when we want to see a series of heads or flips, its probability is given by:



$$P(y_1, y_2, \dots, y_n | \theta) = \prod_1^n P(y_i | \theta)$$

$$P(y_1, y_2, \dots, y_n | \theta) = \prod_1^n \theta^{y_i} \cdot (1 - \theta)^{1-y_i}$$

Furthermore, if we are interested in the probability of number of heads z turning up in N number of flips then the probability is given by:

$$P(z, N | \theta) = \theta^z \cdot (1 - \theta)^{N-z}$$

4.2. Prior Belief Distribution

This distribution is used to represent our strengths on beliefs about the parameters based on the previous experience.

But, what if one has no previous experience?

Don't worry. Mathematicians have devised methods to mitigate this problem too. It is known as *uninformative priors*. I would like to inform you beforehand that it is just a misnomer. Every uninformative prior always provides some information even the constant distribution prior.

Well, the mathematical function used to represent the prior beliefs is known as *beta distribution*. It has some very nice mathematical properties which enable us to model our beliefs about a binomial distribution.

Probability density function of beta distribution is of the form :

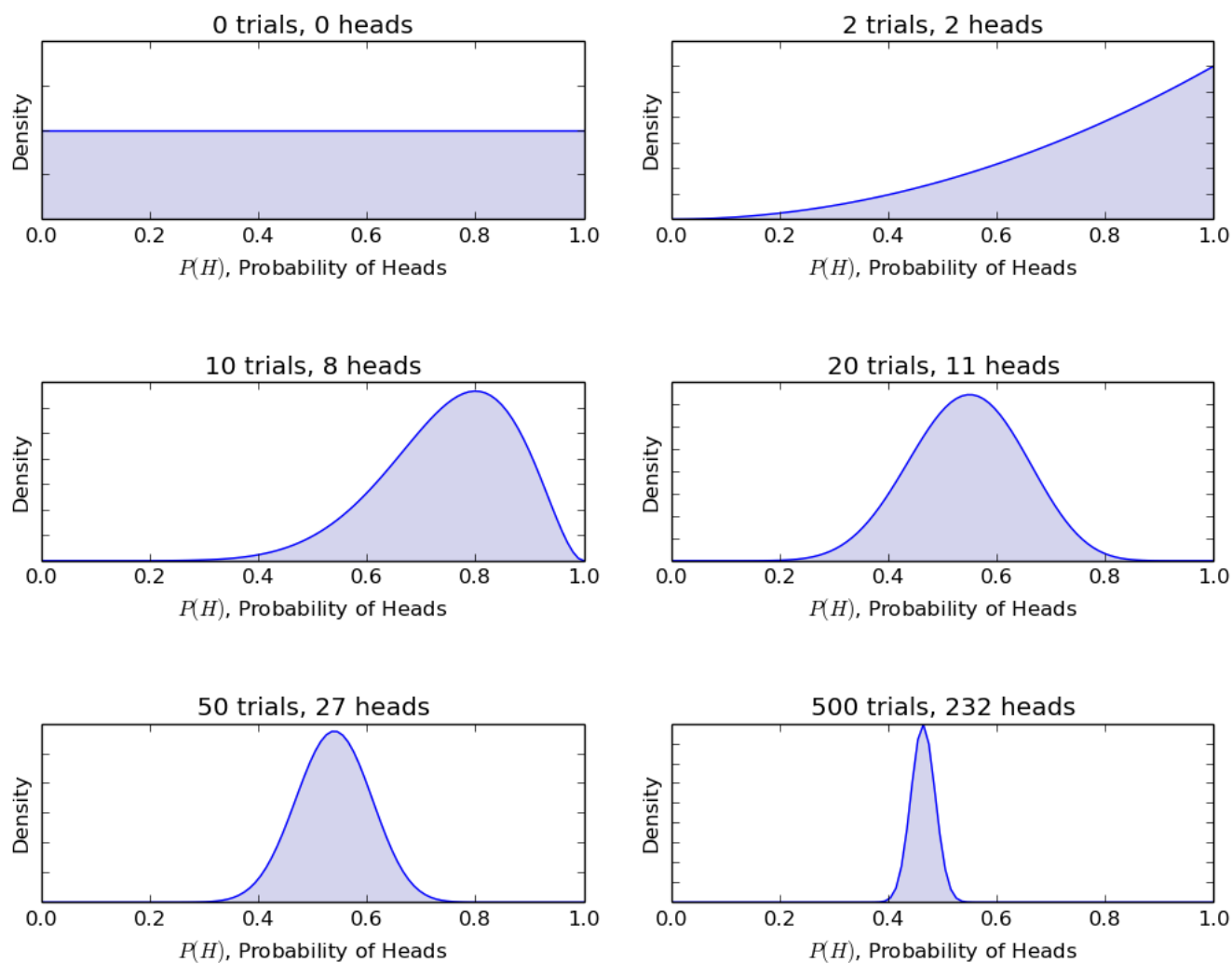


$$x^{\alpha-1} \cdot (1-x)^{\beta-1} / B(\alpha, \beta)$$

where, our focus stays on numerator. The denominator is there just to ensure that the total probability density function upon integration evaluates to 1.

α and β are called the shape deciding parameters of the density function. Here α is analogous to number of heads in the trials and β corresponds to the number of tails. The diagrams below will help you visualize the beta distributions for different values of α and β





You too can draw the beta distribution for yourself using the following code in R:



```

> library(stats)
> par(mfrow=c(3,2))
> x=seq(0,1,by=0.1)
> alpha=c(0,2,10,20,50,500)
> beta=c(0,2,8,11,27,232)
> for(i in 1:length(alpha)){
  y<-dbeta(x,shape1=alpha[i],shape2=beta[i])
  plot(x,y,type="l")
}

```

Note: α and β are intuitive to understand since they can be calculated by knowing the mean (μ) and standard deviation (σ) of the distribution. In fact, they are related as :

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}}$$

If mean and standard deviation of a distribution are known , then there shape parameters can be easily calculated.

Inference drawn from graphs above:

