Modification of posterior probability variable with frequency factor according to Bayes Theorem
Bayes Teoremine göre frekans faktörü ile arka olasılık değişkeninin modifikasyonu

Mehmet Sait VURAL[1]*, Muhammed TELÇEKEN[2]
[1]Department of Computer Engineering, Gaziantep Islam Science and Technology University, Gaziantep, Turkey
[2]Department of Computer Engineering, Sakarya University of Applied Sciences, Sakarya, Turkey
ORCIDs: 0000-0003-2144-5474, 0000-0001-5223-2856
E-mails: mehmetsait.vural@gibtu.edu.tr , muhammedtelceken@subu.edu.tr
*Corresponding author.

***Abstract*—Probability theory is a branch of science that statistically analyzes random events. Thanks to this branch of science, machine learning techniques are used inferences for the prediction or recommendation system. One of the statistical methods at the forefront of these techniques is Bayesian theory. Bayes is a simple mathematical formula used to calculate conditional probabilities and obtain the best estimates. The two most important parts of the formula are the concepts of a priori probability and posterior/conditional probability. In a priori probability, the most rational assessment of the probability of an outcome is made based on the available data, while in posterior probability, the probability of the event occurring is calculated after considering all evidence or data. In this study, a new mathematical model is presented to calculate the posterior probability variable of Bayesian theory more precisely. According to this new mathematical model, equal priority probabilities of some variables should be recalculated according to frequency. Calculations are applied to two nodes. The first of these two nodes is the node consisting of the existing data, and the second is the queried node. The positive frequency value will be applied when the variables consisting of existing data and having the same a priori probabilities are found at the questioned node, and negative frequency value will be applied for the other variables. Thus, while calculating a standard probability value according to Bayesian Theory, frequency-based values are taken into account with the help of the newly created mathematical model. With the help of these frequencies, the modification of the system reveals more precise results according to these two basic principles. The results obtained were tested with the cross validation method and high accuracy rates were determined.**

Keywords*—Bayesian Theory; Machine Learning; Mathematical Model

***Özetçe*—Olasılık teorisi, rastgele olayları istatistiksel olarak analiz eden bir bilim dalıdır. Bu bilim dalı sayesinde makine öğrenmesi teknikleri tahmin veya öneri sistemi için çıkarımlarda kullanılmaktadır. Bu tekniklerin başında istatistiksel yöntemlerden biri Bayes teorisidir. Bayes, koşullu olasılıkları hesaplamak ve en iyi tahminleri elde etmek için kullanılan basit bir matematiksel formüldür. Formülün en önemli iki kısmı, önsel olasılık ve sonsal/koşullu olasılık kavramlarıdır. Önsel olasılıkta, bir sonucun olasılığının en rasyonel değerlendirmesi mevcut verilere dayalı olarak yapılırken, sonraki olasılıkta olayın meydana gelme olasılığı tüm kanıtlar veya veriler dikkate alınarak hesaplanır. Bu çalışmada, Bayes teorisinin sonsal olasılık değişkenini daha kesin olarak hesaplamak için yeni bir matematiksel model sunulmaktadır. Bu yeni matematiksel modele göre bazı değişkenlerin eşit öncelikli olasılıkları frekansa göre yeniden hesaplanmalıdır. Hesaplamalar iki düğüme uygulanır. Bu iki düğümden ilki mevcut verilerden oluşan düğüm, ikincisi ise sorgulanan düğümdür. Mevcut verilerden oluşan ve aynı a priori olasılıklara sahip değişkenler sorgulanan düğümde bulunduğunda pozitif frekans değeri, diğer değişkenler için negatif frekans değeri uygulanacaktır. Böylece Bayes Teorisine göre standart bir olasılık değeri hesaplanırken yeni oluşturulan matematiksel model yardımıyla frekans bazlı değerler dikkate alınmaktadır. Bu frekanslar yardımıyla sistemin modifikasyonu bu iki temel prensibe göre daha kesin sonuçlar ortaya koymaktadır. Elde edilen sonuçlar çapraz doğrulama yöntemi ile test edilmiş ve yüksek doğruluk oranları belirlenmiştir.**

Anahtar Kelimeler*—Bayes Teorisi; Makine öğrenme; Matematiksel model

I. Introduction

Human beings are always asking questions about the future today. What will happen tomorrow? What are my advantages or disadvantages for a startup? Humanity is in a constant search for future plans and predictions. Based on past experiences, human beings always go through a decision-making process about the benefits of an uncertain structure. At this point, statistics and probability theory is appeared as facts that help us and answer these questions. Statistical decision making theory is one of the solutions to this problem. Statistical decision theory deals with situations where decisions must be made in a situation of uncertainty, and its purpose is to provide a rational framework for dealing with such situations. In summary, Bayesian paradigm is a concept related to uncertainty [1]. That is why methods suitable for the uncertainty problem should be used to solve this problem.

