BAYESIAN INFERENCE

Amanpreet Singh Saini
2023MAS1001

Department of Mathematics
IIT ROPAR

Supervisor - **Dr. Deepak Singh**

September 24, 2024

Contents

BAYES' THEOREM

Bayes' Theorem

What exactly is Bayes' Theorem?

There are two ways to think about Bayes' Theorem:

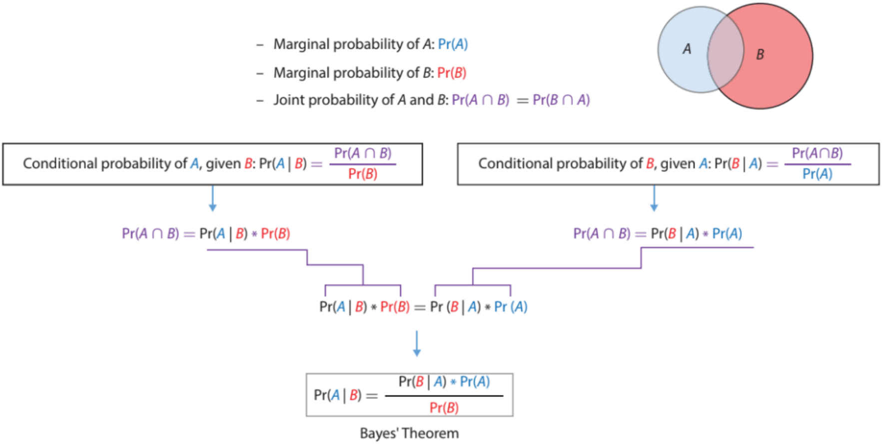
- It describes the relationship between $\Pr(A)$, $\Pr(B)$, $\Pr(A|B)$ and $\Pr(B|A)$
- It expresses how a subjective degree of belief should rationally change to account for evidence

First interpretation

Bayes' Theorem is a relationship between $\Pr(A)$, $\Pr(B)$, $\Pr(A|B)$ and $\Pr(B|A)$

First interpretation

Bayes' Theorem is a relationship between $\Pr(A)$, $\Pr(B)$, $\Pr(A|B)$ and $\Pr(B|A)$



EXAMPLE

Suppose we have one type of conditional probability, $\Pr(A|B)$, and we have to find the another, $\Pr(A|B)$.

	A	$\sim A$	Sum
B	0.05	0.10	0.15
$\sim B$	0.65	0.20	0.85
Sum	**0.70**	**0.30**	**1.00**

EXAMPLE

Suppose we have one type of conditional probability, $\Pr(A|B)$, and we have to find the another, $\Pr(A|B)$.

	A	$\sim A$	Sum
B	0.05	0.10	0.15
$\sim B$	0.65	0.20	0.85
Sum	**0.70**	**0.30**	**1.00**

Using Bayes' Theorem

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)} \tag{1}$$

$$\Pr(A|B) = \frac{\frac{0.05}{0.7} \times 0.7}{0.15} = 0.333 \tag{2}$$

EXAMPLE

Suppose we have one type of conditional probability, $\Pr(A|B)$, and we have to find the another, $\Pr(A|B)$.

	A	$\sim A$	Sum
B	0.05	0.10	0.15
$\sim B$	0.65	0.20	0.85
Sum	**0.70**	**0.30**	**1.00**

Using Bayes' Theorem

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)} \tag{1}$$

$$\Pr(A|B) = \frac{\frac{0.05}{0.7} \times 0.7}{0.15} = 0.333 \tag{2}$$

In this case:

$$\Pr(A|B) = \frac{0.05}{0.15} = 0.333 \tag{3}$$

Second Interpretation

The Theorem can express how a subjective degree of belief should rationally change to account for evidence.

Second Interpretation

The Theorem can express how a subjective degree of belief should rationally change to account for evidence.

Here's the theorem we'll be using for second interpretation:

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)} \tag{4}$$

Second Interpretation

The Theorem can express how a subjective degree of belief should rationally change to account for evidence.