1. When there was no toss we believed that every fairness of coin is possible as depicted by the flat line.
2. When there were more number of heads than the tails, the graph showed a peak shifted towards the right side, indicating higher probability of heads and that coin is not fair.
3. As more tosses are done, and heads continue to come in larger proportion the peak narrows[^] increasing our confidence in the fairness of the coin value.

4.3. Posterior Belief Distribution

The reason that we chose prior belief is to obtain a beta distribution. This is because when we multiply it with a likelihood function, posterior distribution yields a form similar to the prior distribution which is much easier to relate to and understand. If this much information whets your appetite, I'm sure you are ready to walk an extra mile.

Let's calculate posterior belief using bayes theorem.

Calculating posterior belief using Bayes Theorem

$$\begin{aligned} P(\theta|z, N) &= P(z, N|\theta)P(\theta)/P(z, N) \\ &= \theta^z(1 - \theta)^{N-z}.\theta^{\alpha-1}(1 - \theta)^{\beta-1}/[B(\alpha, \beta)P(z, N)] \\ &= \theta^{z+\alpha-1}(1 - \theta)^{N-z+\beta-1}/[B(z + \alpha, N - z + \beta)] \end{aligned}$$

Now, our posterior belief becomes,

$$P(\theta|z + \alpha, N - z + \beta)$$

This is interesting. Just knowing the mean and standard distribution of our belief about the parameter θ and by observing the number of heads in N flips, we can update our belief about the model parameter(θ).

Lets understand this with the help of a simple example:

Suppose, you think that a coin is biased. It has a mean (μ) bias of around 0.6 with standard deviation of 0.1.

Then ,


$$\alpha = 13.8, \beta = 9.2$$

i.e our distribution will be biased on the right side. Suppose, you observed 80 heads ($z=80$) in 100 flips ($N=100$). Let's see how our prior and posterior beliefs are going to look:

$$\text{prior} = P(\theta | \alpha, \beta) = P(\theta | 13.8, 9.2)$$

$$\text{Posterior} = P(\theta | z + \alpha, N - z + \beta) = P(\theta | 93.8, 29.2)$$

Lets visualize both the beliefs on a graph:

 prior, posterior

The R code for the above graph is as:

```
> library(stats)
> x=seq(0,1,by=0.1)
> alpha=c(13.8,93.8)
> beta=c(9.2,29.2)
> for(i in 1:length(alpha)){
  y<-dbeta(x,shape1=alpha[i],shape2=beta[i])
  plot(x,y,type="l",xlab = "theta",ylab = "density")
}
```

As more and more flips are made and new data is observed, our beliefs get updated. This is the ~~real~~ power of Bayesian Inference.