At this point, Bayes Theory, one of the most used methods, will be discussed in this study. The Bayesian approach is a special way of formulating and dealing with statistical decision problems. More specifically, it provides a method of effectively combining existing observations to allow rational generation of optimal decision criteria. The most important reason behind this method is that Bayesian theory occupies an important place in inference problems. Possible results about how to predict an unknown or unforeseen event are revealed by this theory. These inferences and foresights are used in medicine, finance and different interdisciplinary studies. Bayesian theory is consisted of calculating the probabilities for the relations between the nodes. As priori and posterior/conditionally unknown calculations of the variables in an existing node are made. However, some problems are encountered in these calculations. The main problem is that all the variables with equal a priori probability values from the available data are treated equally. For a variable calculated from the available data; If this variable is also present in the event in questioned, then the value of this variable must be different from other variables with equal a priori probability. Bayesian theory rates it with the same value and thus, a situation contrary to reality emerges. The problem in the focus of the study is to offer a solution to this problem [2]. It is aimed to present a more realistic approach to this problem with a different approach and a mathematical method in Bayesian theory. Both theoretical and applied studies on Bayesian theory are examined. It is not uncommon for scientists to find support for a theory in evidence known long before the theory was introduced, so they intuitively present already known or old evidence and hypotheses that confirm new theories [3].

Studies are more on these evidences and hypotheses. Especially in the literature review, studies on this concept are quite extensive. However, these studies are mostly studies that give detailed information about Bayesian theory, formulas and models [4], [5]. In the studies examined, when a conclusion is to be drawn from the observed data, statistics, signal processing, speech analysis, image processing, computer vision, astronomy, telecommunications, neural networks, pattern recognition, machine learning, artificial intelligence, psychology, sociology, medical decision making, econometrics and Bayesian principles and tools are used in fields such as biostatistics. The Bayesian approach appears in studies that indicate that it is important not only at the practical application level, but also at deeper conceptual levels, and that it touches on the fundamental and philosophical aspects of scientific inference [6]. Bayesian probability theory provides a mathematical framework for inference or reasoning using probability. It is widely used in many scientific disciplines today. It is used from astrophysics to neuroscience and often to judge relative validity [7]. It has an important place in modeling as well as the theory and formulas of Bayesian theory.

Bayesian information criterion is one of the most widely known and widely used tools in statistical model selection. The biggest benefit of this modeling is its computational simplicity and effective performance [8]. In a study on the uncertainty of the model parameters, practical methods for modeling with real samples and simulated data and for diagnosing the model as a result of assumed mixed effects were discussed [9]. The Bayesian estimation approach is emphasized, which takes into account the uncertainty in the model form and also the uncertainty of the parameters assigned to each potential model [10]. One of the natural steps in the development of probability theory is to understand the conditioning process in conjunction with Bayesian conditioning. A study of probability measures shows that conditioning a family of probabilities induced by lower bounds based on probabilities of nested events can be reliably addressed in the probability representation itself [11]. In this theory, the concept of posterior/conditional probability rather than a priori probability is a more important concept for an unknown event. One of the objectives of this article is to analyze the frequency evaluation used in optimizing the concept of posterior probability, from a different perspective. There is a study on this subject that generalizes the concepts of posterior distribution of both accuracy and balanced accuracy in a multi-class manner. Here, the relative fractions of the class frequencies in the test data were estimated by replacing the Beta distribution with the Dirichlet distribution [12]. For Bayesian inference, which is also used for conditional probability, a study has been done that proposes the use of combined marginal probabilities [13]. This approach is a study in which complex statistical models are dealt with in the Bayesian framework when the calculation of the exact probability and thus the exact posterior distribution is impractical or even analytically unknown. It is known that Bayesian theory is used in many places in the field of artificial intelligence. In particular, artificial intelligence techniques are used for various inferences and predictions from existing data. In one of the studies conducted with this aspect, the Bayesian technique is described for the unsupervised classification of data and computers [14]. Again, in a study conducted in the field of artificial intelligence, it is shown how this theory can be effective in detecting fake exam takers [15].

