Bayesian Probability Estimation for Reasoning Process

Sara Salehi

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Applied Mathematics and Computer Science

Eastern Mediterranean University
May 2014
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Applied Mathematics and Computer Science.

Prof. Dr. Nazim Mahmudov
Chair, Department of Mathematics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Applied Mathematics and Computer Science.

Prof. Dr. Rashad Aliyev
Supervisor

Examining Committee

1. Prof. Dr. Rashad Aliyev
2. Prof. Dr. Agamirza Bashirov
3. Asst. Prof. Dr. Ersin Kuset Bodur

ABSTRACT

It is a comprehensible fact that people always desire to be able to remove or at least to decrease the level of uncertainty in real world application. In all the areas of science and technology, it is important to have an accurate measurement for evaluating the uncertainty.

Increasing accuracy of measurement includes the identification, analysis and minimization of errors, compute and estimate the result of uncertainties. A probability is the branch of science studying the quantitative inferences of uncertainty. Probability is involved in various fields such as finance, meteorology, engineering, medicine, management etc.

In this thesis, Bayesian probability estimation for reasoning process is analyzed. The conditional, joint, prior, and posterior probabilities are mentioned. The importance of the probability views based on the subjectivity and objectivity, and the properties of these two terms are considered. The Bayesian inference and the generalized Bayes' theorem are discussed.

Keywords: Uncertainty, Bayesian method, subjective and objective probabilities, Bayesian inference, generalized Bayes' theorem

ÖZ

Bilinen bir gerçektir ki insanlar farklı uygulamalarda belirsizlik derecesini yok etmeğe veya en azından küçültmeğe isteklidirler. Bilim ve teknolojinin tüm alanlarında belirsizliği değerlendirmek için hassas ölçüm gereklidir.

Hassas ölçümü yükseltmek amacı ile belirsizliğin tanımlanması, tahlili, hatanın en az olması, sonuçların hesaplanması ve değerlendirilmesi gerekir. Olasılık bir bilim dalı olarak belirsizliğin nicel çıkarımlarını öğrenir. Olasılık finans, meteroloji, mühendislik, tıp ve başka alanlarda yer alır.

Bu tezde Bayes olasılığı uslamlama işlemi için incelenir. Koşullu, bileşik, önsel, ve sonsal olasılıklardan bahsedilir. Öznellik ve nesnelliğe dayanan olasılık görünümlerinin önemi, ve bu iki kavramın özellikleri dikkate alınır. Bayes sonuç çıkarma ve genelleştirilmiş Bayes teoremi tartışılır.

Anahtar Kelimeler: Belirsizlik, Bayes yöntemi, öznel ve nesnel olasılıklar, Bayes çıkarımı, genelleştirilmiş Bayes teoremi

ACKNOWLEDGMENTS

First of all, I want to extend my gratitude to my thesis supervisor Prof. Dr. Rashad Aliyev. Unless his assistance and support in each step of the process, my thesis would not have been completed. I want to thank my lovely family for encouraging and inspiring me to go my own way. Without any doubt, none of this could have happened without complete support of my family.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

Reasoning about uncertainty is a significant and valuable synthesis of the uncertainty in mathematics as it has evolved in a wide variety number of related fields such as computer science, probability, statistics, artificial intelligence, economics, logic, game theory, and even philosophy. Understanding why uncertainty plays a significant role in human affairs is not difficult. For instance, making decisions in everyday life is indivisible from uncertainty. People do not have certain information about what happened in the past because of absence broad and stable data about the past. People do not have a certainty about present affairs due to the lack of appropriate information. Making an appropriate decision in all situations is the most important capability of a person. To understand the capability, firstly it is needed to comprehend the notation of uncertainty. The terms of uncertainty and information are connected to each other strongly. Uncertainty is defined as a lack of information. However, information is used to reduce uncertainty.

The objective of making each system is different, such as forecasting, planning, control etc. Nowadays, uncertainty theory has become a branch of mathematics for modeling human uncertainty.

Ordinary life is not imaginable without uncertainty. However, from the traditional point of view a science without uncertainty is the best choice we may desire, but the complete elimination of the uncertainty is almost impossible.