5. Test for Significance – Frequentist vs Bayesian

Without going into the rigorous mathematical structures, this section will provide you a quick overview of different approaches of frequentist and bayesian methods to test for significance and difference between groups and which method is most reliable.

5.1. p-value

In this, the t-score for a particular sample from a sampling distribution of *fixed size* is calculated. Then, p-values are predicted. We can interpret p values as (taking an example of p-value as 0.02 for a distribution of mean 100) : There is 2% probability that the sample will have mean equal to 100.

This interpretation suffers from the flaw that for sampling distributions of different sizes, one is bound to get different t-score and hence different p-value. It is completely absurd. A p-value less than 5% does not guarantee that null hypothesis is wrong nor a p-value greater than 5% ensures that null hypothesis is right.

5.2. Confidence Intervals

Confidence Intervals also suffer from the same defect. Moreover since C.I is not a probability distribution , there is no way to know which values are most probable.



5.3. Bayes Factor

Bayes factor is the equivalent of p-value in the bayesian framework. Lets understand it in an comprehensive manner.

The *null hypothesis* in bayesian framework assumes ∞ probability distribution only at a particular value of a parameter (say $\theta=0.5$) and a zero probability else where. (M1)

The *alternative hypothesis* is that all values of θ are possible, hence a flat curve representing the distribution. (M2)

Now, posterior distribution of the new data looks like below.



significance, bayes factor, hypothesis testing

Bayesian statistics adjusted credibility (probability) of various values of θ . It can be easily seen that the probability distribution has shifted towards M2 with a value higher than M1 i.e M2 is more likely to happen.

Bayes factor does not depend upon the actual distribution values of θ but the magnitude of shift in values of M1 and M2.

In panel A (shown above): left bar (M1) is the prior probability of the null hypothesis.

In panel B (shown), the left bar is the posterior probability of the null hypothesis.

Bayes factor is defined as the ratio of the posterior odds to the prior odds,

$$BF = \frac{P(M = null|z, N)}{P(M = alt|z, N)} / \frac{P(M = null)}{P(M = alt)}$$

To reject a null hypothesis, a BF <1/10 is preferred.



We can see the immediate benefits of using Bayes Factor instead of p-values since they are independent of intentions and sample size.

5.4. High Density Interval (HDI)

HDI is formed from the posterior distribution after observing the new data. Since HDI is a probability, the 95% HDI gives the 95% most credible values. It is also guaranteed that 95 % values will lie in this interval unlike C.I.

Notice, how the 95% HDI in prior distribution is wider than the 95% posterior distribution. This is because our belief in HDI increases upon observation of new data.




significance, hdi, high density interval

End Notes

The aim of this article was to get you thinking about the different type of statistical philosophies out there and how any single of them cannot be used in every situation.

It's a high time that both the philosophies are merged to mitigate the real world problems by addressing the flaws of the other. Part II of this series will focus on the Dimensionality Reduction techniques using MCMC (Markov Chain Monte Carlo) algorithms. Part III will be based on creating a Bayesian regression model from scratch and interpreting its results in R. So, before I start with Part II, I would like to have your suggestions / feedback on this article.

Did you like reading this article ? As a beginner, were you able to understand the concepts? Let me know in comments. 

You can test your skills and knowledge. Check out Live Competitions (<http://datahack.analyticsvidhya.com/contest/all>) and compete with best Data Scientists from all over the world.

You can also read this article on Analytics Vidhya's Android APP



(https://play.google.com/store/apps/details?id=com.analyticsvidhya.android&utm_source=blog_article&utm_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1)


Share this:

 (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/?share=linkedin&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/?share=facebook&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/?share=twitter&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/?share=pocket&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/?share=reddit&nb=1>)

TAGS : BAYES INFERENCE ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BAYES-INFERENCE/](https://www.analyticsvidhya.com/blog/tag/bayes-inference/)), BAYES THEOREM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BAYES-THEOREM/](https://www.analyticsvidhya.com/blog/tag/bayes-theorem/)), BAYESIAN ANALYSIS