Bayesian theory as artificial intelligence is examined within the framework of Bayesian belief networks. Studies have been conducted that address the problem of learning Bayesian network structures from data using an information theoretic dependency analysis approach and show that by developing algorithms, exponential complexity can be avoided in various tests [16]. Recently, several researchers have also studied the conditional probability distributions stored at each node in Bayesian networks. In these studies, firstly, it was determined what kind of data usage techniques are used to learn Bayesian networks with compact representations. Thus, it is shown how to derive a Bayesian score from a network structure consisting of parameter maps applied with a more general decision graph [17]. In another study proposed to address multi-response optimization problems, both the loss of quality and the reliability of the optimization results were examined. A new Bayesian optimization approach is proposed to address a set of problems involving the correlation, robustness and reliability of the optimization results in the multi-response optimization process [18]–[21]. The major contribution of this paper is to provide a new optimization scheme that combines a loss function approach with a posterior estimation approach within a single Bayesian modeling and optimization framework. The proposed approach not only takes into account the correlation between multiple responses and the uncertainty of the model parameters, but also takes into account the loss of quality and the reliability of the optimization results [22]. There is a similar study presenting an approach to multi-response surface optimization that not only provides optimal operating conditions, but also measures the reliability of an acceptable quality result for any set of operating conditions. With this approach, it is taken into account the correlation structure of the data, the variability of the process distribution, and the model parameter uncertainty [23]. There are few studies presenting a number of objections to Bayesian inference, with a different opposing view to a hypothetical Bayesian theory [24]. In these studies, considering the details of other arguments related to Bayes' method, results that may have different perspectives have been revealed. However, all these studies do not offer a different approach to the theory regarding the aforementioned problem. In summary, there is no study on converting variables with equal prior probability to more optimum values. For this reason, the chosen topic is a concept that makes novel approach of the study.

The work is organized as follows. In the second chapter, after Bayes theory and its formulas and prior and posterior/conditional probability concepts are given, calculating positive and negative frequency values for equal prior probabilities, which is the optimization of Bayesian theory, will be explained. In the third section, the results of the applications for equal frequency factor calculation will be shown with tables and graphics, and the findings obtained in the last section will be discussed in the results section.

II. Materials & Methods

A. Bayesian Theory

Bayes' Theorem is a method that was proposed by a mathematician, Thomas Bayes (1702–1761), and is an increasingly important method. Today, Bayes' theorem is used in many research areas from medicine to environmental science, from technological developments to psychology.In statistics and probability theory, Bayes' theorem is a mathematical formula used to determine the conditional probability of events. Essentially, Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that may be relevant to the event. For some statisticians, Bayes' theorem is of particular significance for solving problem. These statisticians differ on two basic principles. These two basic principles are objective and subjective philosophical approach. In general, he adopts a view that accepts probability values as a subjective value revealed by the observer, not an objective property on a philosophical basis. With this view, according to subjectivist probability thinkers, Bayes' theorem is a fundamental tool that enables updating and changing subjective beliefs about probability value in the light of new evidence. This system of thought forms the basis of a posterior approach. The main purpose in posterior probability theory is to find the questioned event from a known event. For this reason, it is expressed as the concept of conditional probability, since the operation is performed according to a certain condition. The conditional probability values for two events, A and B, differ from each other. this is because event A and event B are different. When the B event is known, the calculated event A and the opposite situation will create different values from each other. But even between these two inverse conditionality, there is a very definite relationship.

Bayesian probability uses theoretical and practical frameworks for reasoning and decision making under uncertainty. Therefore, artificial intelligence researchers actively prefer software based on Bayes in decision making.

Possibility theory is based on two basic principles. These;

- Prior probability
- It is conditional / posterior probability

1) Prior Probability: Pr(A) notation is used to express the probability of a random event A occurring independently of any event. Known as the probability of event A, this expression can be used with prior or unconditional probability names. In a priority probability, the most rational assessment of the probability of an outcome is made based on the available data. Prior probability is expressed as the ratio of a questioned event to all events (1).

$$Pr(A) = \frac{\#A}{N} \tag{1}$$

In the a priority probability formula defined as Pr, the total occurrence of event A is expressed with $\#A$, while the variable N shows the total number of events that have occurred. Therefore, the formula for Pr is; It is of great importance as the prior probability value must be known before calculating a conditional/posterior probability concept.

2) Conditional/Posterior Probability: In conditional/posterior probability, the probability of the event occurring is calculated after considering all evidence or data. Prior probabilities are not sufficient to express the probability that a random event A will occur due to a different random event B. Therefore, the notation P(A|B), called conditional or posterior probability, is used. The conditional probability of event A according to a known event B is shown below (2).

$$P(A\backslash B) = \frac{P(A \cap B)}{P(B)} \tag{2}$$

3) Bayesian Formula: Bayes' theorem allows us to relate the conditional probability of two events. This concept; It is concerned with determining the probability of event A for given event B and the probability of event B for given event A. These are not equal expressions but are related to each other because of the above equation. However, events A and B are independent events. The probability of the outcome of event A does not depend on the probability of the outcome of event B. It is showed mathematical formula of the structures mentioned below (3-5).

