Bayes' Theorem and Operations Research

Sybba[1], Priyanka Patel[2], Rahul Girdhar[3], Rahul Mehta[4], Ramneek Hora[5], Rhea Gupta[6] & Prof. Tehreem Bardi [7]

[1,2,3,4,5,6]BBA, SECOND YEAR, SEMESTER- III (2017-18)
Shri Vile ParleKelavani Mandal's
NMIMS Anil Surendra Modi School of Commerce
V. L. Mehta Road, Vile Parle (W), Mumbai - 400 056 INDIA

Abstract : The problem of decision making is often a part of the human activities. The major problem faced by humans is the problem of uncertainty. Uncertainty is mainly referring to inadequate amount of information but it may also occur at times where we have enough amount of information or uncertainty can even be increased or vanished if we add some new information to the existing one. This concept of uncertainty is faced every day in our day to day life. Many businesses find ways to tackle the uncertainty by different methods. One of the most successful methods of tackling uncertainty is by using Bayesian Theorem for conditional probability density function. This concept mainly focuses on collecting some previously existing data and then analyzing it to either find a pattern or to get an estimate of the current scenario. Using the Bayesian Theorem and using it to minimize uncertainty can be of great help to businesses.

Keywords : Uncertainty, Conditional Probability, Venn diagram, Prior Probability, Posterior Probability, Likelihood Ratio and Normalizing Constant.

INTRODUCTION

In most of the practical situations, the true data for the event is not available, and the only information provided is usually the time-series of historic data available. It is almost impossible for someone to know the prediction of any data before the event even happening. Therefore, we use Bayesian approach by assuming the current data depending on the prior data of the unknown parameter. This helps in solving the one central data in operational statistics in finding certain stats within which we can find an operational data that maximizes the expected profit uniformly for each and every possible value of the unknown parameters.

From identifying the problem to selection of a proper strategy for the event, Bayesian approach makes all the tasks easier for us.

While some people may criticize this approach because stages in this process are performed in parallel than in sequence, a more realistic model can allow several parts of the decision-making process to occur in different order in different decisions.

The Bayesian approach puts forward an effective mathematical structure with its fundamental need to fix complicated relationships among random variables in a straightforward way.

Since Bayes theorem is based on the concept of using a given data. Therefore, it is derived from the method of conditional probability.

The main objective of this research paper will be studying Bayes theorem in Operation research, deriving it from conditional probability,visualizing it using Venn diagrams and analyzing its application in some fields. This entire study will be followed by some real-life examples.

METHODOLOGY

Statement of Bayes' Theorem

Bayes' theorem relates the *conditional and marginal probabilities* of random events A and B:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Each term in Bayes' theorem has a conventional name:

• P (A) is called the *prior probability or marginal probability of A*. It is "prior" in the way that it does not take any information about B into consideration.

• P (A | B) is called the *conditional probability of A, given B*. It is also called the *posterior probability*

as it is derived from or depends upon the specified value of B.

• P (B | A) is called the *conditional probability of B, given A*.

• P (B) is called the *prior or marginal probability of B*, and acts as a *normalizing constant*.

Derivation from Conditional Probabilities

To derive this theorem, we begin with the definition of conditional probability.

A conditional probability is an expression of how one event is probable given that some other event has occurred. For instance, "what is the probability that the road is wet?" will have a different answer than "what is the probability that the road is wet *given that* it rained earlier?"

The conditional probability will hence be:

P (A | B) = P (A∩B) P (B).

Keeping in mind that, "the road is wet" is event A and "it rained earlier" is event B, this expression reads as "the probability the road is wet given that it rained earlier is equal to the probability that the road is wet and it rains over the probability that it rains."

Likewise, the conditional probability of event B given event A will be:

 P (B | A) = P (A∩B) P (A).

Rearranging and combining these equations, we find P (A | B) P (B) = P (A ∩ B) = P (B | A) P (A). This is sometimes called the product rule for probabilities.

Dividing both sides by P(B), providing that it is not a zero, we obtain Bayes' theorem:

P (A | B) = $\underline{P\ (B\ |\ A)}$ * P (A)

   P (B)

There are many equivalent ways to write the Bayes' Theorem

While the above equation can be applied to any probability distribution over events A and B, it has a particular interpretation in the case where A represents a hypothesis H and B represents some observed evidence E. In this case, the formula can be written as:

P (H | E) = $\underline{P\ (E\ |\ H)}$ * P(H)

   P (E)

This relates the probability of the hypothesis before getting the evidence P (E) to the probability of the hypothesis after getting the evidence, P (H | E). For this reason, P (H) is called the **prior probability**, while P (H | E) is called the **posterior probability**. The factor that relates the two, P (E | H) / P (E), is called the **likelihood ratio**. Using these terms, Bayes' theorem can be rephrased as "the posterior probability equals the prior probability times the likelihood ratio."