[\(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BAYESIAN-ANALYSIS/\)](https://www.analyticsvidhya.com/blog/tag/bayesian-analysis/), [BAYESIAN STATISTICS \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BAYESIAN-STATISTICS/\)](https://www.analyticsvidhya.com/blog/tag/bayesian-statistics/), [CONDITIONAL PROBABILITY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/CONDITIONAL-PROBABILITY/\)](https://www.analyticsvidhya.com/blog/tag/conditional-probability/), [FREQUENTIST \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/FREQUENTIST/\)](https://www.analyticsvidhya.com/blog/tag/frequentist/), [P VALUE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/P-VALUE/\)](https://www.analyticsvidhya.com/blog/tag/p-value/), [PDF \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/PDF/\)](https://www.analyticsvidhya.com/blog/tag/pdf/), [PROBABILITY DISTRIBUTION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/PROBABILITY-DISTRIBUTION/\)](https://www.analyticsvidhya.com/blog/tag/probability-distribution/), [STATISTICS \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/STATISTICS/\)](https://www.analyticsvidhya.com/blog/tag/statistics/)

NEXT ARTICLE

Operations analytics case study (level : hard)

(<https://www.analyticsvidhya.com/blog/2016/06/operations-analytics-case-study-level-hard/>)

...

PREVIOUS ARTICLE

Web Analytics – Bangalore (5 -7 years of experience)

(<https://www.analyticsvidhya.com/blog/2016/06/web-analytics-bangalore-5-7-years-experience/>)



(<https://www.analyticsvidhya.com/blog/author/nss/>)



NSS (<https://www.analyticsvidhya.com/blog/author/nss/>).

I am a perpetual, quick learner and keen to explore the realm of Data analytics and science. I am deeply excited about the times we live in and the rate at which data is being generated and being transformed as an asset. I am well versed with a few tools for dealing with data and also in the process of learning some other tools and knowledge required to exploit data.

[in_\(https://in.linkedin.com/in/neeraj-singh-sarwan-a84b6965\)](https://in.linkedin.com/in/neeraj-singh-sarwan-a84b6965)

This article is quite old and you might not get a prompt response from the author. We request you to post this comment on Analytics Vidhya's **Discussion portal** (<https://discuss.analyticsvidhya.com/>) to get your queries resolved

32 COMMENTS



ALEX

[Reply](#)

June 20, 2016 at 6:47 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112415>).

Thx for this great explanation.

I'm a beginner in statistics and data science and I really appreciate it.

If you're interested to see another approach, how toddler's brain use Bayesian statistics in a natural way there is a few easy-to-understand neuroscience courses : http://www.college-de-france.fr/site/en-stanislas-dehaene/_course.htm (http://www.college-de-france.fr/site/en-stanislas-dehaene/_course.htm).



NIKHIL

[Reply](#)

June 20, 2016 at 8:55 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112431>).

Hey one question `difference` -> $0.5 * (\text{No. of tosses}) - \text{no. of heads}$ is it correct?



NSS

[Reply](#)

June 23, 2016 at 8:04 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112573>).

@Nikhil ...Thanks for bringing it to the notice. It should be $\text{no. of heads} - 0.5(\text{No. of tosses})$.



CICEK

[Reply](#)

June 20, 2016 at 12:44 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112440>).

Did you miss the index i of A in the general formula of the Bayes' theorem on the left hand side of the equation (section 3.2)?



NSS

[Reply](#)

June 23, 2016 at 8:07 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112574>).

No, I didn't. 😊



SHARAN LOBANA

[Reply](#)

September 21, 2016 at 1:05 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-116263>).

Are you sure you the 'i' in the subscript of the final equation of section 3.2 isn't required. I think it should be A instead of Ai on the right hand side numerator.

**NSS**[Reply](#)

June 28, 2017 at 10:37 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-131232>).

Yes, It is required. I have made the necessary changes.

**ALPER**[Reply](#)

June 16, 2017 at 6:00 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-130604>).

cicek: i also think the index i is missing in LHS of the general formula in subsection 3.2 (the last equation in that subsection).

**NSS**[Reply](#)

June 28, 2017 at 10:38 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-131233>).

Yes, it has been updated. Thanks for pointing out.

**BHARATH**[Reply](#)

June 20, 2016 at 2:54 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112444>).



Nice visual to represent Bayes theorem, thanks

**NSS**[Reply](#)

June 23, 2016 at 8:09 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112575>).