$$P(A\backslash B) = \frac{P(A) \cap P(B)}{P(B)} = \frac{P(B\backslash A) \times P(A)}{P(B)} \tag{3}$$

$$P(A\backslash B) = \frac{P(B\backslash A) \cap P(A)}{P(B\backslash A) \cap P(A) + P(B\backslash A') \cap P(A')} \quad (4)$$

$$P(Ai\backslash B) = \frac{P(B\backslash Ai) \cap P(Ai)}{\Sigma i(P(B\backslash Ai) \cap P(Ai))} \quad (5)$$

where

- P(A): Probability of event A \ Prior probability of event A
- P(A’): Probability of event A not happening \ Prior probability of event A’
- P(B): Probability of event B \ Prior probability of event B
- P(B’): Probability of event B not happening \ Prior probability of event B’
- P(A\ B):The probability of event A occurring when event B occurs
- P(AxB): It is the probability that both A and B events will occur.

Suppose the event set Ai is part of U. The U symbol denotes the space field.UiAi = U and for any different i and j; Ai $\cap$ Aj $\neq 0$; Then probability sums of each event A will be the probability value of P(U):

$$P(U) = \Sigma i(P(Ai)) \quad (6)$$

Instead of the entire event space, the probability of any particular event can be written. However, in this calculation, it is necessary to find the total for each event A by making one-by-one operations for the possible B event. The total for each event A by making one-by-one operations was calculated using

$$P(B) = \Sigma i(P(B\backslash Ai)) \quad (7)$$

For the $\forall$ i value, the above equation can be written provided that

Although probabilities are introduced over events, most of the discussion is on probabilities of random variables. A random variable is a variable that reports the result of some measurement operation. Therefore, while describing the event A, both the variables in the set A and the concepts of the variables in the event A will be taken into account. When two events are defined within this concept, it is necessary to mention the concept of Node in Bayesian theory. In its simplest form, two nodes must be defined for at least two events. The first of these two nodes is the node value of the known event and the other is the queried node (node) value. The connection between the nodes is shown with arrows and will operate unilaterally in accordance with the concept. Because of the Bayesian structure, double arrows are not used together. Therefore, there must be a link or correlation from event A to event B or from event B to event A.

4) Bayes Theorem in Machine Learning: Bayes’ Theorem is widely used in machine learning to develop more Classifier models. Basically, it is examined under two topics.

- Bayesian Optimization
- Bayesian Belief Networks

a) Bayesian Belief Netwoks: Both theoretical and practical reasons for artificial intelligence to use probabilistic reasoning are used to deal with probabilities. In such a concept for computer technology, we can perform our operations with Bayesian belief networks. Bayesian networks are graphical models used for reasoning under uncertainty, where nodes represent variables and arrows represent direct connections between them. These direct links are usually causal links. In addition, Bayesian belief networks allow for automatic updating of probabilistic beliefs about new information when new information becomes available by quantitatively modeling the connections between variables.

In a Bayesian Network;

- A set of variables
- The graphical structure that connects the variables and
- It consists of a set of conditional probabilities.

The simplest mathematical model created for the Bayesian network is given below (Fig. 1).

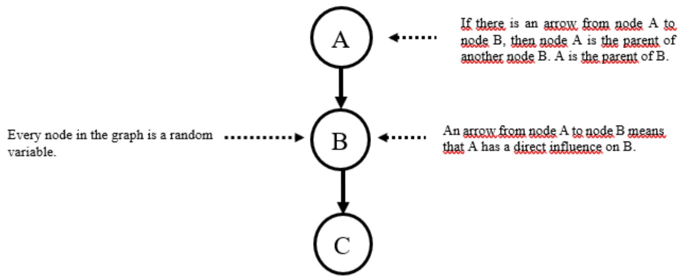

Figure 1: Bayesian network with three nodes

B. Frequency Factor For Conditional Probability

The probability of event B occurring when an event A is known is called the posterior probability. For posterior probability, a priority probabilities have to be calculated first. When calculating the prior probabilities of the variables in an event A, it will be seen that there are variables with equal value. The equal probabilities calculated for the prior probabilities of these variables are called equal frequency values. For equal frequency values of the variables, all variables in the posterior probability calculation should not be taken into account the same. However, according to the Bayesian theorem, they all operate to the same degree. This is not the right approach. Therefore, a more accurate calculation is needed for the event in question in posterior probability. Elements in each set were used as variables on two sets. When questioning the second set from the first set, there must be a difference between the variables with equal a priority probability. This difference is directly related to whether the variable exists in both sets. In summary, it is necessary to recalculate in a different way by distinguishing those who are in the two sets and those who are not. Because the variable in both places should have a stronger value than the other variables. This would be a more accurate approach. In Fig. 2, in order to use the Bayesian formula from node A to node B, the variables and their frequencies (repeat numbers) at each node are given below.