Here's the theorem we'll be using for second interpretation:

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)} \tag{4}$$

Yes, the equation is the same,
but in this second interpretation, **Bayes' Theorem opens up entirely new possibilities for thinking about probabilities**

BAYESIAN INFERENCE

What is Bayesian inference?

Bayesian inference is the process of confronting alternative hypotheses with new data and using Bayes' Theorem to update your beliefs in each hypothesis.

What is Bayesian inference?

Bayesian inference is the process of confronting alternative hypotheses with new data and using Bayes' Theorem to update your beliefs in each hypothesis.

Bayesian inference is the use of Bayes' Theorem to draw conclusions about a set of **mutually exclusive, exhaustive**, alternative hypotheses by linking prior knowledge about each hypothesis with new data.

What is Bayesian inference?

Bayesian inference is the process of confronting alternative hypotheses with new data and using Bayes' Theorem to update your beliefs in each hypothesis.

Bayesian inference is the use of Bayes' Theorem to draw conclusions about a set of **mutually exclusive, exhaustive**, alternative hypotheses by linking prior knowledge about each hypothesis with new data.

$$\boxed{\text{Initial Belief in Hypothesis } i + \text{New Data} \rightarrow \text{Updated Belief in Hypothesis } i.}$$

Prior Probabilities

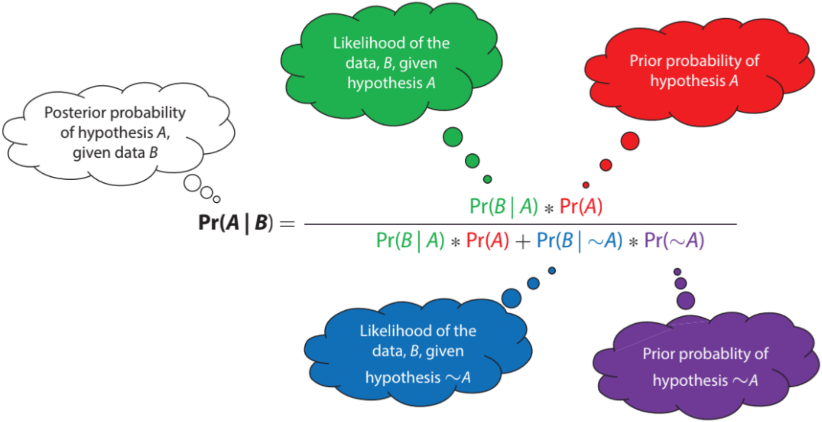
Represent our current belief in each hypothesis , **Prior to data collection**.

Likelihood

Probability of observing the data under each hypothesis.

Posterior Probability

Represents our updated belief in each hypothesis **after new data are collected.**



$$\Pr(A \mid B) = \frac{\Pr(B \mid A) * \Pr(A)}{\Pr(B \mid A) * \Pr(A) + \Pr(B \mid \sim A) * \Pr(\sim A)}$$

The Author Problem: Bayesian Inference with Two Hypotheses

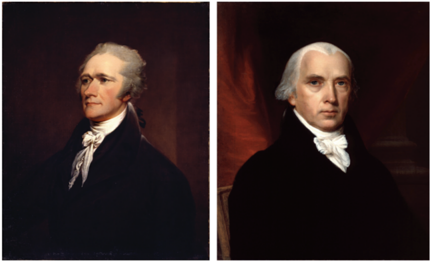


Figure: Alexander Hamilton (left) and James Madison (right).

The authorship of a paper is in dispute between Hamilton and Madison. We have to use a Bayesian analysis to rightfully attribute the authorship of the paper.

We'll use a series of steps to conduct our analysis,

1. Identify your hypotheses.
2. Express your belief that each hypothesis is true in terms of prior probabilities.
3. Gather the data.
4. Determine the likelihood of the observed data under each hypothesis.
5. Use Bayes' Theorem to compute the posterior probabilities for each hypothesis.

STEP 1

Identify our hypotheses

For this paper, we have two hypotheses for authors: Hamilton or Madison.