Here is another equation, that you can use to figure out the above problem. You'll get exactly the same result:

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

The main difference with this form of an equation is that it uses the probability terms *intersection*(∩) and *compliment* (^{c}). Think of it as shorthand: it's the same equation, written in a different way.

In order to find the probabilities on the right side of the equation, use the multiplication rule:

P(B ∩ A) = P(B) * P(A|B)

The two sides of the equation are equivalent, and P(B) * P(A|B) is what we were using when we solved the numerator in the problem above.

P(B) * P(A|B) = 0.01 * 0.9 = 0.009

As the denominator, we have P(B^c ∩ A) as part of the equation. This can be (equivalently) rewritten as P(B^c*P(A|B^c). This gives us:

P(B^c*P(A|B^c) = 0.99 * 0.08 = 0.0792.

Inserting those two solutions in the formula, we get:

0.009 / (0.009 + 0.0792) = **10%.**

Let's look at how does one go about solving a sum:

1)In a particular pain clinic,

- 10% of patients are prescribed narcotic pain killers.
- Overall, five percent of the clinic's patients are addicted to narcotics (including pain killers and illegal substances).
- Out of all the people prescribed pain pills, 8% are addicts.

If a patient is an addict, what is the probability that they will be prescribed pain pills?

Step 1: Figure out what your hypothesis "H" is from the question. That information is in the italicized part of this particular question. The event that happens first (H) is being prescribed pain pills. That's given as 10%.

Step 2: Figure out what your evidence "E" is from the question. That information is also in the italicized part of this particular question. Event E is being an addict. That's given as 5%.

Step 3: Figure out what the probability of E (Step 2) given H (Step 1). In other words, find what (B|A) is. We want to know "Given that people are prescribed pain pills, what's the probability they are an addict?" That is given in the question as 8%, or 0.8.

Step 4: Insert your answers from Steps 1, 2 and 3 into the formula and solve.
P (H | E) = P (E | H) / P (E) * P (H) = 0.08 /0.05 * 0.1 = 0.16

The probability of an addict being prescribed pain pills is 0.16 (16%).

2) Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result?

- One percent of women over 50 have breast cancer.
- Ninety percent of women who have breast cancer test positive on mammograms.

- Eight percent of women will have false positives.

Step 1: Assign events to A or X. We want to know what a woman's probability of having cancer is, given a positive mammogram. For this problem, actually having cancer is A and a positive test result is X.

Step 2: List out the parts of the equation (this makes it easier to work the actual equation):
P(A)=0.01
P(~A)=0.99
P(X|A)=0.9
P(X|~A)=0.08

Step 3: Insert the parts into the equation and solve. Note that as this is a medical test, we're using the form of the equation from example #2:

$(0.9 * 0.01) / ((0.9 * 0.01) + (0.08 * 0.99)) = 0.10.$

The probability of a woman having cancer, given the test result is positive, is 10%.

Visualizing Bayes' Theorem

Venn diagrams are particularly useful for visualizing Bayes' theorem, since both the diagrams and the theorem are about looking at the intersections of different spaces of events.

For example, A disease is present in 5 out of 100 people, and a test that is 90% accurate (meaning that the test produces the correct result in 90% of cases) is administered to 100 people. If one person in the group tests positive, what is the probability that this one person has the disease?