Statistical methods are applied to problems that include systems with components the behavior of which is random. Additionally, statistical methods deal with problems in business fields such as marketing, investment and insurance. At the beginning of the 20th century, there was no positive attitude about the concept of uncertainty when statistical methods were accepted by the scientific community.

Uncertainty relates to conditions that are not exactly measurable or quantifiable and not controllable by human. Uncertainty occurs in the condition of lacking the complete information such as deficit of perfect information regarding models, phenomena, data or process to precisely determine future outcomes. Besides incomplete information, uncertain variables are usually subjects to a certain level of errors because of their randomization characteristics. Reducing the effects of uncertainty in the decision-making process and making the best possible decision among existing options are the main purposes of uncertainty analysis.

Recognizing sources and different uncertain variables are the fundamental steps in risk and reliability analysis. The uncertainty analysis is the process that prevents system failure. The main common steps to evaluate uncertainty are given below [1]:

Step 1: Recognizing sources of uncertainty;

Step 2: Derive the probability distribution function of desired uncertain variables;

Step 3: Insert uncertain variables into the model and estimate results;

Step 4: Finding the most sensitive variables.

According to [2], the terms of uncertainty are used in different ways, and defined by many specialists in statistics and decision making theory as following:

1. Uncertainty: This is a lack of certainty in different situations. If we do not have enough knowledge, it is impossible to describe data exactly, and predict a future outcome;

2. Measuring the uncertainty: Probabilities are set to all the results or all possible states, and the application of a probability density function (PDF) is performed.

The importance of the application of "mathematics of chance" including such fields of mathematics as graph theory, analysis, and mathematical physics, is undeniable.

Bayes theorem is a theorem of probability theory that is named after Reverend Thomas Bayes (1701–1761). Nevertheless, the French mathematician Pierre-Simon Laplace was a pioneer of what is called Bayesian probability these days. Bayes theorem in statistics and probability theory is the result derived from the most basic axioms of probability. It is a result of a mathematical manipulation of conditional probabilities.

Bayes' theorem is an important method aiming to understand the provided evidences, what are really known and other information. Additionally, Bayes' theorem is used to define the existence of relationships within an array of simple and conditional probabilities. It helps making "conditional probabilities" into the conclusions. Bayesian probability belongs to the group of evidential probability and is one of the various interpretations of the idea of probability. To compute the probability of a hypothesis or a problem, the Bayesian probability identifies some prior probabilities which will be updated in the light of related data. Bayes' theorem is used in different topics; its range varies from marine biology to the spam blockers from an email by the evolvement of the Bayesian approach. As a view of science, it is used to make a clear relationship between theory and evidence. By the use of Bayes' theorem, many insights into the philosophy of science which involves falsification and confirmation, and many other different topics can be made more accurate. Bayes' theorem has not been overlooked firstly and will not be the last in probability and uncertainty questions.

According to [3], both of the Bayesian method and classical method have advantages and disadvantages, and also they have some similarities. The results of both Bayesian method and classical method are similar to each other when the sample size is large. Some advantages of Bayesian analysis are:

- Within a solid decision theoretical framework, Bayesian method provides a way to combine prior information with data. By solid decision theoretical structure, this method makes the principle and natural way for combining prior information with

data. The previous information of parameter can be formed as prior information for the next analysis. With this method, the last posterior distribution can be used as a prior distribution when new observations become available;

- Without any reliance on an asymptotic approximation, this method provides inferences that are conditional and exact on data. Both small and large sample inferences proceed in the same manner. Bayesian analysis can also estimate any function of the parameters;

- Whereas the classical method, Bayesian analysis obeys the likelihood principle. The likelihood principle is not appropriate for the classical inference;

- It gives answers like "the parameter θ has a probability of 0.65 of increasing in 65% credible interval";

- A wide range of models can be conveniently set.

Using Bayesian analysis has also some disadvantages such as:

- It cannot say anything about selecting a belief since there is no proper way to select a prior. If it is not done with caution, it might generate misleading results;

- From a point of the practical view, sometimes it might be problematic to convince the subject matter expert who disagrees with the validity of the selected prior. It can yield posterior distributions which are affected by the priors;

- Sometimes in a model with a large number of parameters, it comes with a high computational cost. Additionally, by using the same random seed the simulation provides a little bit different answer. The slight differences in simulation outcomes do not contradict earlier claims that Bayesian inferences are exact. Given the likelihood function and also the priors, the posterior distribution of the parameter is precise, whereas estimates of posterior quantities by simulation-based way can vary as a result of the random number generator which is used in the process.