- Hamilton = Hamilton hypothesis
- Madison = Madison hypothesis

STEP 1

Identify our hypotheses

For this paper, we have two hypotheses for authors: Hamilton or Madison.

- Hamilton = Hamilton hypothesis
- Madison = Madison hypothesis

Note

- The hypotheses are exhaustive and mutually exclusive.
- Since there are only two hypotheses, it makes sense that:

$$\Pr(\text{Hamilton}) = \Pr(\neg\text{Madison}) \tag{5}$$

And similarly:

$$\Pr(\text{Madison}) = \Pr(\neg\text{Hamilton}) \tag{6}$$

STEP 2

Set Prior Probabilities

We express our belief that each hypothesis is true in terms of prior probabilities.

For example

- $\Pr(\text{Hamilton}) = 0.1$ and $\Pr(\text{Madison}) = 0.9$
- $\Pr(\text{Hamilton}) = 0.5$ and $\Pr(\text{Madison}) = 0.5$
- $\Pr(\text{Hamilton}) = 0.7$ and $\Pr(\text{Madison}) = 0.3$
- $\Pr(\text{Hamilton}) = 0.75$ and $\Pr(\text{Madison}) = 0.25$

- Suppose we set the odds at 50:50 ,

$$\Pr(\text{Hamilton}) = 0.5 \text{ and } \Pr(\text{Madison}) = 0.5$$

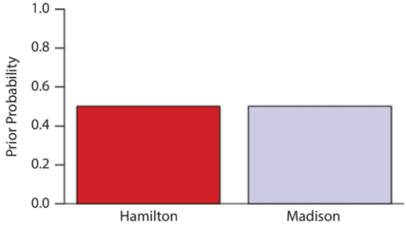


Figure: Prior distributions.

STEP 3

Gather the data.

We will calculate the frequency of the use of **'upon'** in the paper.

The word 'upon' appeared twice in the paper in question. The rate of 'upon' is calculated as:

$$\frac{\text{upons}}{\text{total words}} = \frac{2}{2008} = 0.000996 \tag{7}$$

Standardize this rate to 1000 words,

$$0.000996 \times 1000 = 0.996 \tag{8}$$

Data is given as

$$\text{Rate} = 0.996$$

STEP 4

Determine the **likelihood**

$$\Pr(\text{Data} \mid \text{Hamilton})$$

and

$$\Pr(\text{Data} \mid \text{Madison})$$

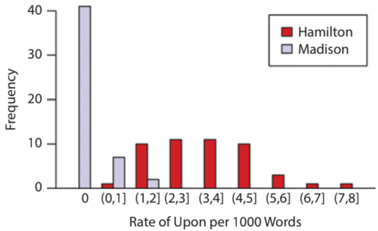
STEP 4

Suppose we calculate the standard rate of use the word 'upon' in each paper of the two authors .

The data might look like that shown in table.

Rate	Hamilton	Madison
0 (exactly)	0	41
(0,1]	1	7
(1,2]	10	2
(2,3]	11	0
(3,4]	11	0
(4,5]	10	0
(5,6]	3	0
(6,7]	1	0
(7,8]	1	0
	48	**50**