The intuitive answer is that the one person is 90% likely to have the disease. But we can visualize this to show that it's not accurate. First, draw the total population and the 5 people who have the disease:
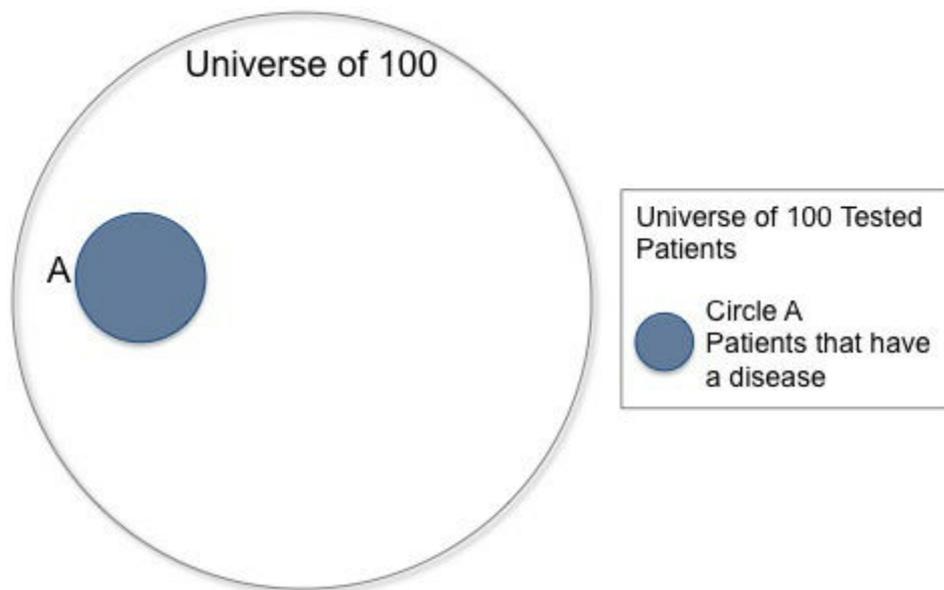
The circle A represents 5 out 100, or 5% of the larger universe of 100 people.

Further, overlay a circle to represent the people who get a positive result on the test. We know that 90% of those with the disease will get a positive result, so need to cover 90% of circle A, but we also know that 10% of the population who does not have the disease will get a positive result, so we need to cover 10% of the non-disease carrying population (the total universe of 100 less circle A).
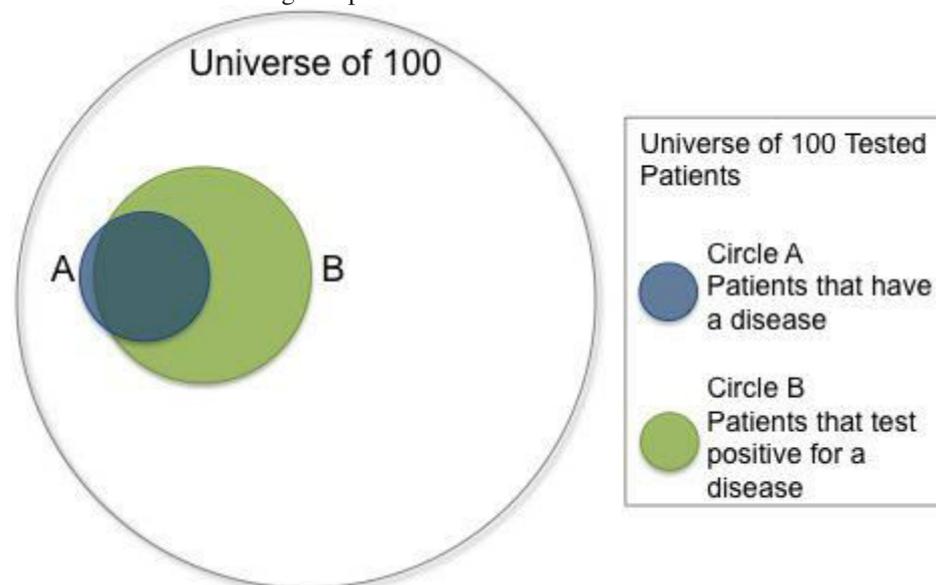
Circle B is covering a substantial portion of the total population. It actually covers more area than the total portion of the population with the disease. This is because 14 out of the total population of 100 (90% of the 5 people with the disease + 10% of the 95 people without the disease) will receive a positive result. Even though this is a test with 90% accuracy, this visualization shows that any one patient who tests positive (Circle B) for the disease only has a 32.14% (9 in 28) chance of actually having the disease.

ANALYSIS

In this section, we shall analyze how Bayes Theorem is applied in models, such as Bayesian Search Theory, Naive Bayesian Approach, in areas such Geological Mapping, Spam Filtering.

Bayesian Search Theory

In this sector, to search for a missing vessel or an aircraft in the ocean, Bayesian Search Theory is applied. It provides a principled and successful recipe for planning searches for lost aircraft and other vessels. It has been used for multiple searches, such as the US Navy Submarine USS Scorpion, Air France Flight AF477, and is currently being used to locate the wreckage of Malaysian Airlines Flight MH370.