Chapter 2

REVIEW OF EXISTING LITERATURE ON BAYESIAN PROBABILITY AND ITS APPLICATIONS

Bayesian estimation techniques are used in [4] to evolve a dynamic stochastic general equilibrium (DSGE) model for an open economy and the estimation is performed on Euro. Based on the DSGE model, some open economy features such as a number of nominal and real frictions that have verified to be essential for the empirical fit of closed economy methods are incorporated. The evolvement of the standard DSGE model for an open economy is realized.

[5] demonstrates how Bayesian method is used to calculate a small scale, structural general equilibrium model. The monetary policy of four countries Australia, Canada, New Zealand, and U.K. is compared. The outcome of this study is that the central banks of Australia, New Zealand, and U.K consider the nominal exchange rate in their strategic policy, but the bank of Canada does not consider nominal exchange rate.

The properties of Bayesian approach are studied in [6] to estimate and compare the dynamic equilibrium economies. If even the models are non-nested, nonlinear, and misspecified, both tasks can be done.

The simulation based Bayesian inference procedures in a cost system are investigated in [7]. The reason of using the cost function and the cost share equations augmented in Bayesian inference procedures is to accommodate technical and allocative inefficiency. The way of estimating a translog system in a random effects framework is also represented.

According to [8], Bayesian approach is presented in order to investigate aggregate level sales data in a business with different kinds of products. A reparameterization of the covariance matrix is introduced, and it is also illustrated that this method is suitable with both actual and simulated data. In addition, based on the sampling experience, it is shown that this approach could be suitable for those who want to exchange one additional distributional assumption to raise efficiency in estimation.

A new Bayesian formulation in order to get the spatial analysis of binary choice data based on a vector multidimensional scaling procedure is presented in [9]. Approximation of a covariance matrix is permitted by the computational procedure. The posterior standard errors can be calculated.

[10] focuses on determination the exchange rate target zone models and also rational expectations models by developing a Bayesian approach. In addition, it can incorporate a stochastic realignment risk by introducing a simultaneous-equation target zone model.

A Bayesian approach for semiparametric regression in multiple equation models is given in [11]. The developed empirical Bayesian approach is used for estimation the prior smoothing hyperparameters from the data.

The shock and friction in two different economy areas such as US and euro over a common sample period (1974–2002) to estimate a DSGE model are compared in [12]. Differences in both shocks and the propagation mechanism of shocks, can affect the differences in business cycle behavior. In order to clarify which of them affects exactly on the business cycle, the structural estimation methodology is used.

According to [13], one of the ways of accounting in the uncertainty model, mostly in regression models for finding the determinants of economic growth is Bayesian model averaging (BMA). In order to do BMA, a prior distribution in two different parts should be specified, and the first one is a prior for the regression parameters and the second one is a prior over the model space.

The general idea of the paper [14] is to introduce a Bayesian posterior simulator in order to fit a model which allows a nonparametric behavior of the body mass index (BMI) variable, and also whose execution needs only Gibbs steps. In order to prove the result, data from the British Cohort Study in 1970 was used. The outcomes demonstrate that there are nonlinearities in the relationship between log salaries and BMI that is different across women and men.

In [15], Bayesian model averaging approach is used to predict realized volatility. Compared to benchmark models, this approach provides improvements in point forecasts.

The main point of the paper [16] is a first order autoregressive non-Gaussian model which can be used for analyzing the panel data. This modeling approach is considered to get sufficient flexibility without losing interpretability and computational ease. The model combines individual effects and covariates and it is noticed to the elicitation of the prior.

Monte Carlo (MC) method is used to draw parametric values from a distribution defined on the structural parameter space of an equation system [17]. The MC method is successful in some existing difficulties of applying Bayesian method to medium size models.