STEP 4

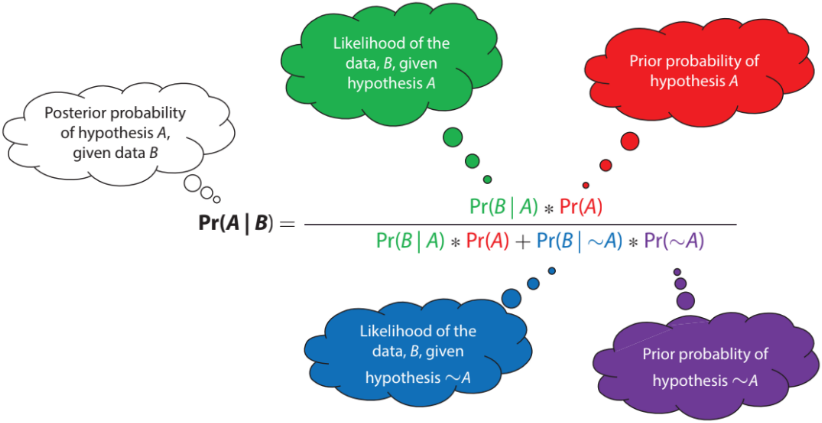


$$\Pr(\text{Data} \mid \text{Hamilton}) = \frac{1}{48} = 0.021 \tag{9}$$

$$\Pr(\text{Data} \mid \text{Madison}) = \frac{7}{50} = 0.140 \tag{10}$$

STEP 5

Use Bayes' Theorem to compute the updated Probabilities



$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \ast \Pr(A)}{\Pr(B \mid A) \ast \Pr(A) + \Pr(B \mid \sim A) \ast \Pr(\sim A)}$$

STEP 5

$$\Pr(\text{Hamilton} \mid \text{Data}) = \frac{0.021 \times 0.5}{0.021 \times 0.5 + 0.140 \times 0.5} = \frac{0.0105}{0.0805} = 0.1304 \tag{11}$$

The posterior probability that the author of the paper in question is Hamilton:

$$\Pr(\text{Hamilton} \mid \text{Data}) = 0.1304 \tag{12}$$

Since we have only two mutually exclusive hypotheses,

$$\Pr(\text{Madison} \mid \text{Data}) = 0.8696 \tag{13}$$

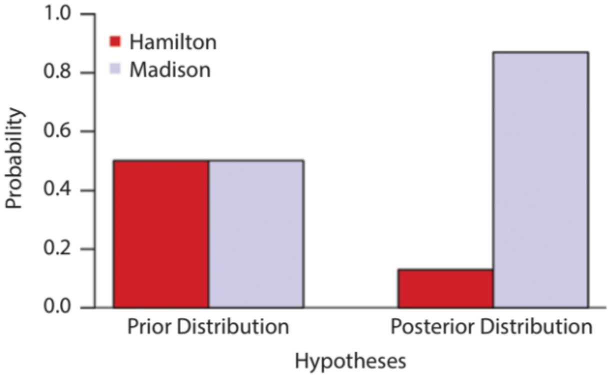


Figure: Prior and posterior distributions

QUESTION: How do we set the prior probabilities?

let's see what happens when we use a different set of priors instead:

$$\Pr(\text{Hamilton}) = 0.75 \text{ and } \Pr(\text{Madison}) = 0.25$$

QUESTION: How do we set the prior probabilities?

let's see what happens when we use a different set of priors instead:

$$\Pr(\text{Hamilton}) = 0.75 \text{ and } \Pr(\text{Madison}) = 0.25$$

$$\Pr(\text{Hamilton} \mid \text{Data}) = \frac{0.021 \times 0.75}{0.021 \times 0.75 + 0.140 \times 0.25} = \frac{0.01575}{0.05075} = 0.3103 \quad (14)$$

The answer now is **0.3103**, whereas our first result was **0.1304**.

QUESTION: How do we set the prior probabilities?

let's see what happens when we use a different set of priors instead:

$$\Pr(\text{Hamilton}) = 0.75 \text{ and } \Pr(\text{Madison}) = 0.25$$

$$\Pr(\text{Hamilton} \mid \text{Data}) = \frac{0.021 \times 0.75}{0.021 \times 0.75 + 0.140 \times 0.25} = \frac{0.01575}{0.05075} = 0.3103 \quad (14)$$

The answer now is **0.3103**, whereas our first result was **0.1304**.

NOTE

- In both cases, the analysis now suggests that there is a greater probability that James Madison was the author.
- However, the posterior probability for the Madison hypothesis is smaller when the odds are stacked against him in the second example.
  In other words, priors matter

QUESTION: Do the likelihoods of the data have to add to 1.0?

QUESTION: Do the likelihoods of the data have to add to 1.0?