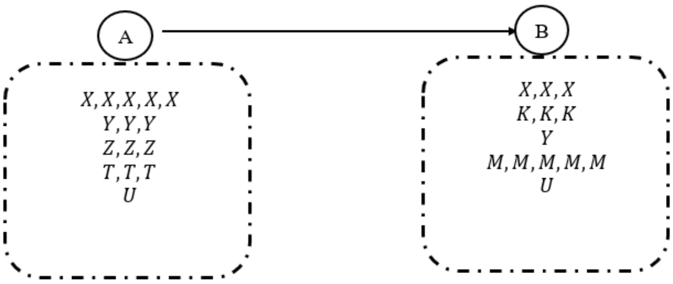

Figure 2: Frequency values of variables in nodes

The probability of the event B(Node2) from the data of the node A (Node1) given in the Bayes formula to be created: P(B\A) can be seen, for these two nodes whose prior probabilities are calculated, there are 3 variables that are equal to the prior probabilities of the variables in the known node. Whether this equivalent event is a covariate for nodes A and B should be determined as an extra factor when participating in the calculation. Thus, the results are more realistic and the probability values are determined by calculating with precise ratios. As seen in Figure-2 above, when the a priority probabilities of the variables used to query the B event from the A event in the correlation between the two nodes are revealed, It is checked whether the variables in node A are in node B. If there are variables at both nodes, the frequency calculation will be applied to these variables. As a result, it is seen that the a priority probabilities of the variables Y,Z,T among the variablesX,Y,Z,T,U in the A node are equal. When calculating the probabilities of these variables, they should not be evaluated the same for the event B. Therefore, a separate calculation of these variables is needed for reaching more accurate results. In this calculation, the common of the variables for node A and node B must be calculated differently, which is called the frequency factor. These frequency factors will operate in a more realistic philosophy with a more precise ratio for the variables as opposed to Bayesian calculation. For example, suppose there are criminals at node A and there is a theft at node A. On the other hand, let B be a drug crime. The question is: How to calculate the probability of committing a drug crime by those who commit theft crime according to Bayes theory? Here Bayesian theory makes no difference for computation and operates in the same way. This error should be eliminated. Therefore, a different approach should be established and the frequency factor should be calculated. Thus, even if the a priority probabilities of X,Y,Z criminals are the same, it would not be correct to consider them all to the same degree. Because if the criminal named Y has committed a drug crime before, it is the most logical and most likely solution that this criminal is more likely to commit a drug crime from theft than other criminals. This is why, unlike Bayes, the frequency factor should be added to this theory.

C. Frequency Factor For Prior Probability

Frequency Factor is the labeling of the variables between two nodes as positive and negative frequency factors according to whether they exist for the questioned node and calculating accordingly. These two concepts are explained in detail below. One of them is the positive Frequency factor and the other one is the negative frequency factor. Positive Frequency factor is applied to the common variables in both nodes. Negative Frequency factor is applied only for the variable that is in the current node but not in the questioned node. The two nodes used here are as follows. $Node_1$ is the current node or given node and $Node_2$ is the node to be queried calculated.

1) Calculation Of Positive Frequency Factor: In order to optimize the Bayesian approach and make it more realistic, frequency-based derived formulas have been created. The purpose of these formulas is to determine that some variables are in both events and that these variables have an advantage over other variables. In such a case, the positive frequency value should be calculated for that variable. In such an event, a positive frequency value will be applied to the common variables for the queried node from the known node between the two nodes. This formula will be applied if at least one of the variables whose equivalent prior probabilities are calculated at the known node, that is, at least one of the variables with the same prior probabilities, exists in both nodes. The difference between the positive frequency, called P^+, and the variables with equal prior probability will also emerge. Frequently used in this study, Node1 indicates the known or current event. Node2 will be the queried or targeted node (8).

$$P(Node2\backslash Node1) = P(QueriedNode\backslash Givennode) \quad (8)$$

In the P^+ formula (9),if the variable in Node1 is also in Node2; Pri(Prior probabilty), which is the ratio of the variable in the total number of elements, should be calculated before applying this formula (10).The formula will become more realistic by adding the probability of the same variable in the other node as a positive frequency value. This formula has three important parts. The first of these(P') will be the ratio of the frequency value of the variable in Node1 to the total number of elements in the same node. Frequency here means how many times the variable in any node is repeated. In the next step, the frequency value of the same variable in Node1 will be determined by the ratio of the total number of elements in all nodes(Pri). This section will result in the ratio of the found value to the total number of variables with equivalent prior probability(11). Finally, the positive frequency value will be calculated by the ratio of the number of frequencies in the queried node2 of same variable to the total number of elements in node2(Q).