The theory states that there are two factors which contribute to searching the missing vessel in a given location; the probability that the object is in a given location, and the probability of detecting said object in said searched location. You would calculate these probabilities and assuming you don't find the missing object in this location, you can update the probability of detection.

You can divide the search area into a grid of cells, as below. This is known as a **Probability Distribution Function (PDF)**

i = index of grid cell

Y_i = object's true existence in this cell (1 means it is located here)

π_i = probability of occurrence = $\Pr(Y_i = 1)$

An example of a PDF is of the USS Scorpion. It was found just a few nautical miles off in the highest probability cell i.e. 85.

NOTE: CONVERT NUMBERS TO PROBABILITIES BY DIVIDING BY 10,000.
SHADING INDICATES MAGNITUDE AS FOLLOWS:

NOTE: * INDICATES LOCATION OF SCORPION.

- BLANK — 0 ≤ NUMBER ≤ 10
- [shaded] — 10 < NUMBER ≤ 100
- [shaded] — 100 < NUMBER < 1000
- [shaded] — 1000 < NUMBER ≤ 10,000

Row	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O			
											3	11	14	24	6			
													5	7	1			
1											5	26	35	22	26	9	1	
2										18	46	74	42	18	10	4	2	
3								8	60	140	99	45	20	4	2	1	1	
4	2	21	137	16	7	1	20	215	239	105	30	5	3	1	1	1		
5	18	40	46	747	30	250	205	571	277	38	5	2	1	1	1			
6	14	326	3	1	28	31	63	85*	62	1	8	7	10	7	3	4		
7	359	175	174	096	282	245	82	71	65	35	27	9	12	6	5	4		
8	24	25	42	82	297	230	129	115	61	33	14	14	10	6	2	5	1	
9	17	25	20	20	20	19	55	99	46	30	14	15	3	5	1	6		
10	2	13	14	25	20	24	45	34	27	19	15	5	7	5	5	1		
11		7	13	12	9	1	3	3	11	14	5	4	3	2	1			
12								1	4	4	10	5	4	1				
									1	3	2							
										3	2							

To show how it done, we shall simulate how the Air France AF447 was recovered by the French Bureau of Enquiries and Analytics (BEA) using systems such as Search and Rescue Optimal Planning System (SAROPS).

First gathering all the intel, they could from the crash, experts had concluded that the flight could not have gone on for more than 40 nautical miles from the Last Known Point (LKP) they were contacted at. Therefore, a search area of roughly 5,025 nm2. Using their intel, they organized consistent scenarios, quantified the uncertainties into probability distributions, weighing the relative likelihood of detection in these scenarios, and producing a prior PDF. They updated the PDFs along with each unsuccessful search. The posterior object location was as follows: -

$$w_n^1 = \frac{\left(1 - p_d^1(n)\right) w_n}{\sum_{n'=1}^{N} \left(1 - p_d^1(n')\right) w_{n'}} \quad \text{for } n = 1, \ldots, N$$

*w1n is the posterior probability here that particle N represents the object's location