Observed Rate	Hypothesis	Prior	Likelihood	Prior $*$ L	Posterior
0.996	Hamilton	0.5	0.0210	0.0105	0.1304
0.996	Madison	0.5	0.1400	0.0700	0.8696
	Sum	**1.0**	**0.1610**	**0.0805**	**1.0000**

Observed Rate	Hypothesis	Prior	Likelihood	Prior $*$ L	Posterior
0.996	Hamilton	0.75	0.0210	0.0158	0.3103
0.996	Madison	0.25	0.1400	0.0350	0.6897
	Sum	**1.00**	**0.1610**	**0.0508**	**1.0000**

QUESTION: Do the likelihoods of the data have to add to 1.0?

Observed Rate	Hypothesis	Prior	Likelihood	Prior $*$ L	Posterior
0.996	Hamilton	0.5	0.0210	0.0105	0.1304
0.996	Madison	0.5	0.1400	0.0700	0.8696
	Sum	**1.0**	**0.1610**	**0.0805**	**1.0000**

Observed Rate	Hypothesis	Prior	Likelihood	Prior $*$ L	Posterior
0.996	Hamilton	0.75	0.0210	0.0158	0.3103
0.996	Madison	0.25	0.1400	0.0350	0.6897
	Sum	**1.00**	**0.1610**	**0.0508**	**1.0000**

In Bayesian analysis, the sum of the prior and posterior probabilities across hypotheses, must be 1.0.
This is not true for the likelihoods of observing the data under each hypothesis.

DISCRETE HYPOTHESES

For n discrete hypotheses $(\mathrm{H}_i, i = 1, 2, 3, 4.......n)$

Bayes' Theorem can be expressed as:

$$\Pr(\mathrm{H}_i \mid \text{data}) = \frac{\Pr(\text{data} \mid \mathrm{H}_i) * \Pr(\mathrm{H}_i)}{\sum_{j=1}^{n} \Pr(\text{data} \mid \mathrm{H}_j) * \Pr(\mathrm{H}_j)} \tag{15}$$

Here both the **prior** $P(\theta)$ and the **posterior** $P(\theta \mid \text{data})$ distributions are **pmf**s.

CONTINUES HYPOTHESIS

When the hypotheses for a parameter θ are infinite.

Bayes' Theorem can be expressed as:

$$P(\theta \mid \text{data}) = \frac{P(\text{data} \mid \theta) * P(\theta)}{\int P(\text{data} \mid \theta) * P(\theta) d\theta} \tag{16}$$

Here both the **prior** $P(\theta)$ and the **posterior** $P(\theta \mid \text{data})$ distributions are **pdf**s.

BAYESIAN CONJUGATES

CONJUGATE PRIOR

A conjugate prior is a probability distribution that, when used with a likelihood function, results in a posterior distribution that is in the same family as the prior. This property is called **conjugacy**

CONJUGATE PRIOR

A conjugate prior is a probability distribution that, when used with a likelihood function, results in a posterior distribution that is in the same family as the prior. This property is called **conjugacy**

There are cases where we can use a particular distribution as a **prior distribution**, collect data of a **specific flavor**, and then derive the posterior distribution with a closed-form solution.

In these special cases, the prior and posterior probabilities follows the same distributions, but they may differ in parameters.

Some examples of Bayesian Conjugates

- beta prior $+$ binomial data $\rightarrow$ beta posterior
- gamma prior $+$ Poisson data $\rightarrow$ gamma posterior
- normal prior $+$ normal data $\rightarrow$ normal posterior

The White House Problem:
The Beta-Binomial Conjugate

Figure: Shaquille O'Neal

NBA star Shaquille O'Neal and a friend debated whether or not he could get into the White House without an appointment. The wager: 1000 push-ups.

QUESTION: What probability function would be appropriate for Shaq's bet?

Shaq is going to attempt to get into the White House, so that represents a **trial**. He will either **succeed** or **fail**. This is a **binomial problem**.

Binomial pmf:

$$f(y; n, p) = \binom{n}{y} p^y (1-p)^{(n-y)} \quad , y = 0, 1, \ldots, n \tag{17}$$

The number of trials is denoted n.
The number of observed successes is denoted as y.