$$P^+(variable) = P' + \frac{pri}{n1} + \frac{Q}{n3} \quad (9)$$

$$P'(variable) = \frac{\#Node1(variable)}{\#kNode1} \quad (10)$$

$$Pri = \frac{\#Node1(variable)}{N} \quad (11)$$

If the above expression is explained again, the concept of Node1(variable) used is shown how many times the variable in the same node is repeated. As a result, this calculation is expressed as calculating the frequency value of the variable in Node1. # kNode1 will give the total number of elements in Node1(11). n1 in (9) the total number of variables with equivalent prior probability in Node1. N represents the total number of elements used in all nodes. While the # kNode1 value indicates all the number of elements in Node1, and n1 indicates the total number of elements in all nodes, these two values are used as different structures used for the same variable. Q is the value to be found from the number of repetitions of the same variable in node2. The last value n3 used is the number of elements in the queried node (Node2).

2) Calculation Of Negative Frequency Factor: Contrary to the Bayesian approach, the P+ formula is used to find that variable in the questioned node (Node2), which is one of the variables with equivalent prior probabilities for the known node(Node1). However, a different strategy should be followed when calculating the probability value of the variables that are only in Node1 but not in the queried Node2 for the same node. This means to obtain a more minimized value from the variable calculated as P+. Therefore, the P- formula shown below will be used for variables that have equivalent probabilities in Node1 but not in Node2 (12).

$$P^{-} = P' - \frac{Pri}{n2} \tag{12}$$

In P- formula, if the variable in Node1 is not present in Node2, then the above calculated formula will be calculated. Here, the Pri value is the n2 ratio of the total number of variables that have the same a priority probability in $Node_1$ but not in $Node_2$. By calculating both frequency factors instead of Bayesian formula, the possible value of the variable for the questioned event will be calculated.

According to the above example, the image set of the values for which P+ and P- values were calculated in Node1 will be as follows. The prior probability values of the Y,Z,T variables in Node1 are equal to each other and have a value of 3⁄15. However, when the above formulas come into play and are analyzed in Node1 according to positive and negative values, Y, Z and T variables will have different values. Since the common element in both node1 and node2 is the Y variable, the P+ formula will be used. Since other variables are only in node1 and not in node2, the previous values will be recalculated for Z and T by using the P- formula Fig 3 and Fig 4.

As seen in Fig. 3, Y, Z and T variables in A node have equal a priori probability and have a value of 0.2. The formula P+ will be used for the variable named Y and P- formula will be used for the other two variables. Contrary to Bayes' theorem, it would be a realistic approach to have a higher probability of Y variable from these 3 variables than other variables, while the other two variables should also be minimized. The final values of the variable calculated according to the P+ and P- formulas are shown in the table below (Fig. 4).

III. Experimental Results

When positive and negative frequency calculations are used depending on the frequency factor, it has been seen that the

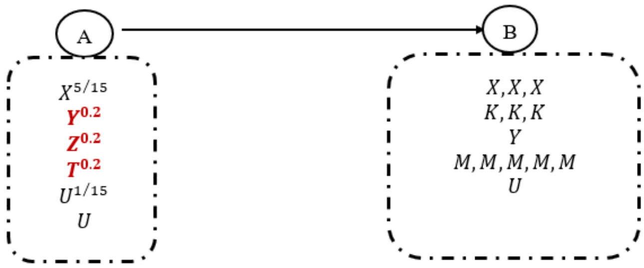

Figure 3: Prior probabilities of the variables in Node A according to Bayes theory

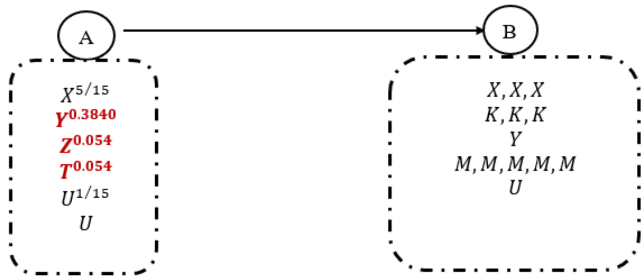

Figure 4: Probabilities of variants in A node according to Positive and Negative frequency factor

possible results of the variables are produced more sensitively and close to reality. These calculations were performed separately for various parameters such as total number of data (N), equivalent a priori probability values for various tests. An increase in the number of N used does not mean that the number of equivalent a priori probability increases at that rate. Since the operations are distributed with a random system, it is related to the possible number of repetitions (frequency) of the variable used in other variables. Even in the simplest manual calculation shown in Figure-3,4, it is seen that the Frequency factor is expressed with a more optimum value. Below, the increase and decrease changes according to Bayes and Frequency factor for these 3 variables are shown in Table I.