Thanks Bharath..... It keeps us motivated.

**ANJA REBBER**[Reply](#)

June 20, 2016 at 8:01 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112452>).

I will let you know tomorrow! Because tomorrow I have to do teaching assistance in a class on Bayesian statistics. I will try to explain it your way, then I tell you how it worked out.
For me it looks perfect! Thanks!

**MELISSA L MOSER**[Reply](#)

June 21, 2016 at 6:01 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112471>).

Thorough and easy to understand synopsis. Good stuff. Thanks.

**GAURAV**[Reply](#)

June 21, 2016 at 9:25 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112479>).

It was a really nice article, with nice flow to compare frequentist vs bayesian approach. I will look forward to next part of the tutorials.



**SACHIN HOSAMANI**[Reply](#)

June 21, 2016 at 9:28 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112480>).

Excellent article. I didn't know much about Bayesian statistics, however this article helped me improve my understanding of Bayesian statistics.

**FANG XIANFU**[Reply](#)

June 21, 2016 at 1:31 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112492>).

Without wanting to suggest that one approach or the other is better, I don't think this article fulfilled its objective of communicating in "simple English".

The communication of the ideas was fine enough, but if the focus is to be on "simple English" then I think that the terminology needs to be introduced with more care, and mathematical explanations should be limited and vigorously explained.

**DAVIDE CALDARA**[Reply](#)

June 21, 2016 at 4:24 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112497>).

It's a good article.

As a beginner I have a few difficulties with the last part (chapter 5) but the previous parts were really good



**SCOTT**[Reply](#)

June 22, 2016 at 1:48 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112513>).

Very nice refresher. Thank you and keep them coming.

**SAILESH**[Reply](#)

June 22, 2016 at 2:38 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112514>).

Thank you, NSS for this wonderful introduction to Bayesian statistics. The visualizations were just perfect to establish the concepts discussed. Although I lost my way a little towards the end (Bayesian factor), appreciate your effort!

**ROEL**[Reply](#)

June 22, 2016 at 9:46 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112532>).

this 'stopping intention' is not a regular thing in frequentist statistics. In fact I only hear about it today. It sort of distracts me from the bayesian thing that is the real topic of this post. Perhaps you never worked with frequentist statistics?

**NSS**[Reply](#)

June 23, 2016 at 8:23 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112577>).

@Roel



Irregularities is what we care about ? Isn't it ? and well, stopping intentions do play a role. What if as a simple example: person A performs hypothesis testing for coin toss based on total flips and person B based on time duration . Do we expect to see the same result in both the cases ?

**ROEL**[Reply](#)

June 22, 2016 at 9:59 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112533>).

Some small notes, but let me make this clear: I think bayesian statistics makes often much more sense, but I would love it if you at least make the description of the frequentist statistics correct. Also let's not make this a debate about which is better, it's as useless as the python vs r debate, there is none.

"In this, the t-score for a particular sample from a sampling distribution of fixed size is calculated. Then, p-values are predicted. We can interpret p values as (taking an example of p-value as 0.02 for a distribution of mean 100) : There is 2% probability that the sample will have mean equal to 100."

This is incorrect. a p-value says something about the population. You inference about the population based on a sample. If mean 100 in the sample has p-value 0.02 this means the probability to see this value in the population under the nul-hypothesis is .02. Which makes it more likely that your alternative hypothesis is true.

"sampling distributions of different sizes, one is bound to get different t-score and hence different p-value. It is completely absurd."

correct it is an estimation, and you correct for the uncertainty in

I know it makes no sense, we test for an effect by looking at the probability of a score when there is no effect. if that is a small change we say that the alternative is more likely.



**NSS**[Reply](#)

June 23, 2016 at 8:17 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112576>).

@Roel

I agree this post isn't about the debate on which is better- Bayesian or Frequentist.

And I quote again- "The aim of this article was to get you thinking about the different type of statistical philosophies out there and how any single of them cannot be used in every situation".

Regarding p-value , what you said is correct- Given your hypothesis, the probability.....

But generally, what people infer is – the probability of your hypothesis,given the p-value.....

But, still p-value is not the robust mean to validate hypothesis, I feel.

I would like to hear more.

Thanks for commenting. 😊

**JOSÉ AVILA**[Reply](#)

June 23, 2016 at 4:18 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112601>).