If we knew that Shaq's probability of success is say, $\mathbf{0.7}$, and that Shaq had only one chance, If we knew that Shaq's probability of success is say, **0.7**, and that Shaq had only one chance,

The probability that Shaq would succeed is:

$$f(1;1,0.7) = \binom{1}{1} 0.7^1 (1-0.7)^{(1-1)} = 0.7 \tag{18}$$

The probability that Shaq would fail is:

$$f(0;1,0.7) = \binom{1}{0} 0.7^0 (1-0.7)^{(1-0)} = 0.3 \tag{19}$$

If we knew that Shaq's probability of success is say, **0.7**, and that Shaq had only one chance,

The probability that Shaq would succeed is:

$$f(1;1,0.7) = \binom{1}{1} 0.7^1 (1-0.7)^{(1-1)} = 0.7 \tag{18}$$

The probability that Shaq would fail is:

$$f(0;1,0.7) = \binom{1}{0} 0.7^0 (1-0.7)^{(1-0)} = 0.3 \tag{19}$$

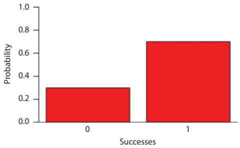


Figure: Binomial distribution: n = 1, p = 0.7

QUESTION: But what about Shaq's friend?

Shaq's friend, however, believes Shaq's chances are much lower than 0.7.
If p is, say **0.1**, then Shaq's chances would look much different.

QUESTION: But what about Shaq's friend?

Shaq's friend, however, believes Shaq's chances are much lower than 0.7.
If p is, say **0.1**, then Shaq's chances would look much different.

The probability that Shaq would succeed is:

$$f(1;1,0.1) = \binom{1}{1} 0.1^1 (1-0.1)^{(1-1)} = 0.1 \tag{20}$$

We don't know what **p** (the probability of success) is.
So, here we are confronted with a **parameter estimation problem**.

Our goal here is to use a Bayesian inference approach to estimate the probability that a celebrity can get into the White House without an invitation.
We'll use the same steps :

1. Identify your hypotheses— for p, ranging from 0 to 1.00.
2. Express your belief that each hypothesis is true in terms of prior densities.
3. Gather the data—Shaq makes his attempt, and will either fail or succeed.
4. Determine the likelihood of the observed data, assuming each hypothesis is true.
5. Use Bayes' Theorem to compute the posterior densities for each value of p (i.e., the posterior distribution).

Now, let's go through the steps one at a time.

Step 1. What are the hypotheses for p?

We know that **p**, the probability of success in the binomial distribution, can take on any value between 0 and 1.0.

$\Longrightarrow$ We have the full range of hypotheses between 0 and 1, which is infinite ($p = 0.01$, $p = 0.011$, $p = 0.0111$, etc.).

Step 2. What are the prior densities for these hypotheses?

We need to assign a prior for each hypothesized value of p.
Here, we will use the **beta distribution** to set prior probabilities for each and every hypothesis for p.

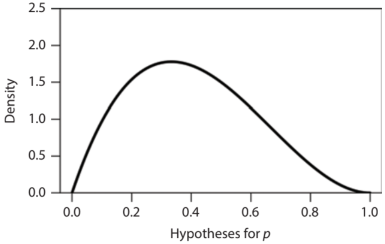


Figure: Example of a beta probability distribution

QUESTION: What is the beta distribution?

In probability theory and statistics,

The beta distribution is a family of continuous probability distributions defined on the interval (0, 1) parameterized by two positive shape parameters, typically denoted by alpha (α) and beta (β).

$$X \sim \text{beta}(\alpha, \beta)$$

QUESTION: What is the beta distribution?

In probability theory and statistics,

The beta distribution is a family of continuous probability distributions defined on the interval (0, 1) parameterized by two positive shape parameters, typically denoted by alpha (α) and beta (β).

$$X \sim \mathrm{beta}(\alpha, \beta)$$

The beta probability density function

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1. \tag{21}$$

where beta function, B, is a normalization constant to ensure that the area under the curve is 1.