The similarities and differences between Bayesian and classical methods are studied in [18]. It is shown that both results in virtually equivalent conditional estimate partworths for customer. Therefore, selecting Bayesian or classical estimation becomes one of implementation conveniences rather than parametric usefulness.

Bayesian method is used in [19] to get quintile regression for dichotomous response data. This view to the regression has problems in making inferences on the parameters and also in the optimizing the objective function. The problems to be set

by Bayesian method are avoided by accepting additional distribution assumption on the error.

Because of the computational efficiency, direct theoretical base, and comparative accuracy, the naive Bayes classifier is useful in an interactive application [20]. This paper also compares and contrasts the lazy Bayesian rule (LBR) and the tree-augmented naïve Bayes (TAN), finding that these two techniques have comparable accuracy to be selected on the base of the computational profile. In order to classify the small number of objects, it is desirable to use LBR while TAN is used with a large number of objects.

The application of Bayesian decision theory is useful for making an effective cooperation of multiple decentralized components in a job scheduling, so it is necessary to have a heuristic matching a process dynamically [21]. The important points of using Bayesian decision theory are that its rules and principles are applied as a systematic approach to complicated decision making under conditions of incomplete information.

Bayesian estimation is provided in [22] to loosen the problem when it deals with lack of information. The default data can be extremely sparse mainly when reducing to issues with specific characteristics. Using Bayesian estimation techniques is adopted since classical tools result larger estimation errors.

Abnormal returns cause hamper in the study of statistical inference, in the long-horizon event. An approach that controls other popular testing methods, and also overcomes these difficulties is presented in [23]. The usefulness of the methodology is illustrated.

Model such as the method of maximum likelihood is developed for the spatial representation of market structure [24]. A Bayesian estimation method is provided in to overcome the traditional problems associated with estimating models with such correlated alternatives.

Based on [25], for solving the lack of loss data in operational risk management, which can affect the parameter estimates of the marginal distributions of the losses, Bayesian method and simulation tool should be used. By using Bayesian method, it is allowed to integrate the scarce and, sometimes, incorrect and imperfect quantitative data. Markov chain and Monte Carlo (MCMC) simulations are required to estimate the parameters of the marginal distributions.

In order to combine expert opinions and historical data, Bayesian inference is an appropriate statistical technique [26]. Bayesian inference methods are illustrated for operational risk quantification.

The Bayesian hierarchical structure is described in [27] in order to model calibration from historical rating transition data. The way of assessing to the predictive performance of the model, under the condition of lack of event data, is indicated.

Geographic information system (GIS)-based Bayesian approach for intra-city motor vehicle crash analysis in Texas during the five years crash data is presented in [28]. This method is suitable in estimating the relative crash risks, and in eliminating the uncertainty of estimates.

A model for time series of continuous results that could be defined as density regression on lagged terms or fully nonparametric regression is presented in [29].

[30] demonstrates how subnational population estimation can be performed within a formal Bayesian framework. A major part of the framework is a demographic account providing a whole description of the demographic system. A system model describes regularities within the demographic account, and an observation model describes the relationship between observable data and the demographic account. For the illustration of the model, data for six regions within New Zealand is used.

By growing problem of junk email on the internet, methods for the automated construction of filters in order to eliminate unwanted emails are examined in [31]. It is possible to use probabilistic learning methods in joining with a notation of differential misclassification cost to make filters that are specifically suitable for the nuance of this task.

The intuitive technical approaches are used in inference systems, in artificial intelligence because of the absence of sufficient statistical samples compels reliance on the subjective judgment of specialists. A subjective Bayesian inference method is

described in [32] to indicate some advantages of both formal and informal approaches.

Bayesian methods are developed for combining models and applied using various time series models which yield forecasts of output growth rates for eighteen countries, 1974–87. These odds are used in predictive tests to make a decision whether or not to combine forecasts of alternative models [33]. The Bayesian and non-Bayesian methods combine models, and represent the application of forecasting international growth rates.

An autoregressive, leading indicator (ARLI) model are described in two forms for forecasting of growth rates of 18 countries for the years 1974–1986 [34]. For computing probabilities of downturns and upturns, Bayesian predictive densities are used.