This is a really good post! Thanks for share this information in a simple way!

I have some questions that I would like to ask!

1) I didn't understand very well why the C.I. "do not provide the most probable value for a parameter and the most probable values". Before to read this post I was thinking in this way: the real mean of population is between the range given by the CI with a, for example, 95%)



2) I read a recent paper which states that rejecting the null hypothesis by bayes factor at $<1/10$ could be equivalent as assuming a p value <0.001 for reject the null hypothesis (actually, I don't remember very well the exact values, but the idea of making this equivalence is correct? could be good to apply this equivalence in research?)

3) For making bayesian statistics, is better to use R or Phyton? or it depends on each person? Now I m learning Phyton because I want to apply it to my research (I m biologist!)

Thanks in advance and sorry for my not so good english!

Jose Avila



SHANE

[Reply](#)

June 24, 2016 at 12:38 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112641>).

Hi NSS,

A quick question about section 4.2: If alpha = no. of heads and beta = no. of tail

Why the alpha value = the number of trails in the R code:

```
> alpha=c(0,2,10,20,50,500) # it looks like the total number of trails, instead of number of heads....
```

```
> beta=c(0,2,8,11,27,232)
```

I plotted the graphs and the second one looks different from yours...

Thanks,

Shane



**AYUSH MEHTA**[Reply.](#)

June 26, 2016 at 9:52 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112700>).

How can I know when the other posts in this series are released?

**BOOMY**[Reply.](#)

June 26, 2016 at 4:36 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112714>).

Thank you for this Blog. I like it and I understand about concept Bayesian. I can practice in R and I can see something.

I think, you should write the next guide on Bayesian in the next time.
I will wait.

**ASANKA**[Reply.](#)

June 29, 2016 at 8:12 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-112825>).

HI...

Good post and keep it up ... very useful...

**NIKHIL**[Reply.](#)

August 4, 2016 at 7:11 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-114453>).



printer friendly version please!

**NISHTHA (HTTP://WWW.HOPEANDHAPPINESS.CO)**[Reply](#)

March 25, 2017 at 2:13 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-125531>).

Hi NSS

Thanks for the much needed comprehensive article. Please tell me a thing :-

"Since HDI is a probability, the 95% HDI gives the 95% most credible values. It is also guaranteed that 95 % values will lie in this interval unlike C.I."

How is this unlike CI? As far as I know CI is the exact same thing.

**NSS**[Reply](#)

March 25, 2017 at 3:45 pm (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-125535>).

@Nishtha CI is the probability of the intervals containing the population parameter i.e 95% CI would mean 95% of intervals would contain the population parameter whereas in HDI it is the presence of a population parameter in an interval with 95% probability. Both are different things. Hope this helps.

**STEV**[Reply](#)

June 19, 2017 at 9:22 am (<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/#comment-130723>).

Hi, greetings from Latam. I liked this. You've given us a good and simple explanation about Bayesian Statistics. Help me, I've not found the next parts yet.























(<https://www.analyticsvidhya.com/>)



Download App

(https://play.google.com/store/apps/details?

id=com.analyticsvidhya.android)



(https://apps.apple.com/us/app/analytics-vidhya/id1470025572)

Analytics Vidhya

About Us (https://www.analyticsvidhya.com/about-me/)

Our Team (https://www.analyticsvidhya.com/about-me/team/)

Careers (https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/)

Contact us (https://www.analyticsvidhya.com/contact/)

Data Science

Blog (https://www.analyticsvidhya.com/blog/)

Hackathon (https://datahack.analyticsvidhya.com/)

Discussions (https://discuss.analyticsvidhya.com/)

Apply Jobs (https://www.analyticsvidhya.com/jobs/)

Companies

Post Jobs (https://www.analyticsvidhya.com/corporate/)

Trainings (https://courses.analyticsvidhya.com/)

Hiring Hackathons (https://datahack.analyticsvidhya.com/)

Advertising (https://www.analyticsvidhya.com/contact/)

Visit us**in**

(https://www.linkedin.com/company/analytics-

(https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/)

© Copyright 2013-2020 Analytics Vidhya

[Privacy Policy](#) [Terms of Use](#) [Refund Policy](#)

x

-



(<http://play.google.com/store/apps/details?id=com.analyticsvidhya.android>)