QUESTION: How to fix α and β??

$$\alpha = \frac{\beta * \mu}{1-\mu}, \beta = \mu - 1 + \frac{\mu * (1-\mu)^2}{\sigma^2} \tag{22}$$

where μ is the mean and σ^2 is the variance

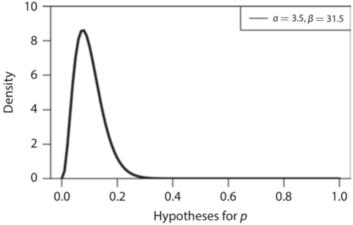
QUESTION: How to fix α and β??

$$\alpha = \frac{\beta * \mu}{1-\mu}, \beta = \mu - 1 + \frac{\mu * (1-\mu)^2}{\sigma^2} \tag{22}$$

where μ is the mean and σ^2 is the variance

KEY POINTS

- The bigger α is relative to β, shift the weight of the curve to the right
- The bigger β is relative to α, shift the weight of the curve to the left.



QUESTION: What prior distribution did Shaq and his friend settle on?

Let's assume they go with a beta distribution with α_0 and β_0 set to 0.5. As p is either really high or really low.

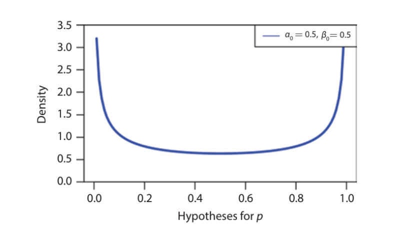


Figure: Prior distribution.

STEP 3 - Collect data

Let's assume that Shaq makes 1 attempt and fails to get in.

In the Binomial function terms,
the number of trials $\mathbf{n = 1}$ and the number of successes $\mathbf{y = 0}$

STEP 4 - Determine the likelihood

We'll determine the likelihood of the observed data, assuming each hypothesis is true.

Now, for each hypothesized value of p, let's compute the binomial likelihood of observing 0 successes out of 1 trial.

$$\mathcal{L}(y; n, p) = \binom{n}{y} p^y (1-p)^{(n-y)} \tag{23}$$

But Because p is a continuous variable between 0 and 1,
we have an infinite number of hypotheses!
(Bayesian inference problem for a **Continuous random variable.)**

Bayes' Theorem in this case is specified as:

$$P(p \mid \text{ data}) = \frac{P(\text{data} \mid p) * P(p)}{\int_0^1 P(\text{data} \mid p) * P(p) dp} \tag{24}$$

But here's the kicker: **The integration of the denominator is often tedious, and sometimes impossible!**

QUESTION: How do we make headway?

QUESTION: How do we make headway?

Start INTEGRATING!

QUESTION: How do we make headway?

Start INTEGRATING!

For this particular problem,
there is an analytical shortcut that makes updating possible in a snap.

The Posterior probability follows the beta distribution with parameters

$$\begin{aligned} \alpha_{\text{posterior}} &= \alpha_0 + y \\ \beta_{\text{posterior}} &= \beta_0 + n - y \end{aligned} \tag{25}$$

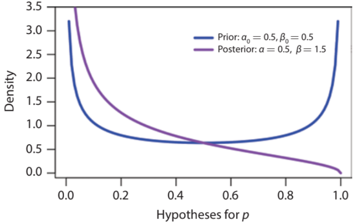
n is the number of trials (**n = 1**),
y is the number of successes (**y = 0**)

Here our prior distribution is a beta distribution with $\alpha_0 = 0.5$ and $\beta_0 = 0.5$
Shaq made an attempt, so n $=$1.
He failed to get into the White house, so y $=$ 0.

$$\begin{aligned} \alpha_{\text{posterior}} &= \alpha_0 + y = 0.5 + 0 = 0.5 \\ \beta_{\text{posterior}} &= \beta_0 + n - y = 0.5 + 1 - 0 = 1.5 \end{aligned} \tag{26}$$

Here our prior distribution is a beta distribution with $\alpha_0 = 0.5$ and $\beta_0 = 0.5$
Shaq made an attempt, so n =1.
He failed to get into the White house, so y = 0.

$$\begin{aligned} \alpha_{\text{posterior}} &= \alpha_0 + y = 0.5 + 0 = 0.5 \\ \beta_{\text{posterior}} &= \beta_0 + n - y = 0.5 + 1 - 0 = 1.5 \end{aligned} \tag{26}$$



QUESTION: What if Shaq makes a second attempt?