Equal variables	Bayesian	Frequency Factor	Increment Rate
Y	0.2	0.3840	0.1840
Z	0.2	0.054	0.146
T	0.2	0.054	0.146

Table I: Variations of variables with equal prior probabilities between Bayesian and Frequency Factor

As can be seen in Table tablo1, a value of 0.2 from the Bayesian approach is found for the X, Y, Z variables, while the value of the variable used in both nodes is calculated more precisely. The probability value of the variable Y, whose positive frequency value is calculated in the table above, did not come out like the equal ratio of other variables according to the bayesian formula. While this ratio increased from 0.2 to

0.3840, the other two variables decreased from 0.2 to 0.054. Because there are y values in both nodes, these results should be obtained in a realistic framework. Thus, it is seen that a more accurate calculation emerges in the light of the results. Likewise, for the other variables with equal frequency (Z and T), the equivalent probability amount found for Bayes is minimized with the P- formula. The values and results are changed according to the number of elements in Node1 used here, the number of elements in Node2 and the amount of the total number of elements, and the frequency values in that node for the variables in each node. More precise values are calculated as the number of repetitions, the number of variables and the total number of data increase. The Y variable not only has an increase of 0.1840, but the difference between them has increased by % 30 with the decrease of the Z and T variables by- 0.146. For this example, larger proportions and different test stages were made for N pieces of data. Equal a priori probability values in node1 are expressed as "Eqq", variables whose positive frequency values are calculated. In this equal a priori probability, the variation of the changes according to the Bayes formula with the frequency factor was found. If it is considered that there are 5 common a priori probability values in node1, each of these 5 values will be considered one by one. For each equal ratio considered, the number of variables will be looked at and the maximum number of variables will be chosen as the pivot. The frequency factor will be applied for the variables of this selected pivot value. As a result, the findings of the positive frequency value for the variable examined and the amount of increase are shown in Table tablo2.

P+	Node1	Node2	Bayes	Frequency Factor
N=50, Eq.deg.*=3	5	7	0.1	0.2013
N=50, Eq.deg.*=5	10	12	0.2	0.39
N=100, Eq. deg.*=3	15	18	0.1667	0.4146
N=100, Eq. deg.*=4	18	10	0.03	0.42
N=100, Eq. deg.*=3	12	20	0.01	0.5382

Table II: Change of equal a priori values for bayesian and frequency factor formula

In general terms, not only for N=50 or N=100, but also for different data numbers, tests and results are shown in the table below. As can be seen, when the system is operated repeatedly and different values are taken into account, it is seen that the values obtained for the frequency factor are more sensitive than the values found for the bayesian. Results are created in a scenario close to reality, even though it consists of random data (Table III).

Test Number	#Data	Equal Variable	Bayes	Frequency Factor
1	10	5	0.4871	0.53
2	20	3	0.67005	0.697
3	50	7	0.5614	0.6126
4	100	6	0.7180	0.7344
5	200	3	0.65	0.820
6	500	6	0.5004	0.73

Table III: Prior probability results according to Bayes and Frequency factor

Below is the graph of the frequency factor modeled for conditional probability (Fig. 5).

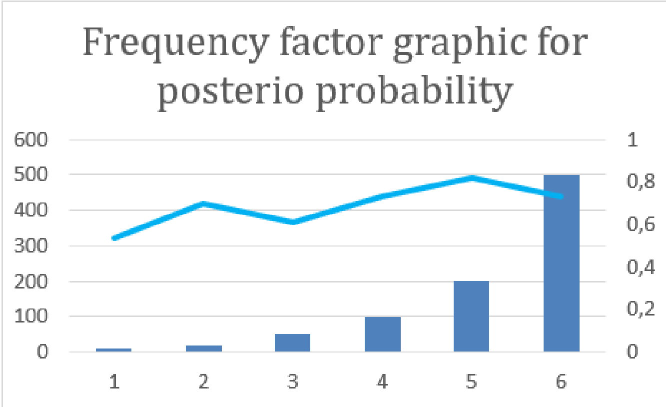

Figure 5: Positive and negative frequency factor for N data

A. Accuracy Rate

Success was determined according to the accuracy rate in the calculations made according to the Bayes network and frequency factor principle. In order to measure the accuracy of the proposed method, the data set is divided into training and test sets using the cross-validation method. The randomly selected % 90 is used for training and the remaining % 10 is used for testing.