They were updated along with the surface searches and three phases of active sonar search for the black boxes and wreckage. The final PDF was as follows:
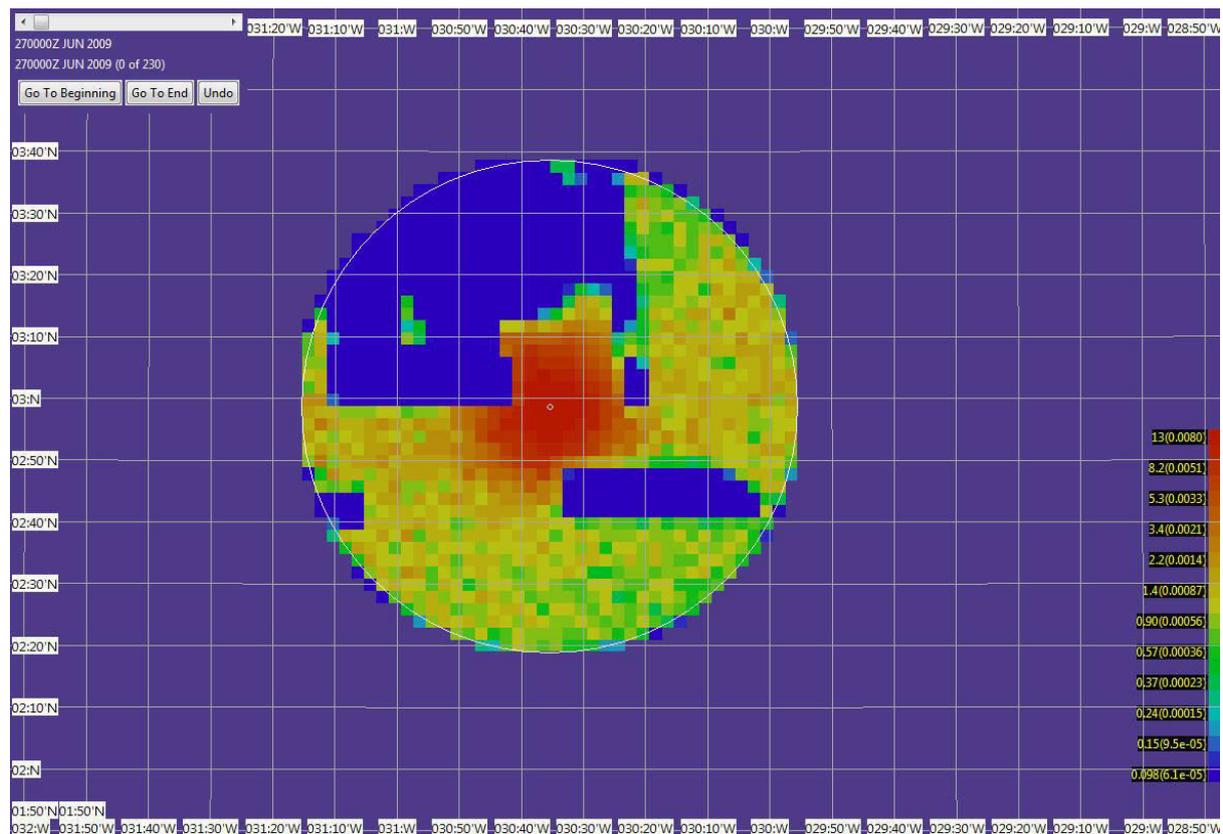
Within one week of effort, the wreckage was found. This is a successful illustration on how a methodical Bayesian approach can help in search planning.

Spam Filtering

The spam mail is a trouble mail that sent to many people, such as one-sided advertising mail, the chain mail, the fictitious claim mail, and computer virus spread by mail. The spam mail becomes a problem because an increase in the network traffic occurs because not only is our other, important mail buried by a large amount of spam mail.

When dealing with spam the **Naive Bayesian Approach** theorem is used to calculate a probability whether a certain message is spam, based on words in the title and message, learning from messages that were identified as spam and messages that were identified as not being spam and as regular mail instead, also known as Ham.

The filter separates the spam mail by the appearance probability of the character string (token) used in the mail.

An example of spam would be:

Paying too much for Viagra? Now you have a chance to receive a free trial!

An example of regular mail would be:

For the Clarins, just take your time. As I have advised her to exercise regularly.

Now here is a training dataset, which shows the appearance of these words in regular and spam mail.

Feature	Appearances in Spam	Appearances in Ham
A	165	1235
Advised	12	42
As	2	579
Chance	45	35
Clarins	1	6
Exercise	6	39
For	378	1829
Free	253	137
Fun	59	9
Girlfriend	26	8
Have	291	2008
Her	38	118
I	9	1435
Just	207	253
Much	126	270
Now	221	337
Paying	26	10
Receive	171	98
Regularly	9	87
Take	142	287
Tell	76	89
The	185	930
Time	212	446
To	389	1948
Too	56	141
Trial	26	13
Vehicle	21	58
Viagra	39	19
You	391	786
Your	332	450

We make the assumption that all the words are independent of each other and are randomly displayed. This way, if a word is deemed "Spam" it does not affect how the other words will be deemed. By Naive Bayes Assumption:

$$P(w_1, \ldots, w_n) = \prod_{i=1}^{n} P(w_i$$

We denote the two classes of email as Spam(S) and Ham(H). We find the two probabilities of P(E|S) and P(E|H). Here, we make the conditional independence i.e. that the appearance of a particular token is independent of other tokens when deemed as spam. Therefore, we derive the following expression:

$$P(E|S) = P(w_1, \ldots, w_n|S)$$
$$= \prod_{i=1}^{n} P(w_i|S)$$

and

$$P(E|H) = P(w_1, \ldots, w_n|H)$$
$$= \prod_{i=1}^{n} P(w_i|H)$$

The probability of getting a spam mail is defined in the following expression:

$$P(w_i|S) = \frac{P(w_i \cap S)}{P(S)} \quad and \quad P(w_i|H) = \frac{P(w_i \cap H)}{P(H)}$$

example, the word *Free* in our training dataset:

$$P(w_8 \cap S) = \frac{253}{432 + 2170}$$

$$P(S) = \frac{432}{432 + 2170}$$

ɘ get

$$P(w_8|S) = \frac{253}{432}$$

A problem that may come up is having extremely small values if we have a huge set of words. Therefore, we add log in the probability ratio.