Chapter 3

BAYESIAN METHOD FOR CALCULATION OF PROBABILITY OF EVIDENCE

3.1 Bayes' Rule

Probabilities represent a set of logical beliefs, and there is a connection between information and probability. Bayes' rule is used to describe a logical way to update beliefs because of new information. Bayes' rule is for the process of inductive learning to make the Bayesian inference. In general, Bayesian methods are obtained from the rules of Bayesian inference. Bayesian methods provide the estimation of parameter with suitable statistical properties, a description of observed data, prediction of missing and unknown data, forecasting of future data, estimation of a model, validation and selection. So, Bayesian methods are derived to exceed the formal task of induction [35].

Bayesian methods make statements about the incomplete and partial available knowledge that is based on data and concerning some unobservable situation in a systematic way, using probability as a measurement. The following reasons show that the probability is a reasonable way for quantifying uncertainty:

1) By analogy: physical randomness causes uncertainty, so uncertainty is described in the language of random events. The use of different terms such as "probably" and

"unlikely" in general speech causes an extension of a formal probability calculus to problem of scientific inference;

2) Axiomatic approach: this approach puts all statistical inferences to the concept of decision making with gains and losses. It is implied that the uncertainty should be defined in terms of probability [36].

One of the explanations of the concept of probability is Bayesian probability that belongs to the category of evidential probabilities. An extension of propositional logic can be a Bayesian probability that enables reasoning with proposition truth or falsity of which is uncertain. Bayesian probability clarifies prior probability, which is then updated into relevant data to evaluate the probability of the hypothesis [37].

The purpose of a statistical analysis to compare with probabilistic modeling is fundamentally an inversion purpose. To clarify, when observing a random phenomenon directed by a parameter θ, statistical methods deduce from these observations an inference which can be a characterization or a summary about θ. In the notation of the likelihood function, this inverting aspect of statistics is obvious since it is rewritten in the sample density in the proper order,

$$\ell(\theta|x) = f(x|\theta),$$

as a function of θ, based on the observed value x [38].

The important concepts in probability theory are events and the probabilities of events. An event is defined with a probability which is allocated to a number between 0 and 1. The mathematical propositions are categorized by the elements in a given set Ω. In this way, each of the predicates defines a subset

$$E = \{x \in \Omega | p(x)\ is\ true\}$$

If two predicates define the same subset, they are equivalent. For clarification, instead of the various equivalent predicates an event is called a subset of Ω.

A random variable X, which is a map from Ω to $\mathbb{R}$, $X: \Omega \rightarrow \mathbb{R}$, is defined on the base space Ω of a probability space(Ω, ε, P). To give emphasis on the importance of the image of X, the name of "variable" is used. A probability measure on the image $X(\Omega)$ is caused by the probability measure of P on Ω.

3.2 Conditional Probability

The term called conditional probability is defined as a practical tool for computing the probability of two or more events. The conditional probability is one of the main important concepts in the probability theory which is defined in elementary statistics.

In general, every subset of the sample space Ω must not necessarily be an event. For example, some of the subsets cannot be measured, where the events are intervals like "between 20 and 35 miles" and unions of such intervals, but not "irrational numbers between 20 and 35 miles" [39].

The probability space (Ω, ε, P) is a space which posses the following three main sections:

1. The sample space indicated by Ω consisting of all the events which are considerable to be happening;

2. All subsets of the sample set. Each subset can be assumed as a set itself and is shown generally by ε;

3. A mapping defined from a subset of a sample set to the set including all probability values belonging to the interval [0,1].

Let's assume that A and B are two events, so A and B belong to the event set, $A, B \in \varepsilon$, such that the probability of B is greater than 0, $P(B) > 0$. The conditional probability of A given B is denoted by $P(A|B)$, and it is defined by

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

This means that B is taken as a certain event, the probability of A is $P(A|B)$. The probability that both of A and B occur is $P(A \cap B)$ on the numerator and the denominator rescales this number in order to find conditional probabilities. In fact, let $P_B: A \in \varepsilon \longmapsto P(A|B) \in [0,1]$. Then (ε, P_B) is a new measure on Ω such that $P_B(B) = 1$ and, more generally, $P_B(C) = 1 \; \forall C \in \varepsilon$ such that $C \supset B$.