We now set a prior distribution based on our updated knowledge,

$$\begin{aligned} \alpha_0 &= 0.5 \\ \beta_0 &= 1.5 \end{aligned} \tag{27}$$

and then collect more data.

Suppose Shaq fails again.

QUESTION: What if Shaq makes a second attempt?

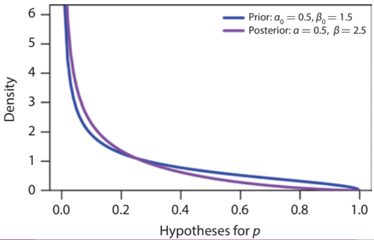
We now set a prior distribution based on our updated knowledge,

$$\begin{aligned} \alpha_0 &= 0.5 \\ \beta_0 &= 1.5 \end{aligned} \tag{27}$$

and then collect more data.

Suppose Shaq fails again.

$$\begin{aligned} \alpha_{\text{posterior}} &= \alpha_0 + y = 0.5 + 0 = 0.5 \\ \beta_{\text{posterior}} &= \beta_0 + n - y = 1.5 + 1 - 0 = 2.5 \end{aligned} \tag{28}$$



QUESTION: Could he have just tried twice before we updated the prior?

QUESTION: Could he have just tried twice before we updated the prior?

- The prior:

$$\begin{aligned} \alpha_0 &= 0.5 \\ \beta_0 &= 0.5 \end{aligned} \tag{29}$$

QUESTION: Could he have just tried twice before we updated the prior?

- The prior:

$$\begin{aligned} \alpha_0 &= 0.5 \\ \beta_0 &= 0.5 \end{aligned} \tag{29}$$

- The data:
  - $n = 2$
  - $y = 0$

QUESTION: Could he have just tried twice before we updated the prior?

- The prior:

$$\begin{aligned} \alpha_0 &= 0.5 \\ \beta_0 &= 0.5 \end{aligned} \tag{29}$$

- The data:
  - n = 2
  - y = 0
- The posterior:

$$\begin{aligned} \alpha_{\text{posterior}} &= \alpha_0 + y = 0.5 + 0 = 0.5 \\ \beta_{\text{posterior}} &= \beta_0 + n - y = 0.5 + 2 - 0 = 2.5 \end{aligned} \tag{30}$$

The same answer we got before.

NOTE-

- The beta distribution is a suitable model for the random behavior of percentages and proportions. These distributions all depend on a parameter, p that can assume values between 0 and 1.
- In Bayesian inference, the beta distribution is the conjugate prior probability distribution for the Bernoulli, binomial, negative binomial and geometric distributions.

ADVANTAGES

ADVANTAGES OF BAYESIAN INFERENCE

- **Incorporates prior information**
  Bayesian analysis allows you to combine prior information with data, which can be useful when dealing with limited data.
- **More accurate estimates**
  Bayesian analysis can provide more accurate estimates of parameters, especially when the number of observations is small
- **Flexibility**
  Bayesian inference allows for a wider range of models, including hierarchical models and models that account for missing
- **Hypothesis testing**
  Bayesian inference allows for the quantification of evidence and tracking of its progression as new data is collected

References

- **Books**
  1. Donovan, T. M., Mickey, R. M. (2019). Bayesian Statistics for Beginners
  2. Winkler, R. L., Hays, W. L. (1975). Statistics: Probability, inference, and decision .
- **Youtube**
  1. 3Blue1Brown
  2. Learn Statistics with Brian
  3. Bary Van Veen


Thank You