$$\log \frac{P(S|E)}{P(H|E)} = \log \left(\frac{P(S)}{P(H)} \prod_{i=1}^{n} \frac{P(w_i|S)}{P(w_i|H)} \right)$$
$$= \log \frac{P(S)}{P(H)} + \sum_{i=1}^{n} \log \frac{P(w_i|S)}{P(w_i|H)}$$

If $P(S|E) > P(H|E)$, it means the result is greater than 0, and will be classified as Spam. If $P(H|E) > P(S|E)$, then the result is less than 0 and will be classified as Ham.

CONCLUSION

Bayes' theorem is an important reality check on our efforts to forecast the future and we tend to either dismiss new evidence, or embrace it as though nothing else matters. Bayesians try to weigh both the old hypothesis and the new evidence in a sensible way.

There is nothing mystical or anything profound about Bayes' theorem although, of course, we've only scratched the surface of how it can be applied in practice. The theorem does not really allow us to calculate anything that we can't find out by other means, but it has the advantage of being notational compact, and works directly with probabilities rather than counts.

The Bayes' theorem responds that intelligence must still do the best it can with what it has. In a world of fallible judgments about evidence, the Bayesian approach is not a path to perfection and it can be at best only a path to improvement. This research on Bayesian method is a mathematical logic to which intelligence can have recourse for substantiating or contradicting the verbalizations of the traditional analysis. When the different approaches lead to discrepant conclusions, intelligence should perhaps undertake to rethink, recalculate, and if possible reconcile.

At the core of Bayesian statistics is Bayes' theorem, which describes the outcome probabilities of random (dependent) events using the concept of conditional probability. Whenever you encounter new evidence, you should revise how probable you think things are. This is true in science, where we perform experiments to test hypotheses; it is true when trying to solve a mystery using evidence, or

trying to avoid getting a goat on a game show. Bayes' theorem is used to update probabilities.

Thus, Bayes' theorem is an extremely useful tool, more often right than wrong, but it is only ever a guide. Many areas of science are adapting to this reworking of an old theory, and it promises to fit alongside the traditional methods very well.

Although Bayes' theorem itself is quite simple, calculating likelihoods, priors and evidences for use in it can be difficult.

REFERENCES

1. https://brilliant.org/wiki/bayes-theorem/
2. http://www.eng.utah.edu/~cs5961/Resources/bayes.pdf
3. http://psychology.wikia.com/wiki/Bayes_theorem
4. http://www.statisticshowto.com/bayes-theorem-problems/
5. https://www.farnamstreetblog.com/2012/12/thomas-bayes-and-bayess-theorem/
6. http://www.kevinboone.net/bayes.html
7. https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol16no2/html/v16i2a03p_0001.htm
8. https://explorable.com/bayesian-probability
9. http://fusion2015.org/wp-content/uploads/2015/07/Fusion2015_Plenary Talk_ColleenKeller.pdf
10. http://fusion2015.org/wp-content/uploads/2015/07/Fusion2015_Plenary Talk_ColleenKeller.pdf
11. http://www2.stat.duke.edu/~banks/130-labs.dir/lab10.dir/Lab10_bayesian_search.pdf
12. https://www.r-bloggers.com/bayesian-search-models/
13. https://www.bea.aero/uploads/tx_elyextendttnews/metron.search.analysis.pdf
14. http://www.cs.ubbcluj.ro/~gabis/DocDiplome/Bayesian/000539771r.pdf
15. https://blog.malwarebytes.com/security-world/2017/02/explained-bayesian-spam-filtering/
16. http://www.statisticshowto.com/bayes-theorem-problems/
17. http://byjus.com/maths/bayes-theorem/
18. Khalid Iqbal, Xu-Cheng Yin, Hong-Wei Hao, QuaziMudassarIlyas and Hazrat Ali; An Overview of Bayesian Network Applications in Uncertain Domainhs
19. Leon Yang Chu, J.GeorgeShanthikumar, Zuo Jun Max Shen; Solving Operational Statistics via a Bayesian Analysis
20. Oleg Pokrovsky; Operational Research Approach to Decision Making