In case of independences of two events such as A and B, the conditional probability of A given B is only based on A. It means that understanding that B has happened cannot make any changes that the probability of A happening.

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

There is a notation that null sets or empty sets are independent of the other sets, in particular, the empty set is independent of itself, so in order to be concerned with empty set or a set with measure 0, it is observed that

$$0 \leq P(A \cap B) \leq P(A) = 0$$

3.3 Bayes Formula and Total Law of Probability

Bayes formula: Let (Ω, ε, P) be a probability space and let $B \in \varepsilon$. In case both $P(A)$ and $P(B)$ are positive, then

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A)$$

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)(1 - P(B))$$

Proposition (Bayes formula): Let(Ω, ε, P) be a probability space and let $A, B \in \varepsilon$ such that $P(A)P(B)P(B^c) > 0$. Then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)(1 - P(B))}$$

It is possible to calculate the probability of B given A in terms of the probability of A given B and the absolute probability of B [39-40].

If A and E are events for which $P(E) \neq 0$, $P(A|E)$ and $P(E|A)$ are related by

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

$$= \frac{P(E|A)P(A)}{P(E)}$$

In particular,

$$\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)}{P(E|B)} \qquad (1)$$

when $P(B) = P(A)$. In order to get the equation (1) by the use of modern axiomatized or the modern expression of the theory, the probability theory becomes insignificant. This theory is one of the major concepts in statistics. The essential fact is expressed by the equation (1), such that for two equal probable causes, the ratio of their probabilities given a particular effect and also the ratio of the probabilities of this effect given the two causes are the same as each other. As a result of making the update of the likelihood of A, at the moment that E has been seen, from $P(A)$ to $P(A|E)$, it is the rule that makes the process to be actual and real.

Let $\{H_k, 1 \leq k \leq n\}$ be a partition of the sample space Ω, and suppose that H_k, $1 \leq k \leq n$, are disjoint sets and that their union equals Ω. Let A be an event. The law of total probability states that

$$P(A) = \sum_{k=1}^{n} P(A|H_k)P(H_k),$$

and Bayes' formula states that

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{k=1}^{n} P(A|H_k)P(H_k)}$$

3.4 Continuous Form of Bayes' Theorem

Thomas Bayes proved the continuous form of equiprobability in 1764. Suppose that x and y are two different random variables with marginal distribution and conditional distribution $g(y)$ and $f(x|y)$, respectively, and the conditional distribution of y given x is defined by

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)\, dy}$$

The mathematician scientists Bayes and Laplace thought that the uncertain model on the parameters Θ could be displayed by a probability distribution π on Θ, called prior distribution, however, this inversion theorem is entirely clear from a probabilistic point of view [38].

The posterior is updated by Bayes theorem from the prior $\pi(\Theta)$ by accounting for the data *x*,

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)\, d\theta}.$$

If Θ is just the only unknown quantity and the data x are available, the posterior distribution entirely describes the uncertainty.

3.5 Bayesian Paradigm

Bayesian paradigm has two primary advantages: (a) Bayesian method follows a simple instruction and recipes when the uncertainty is described by the probability distribution and the statistical inference can be automated, (b) available prior information is included into the statistical model coherently.

The posterior distribution is proportional to the distribution of x conditional upon Θ, it means the prior distribution of Θ, multiplied by the likelihood.

Both parametric statistical model $f(\mathrm{x}|\theta)$ and a prior distribution on the parameters, $\pi(\theta)$, make a Bayesian statistical model.

Bayes' theorem makes the information to be real and actual on θ by the way of taking out the information that is included in the observation x. It also has strongly affected depending on the move that inserts observation (causes) and parameters (effects) on the same level of conceptualization. From the view of statistical

modeling, observations and parameters have a slightly different, due to conditional managements, interplay of their roles.

The famous mathematicians Bayes and Laplace made the Bayesian analysis in particular and modern statistical analysis by the way of imposing this adjustment to the perception of random events. Despite the fact that some of the statisticians accept the probabilistic modeling on the observation(s), they make the boundary between two concepts. Although, in some special cases, the parameter is produced under some actions of many factors that happened at the same time and, therefore, can appear as (partly) random, the parameter cannot be noticed as the outcomes of a random experiment in some cases such as in quantum physics [38].