If the index of the top variable in the common variable list, which is sorted according to the given formulas, is accepted as k, the accuracy rate can be calculated over the total number of k(variables) as follows (14).

$$AccuracyRate(\%) = (100 - \frac{k(variable)}{K})x100 \tag{13}$$

32 tests were carried out and a small section of these tests is presented above. Accordingly, when the frequency factor is not taken into account, the average conditional probability value between two nodes according to the number of N variables and data selected according to Bayes theory has a success rate of % 64.3, according to Bayesian theory, while this rate has increased to % 71 with the frequency factor. Even for the above 6 tests, a Bayes % 58 and Frequency factor % 66.7 success rate was achieved.

References

[1] Lindley DV. Theory and practice of Bayesian statistics. Journal of the Royal Statistical Society. Series D (The Statistician) 1983; 32(1/2): 1-11.

[2] Lavine M, Schervish MJ. (1999). Bayes factors: What they are and what they are not. The American Statistician 1999; 53(2): 119-122.

[3] Eells E. Problems of old evidence. Pacific Philosophical Quarterly 1985; 66(3-4): 283-302.

[4] Wang X, Bradlow E, Wainer H. A general Bayesian model for testlets: Theory and applications. Applied Psychological Measurement 2002; 26(1): 109-128.

[5] de Braganca Pereira CA, Stern JM. Model selection: Full Bayesian approach. Environmetrics 2001; 12(6): 559-568.

[6] Bruyninckx H. Bayesian probability. 2002. Retrived from https://people.cs.kuleuven.be/~danny.deschreye/urks2to4_text.pdf at March 13, 2022.

[7] Olshausen BA. Bayesian probability theory. The Redwood Center for Theoretical Neuroscience, Helen Wills Neuroscience Institute at the University of California at Berkeley, Berkeley, CA, 2004.

[8] Neath AA, Cavanaugh JE. The Bayesian information criterion: Background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics 2012; 4(2): 199-203.

[9] del Castillo E, Colosimo BM, Alshraideh H. Bayesian modeling and optimization of functional responses affected by noise factors. Journal of Quality Technology 2012; 44(2): 117-135.

[10] Rajagopal R, del Castillo E. Model-robust process optimization using Bayesian model averaging. Technometrics 2005; 47(2): 152-163.

[11] Dubois D, Prade H. Bayesian conditioning in possibility theory. Fuzzy Sets and Systems 1997; 92(2): 223-240.

[12] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In 20th International Conference on Pattern Recognition. August 23-26, 2010, Istanbul, Turkey, pp. 3121-3124.

[13] Pauli F, Racugno W, Ventura L. Bayesian composite marginal likelihoods. Statistica Sinica 2011; 21: 149-164.

[14] Cheeseman PC, Kelly J, Self M, Stutz JC, Taylor W, Freeman D. Autoclass: A Bayesian classification system. In Proceedings of the Fifth International Conference on Machine Learning (ML'88). June 1988, pp. 54-64.

[15] Sinharay S, Johnson MS. The use of the posterior probability in score differencing. Journal of Educational and Behavioral Statistics 2021; 46(4): 403-429.

[16] Cheng J, Greiner R, Kelly J, Bell D, Liu W. Learning Bayesian networks from data: An efficient approach based on information theory. Artificial Intelligence 2002; 137(1-2): 43-90.

[17] Chickering DM, Heckerman D, Meek C. A Bayesian approach to learning Bayesian networks with local structure. In Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence (UAI'97), August 1997, pp. 80-89.

[18] Elbir A, Ilhan HO, Aydin MF, Demirbulut YE. The implementation of classification and clustering techniques on Churn analysis. Journal of Intelligent Systems with Applications 2019; 2(1): 72-75.

[19] Sayilgan E, Yuce YK, Isler Y. Frequency recognition from temporal and frequency depth of the brain-computer interface based on steady-state visual evoked potentials. Journal of Intelligent Systems with Applications 2021; 4(1): 68-73.

[20] Degirmenci M, Sayilgan E, Isler Y. Evaluation of Wigner-Ville distribution features to estimate steady-state visual evoked potentials' stimulation frequency. Journal of Intelligent Systems with Applications 2021; 4(2): 133-136.

[21] Kiziltas Ok S, Yeniad M. Diabetes prediction using machine learning techniques. Journal of Intelligent Systems with Applications 2021; 4(2): 150-152.

[22] Wang J, Ma Y, Ouyang L, Tu Y. A new Bayesian approach to multi-response surface optimization integrating loss function with posterior probability. European Journal of Operational Research 2016; 249(1): 231-237.

[23] Peterson JJ. A posterior predictive approach to multiple response surface optimization. Journal of Quality Technology 2004; 36(2): 139-153.

[24] Gelman A. Objections to Bayesian statistics. Bayesian Analysis 2008; 3(3): 445-449.