3.6 Joint Probability Distribution

A model is needed that performs a joint probability distribution for θ and y for making the probability statement about $\mathrm{P}(\theta|y)$. The product of the prior distribution by the data distribution (or sample distribution) makes the joint probability density or mass function

$$\mathrm{P}(\theta, y) = P(\theta)P(y|\theta)$$

By the use of the fundamental property of a conditional probability which is known as the Bayes' rule, the posterior density is defined:

$$P(\theta|y) = \frac{P(\theta, y)}{P(y)} = \frac{P(\theta)P(y|\theta)}{\sum_{\theta} P(\theta)P(y|\theta)}$$

where the summation on the denominator is over all possible values of θ and in the case of continuous θ, the formula is defined by:

$$P(\theta|y) = \frac{P(\theta, y)}{P(y)} = \frac{P(\theta)P(y|\theta)}{\int P(\theta)P(y|\theta)d\theta} \quad (2)$$

The unnormalized posterior density is deduced from (2) by deleting the probability of y which is not based on θ, with fixed y, and considered as a constant:

$$P(\theta|y) \propto P(\theta)P(y|\theta)$$

The formula expresses that the basic task in the Bayesian inference of any specific application is just to evolve the model $P(\theta, y)$, and then do the required calculation to summarize $P(\theta|y)$ in a correct way [36].

3.7 Prior and Posterior Probabilities

Assume that there are n various models denoted as $A_1, A_2, \dots, A_n$. First of all, there is a belief about the credibility and plausibility of these models that are expressed by $P(A_1), P(A_2), \dots, P(A_n)$, and defined as a prior probability in order to express the opinion and belief before or prior to see any data in the model. In the next step, the observed data is denoted by D, and it gives information about the data, and the probabilities of each of them are defined as $P(D|A_1), P(D|A_2), \dots, P(D|A_n)$. These probabilities are called the likelihoods. By the use of Bayes' rule, it will be found out how this data can change the belief about n models after observing the data result, D. The new probability A_i is equivalent to the likelihood multiplied by prior probability.

The updated probability is called the posterior probability due to the fact that it reflects the belief and opinion about the model after data are seen.

$$P(A_i|D) \propto P(A_i)P(D|A_i)$$

Let's consider the first example. Suppose the proportion of an unknown disease in one country is 0.02. Then, the prior probability that a randomly selected subject has the disease is $P_{prior} = 0.02$.

Let's assume now a subject has been positive for the disease. It is known that the accuracy of the test is 97%, and sensitivity of the test is 98%. What is the probability that the subjective has the disease? In another word, what is the conditional probability that a subject has the disease while the test is also positive?

Disease	Positive	Negative	Total
Influenced	0.02*0.98 = 0.0196	0.02*(1-0.98)=0.0004	0.02
Not influenced	0.98*(1-0.97) =0.0294	0.98*(0.97) =0.9506	0.98
Total:	0.0196+0.0294=0.049	0.0004+0.94506=0.951	1.00

The part of really having the disease is 0.0196/0.049 which is equal to 0.40=40%.

Therefore, the posterior (after the test has been performed and known that the test is positive) probability that the person is really having the disease is 0.40.

As the second example let's consider a model of blocking the junk email, and first of all it is needed to recognize whether the email is spam or it contains a word. Let's assume that F is the set of junk emails and E is the set of emails that are containing the word. The purpose of solving this example is to find the probability of F given E that means $P(\mathrm{F}|\mathrm{E})$. By using the Bayes' formula it would be sufficient to have information about:

a) The probability that the word is in the spam message and a non-spam message are by $P(E|F)$ and $P(E|F^c)$, respectively. By the use of statistical analysis of the email, these probabilities can be achieved;

b) The probability of the spam message is shown by $P(\mathrm{F})$. The value of this probability can be obtained on the internet or by statistical analysis of the traffic.

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)(1 - P(F))}.$$

One can calculate the probability of F given E regarding the probability of E given F and the absolute probability of F.

Another example is that assume that there are N coins in a box and just one of them has a head on the both sides. Suppose the coin is taken from the box randomly and

without looking the coin, flipped m times and in all m times, heads come. Calculate the probability that the two headed coin was taken from the box.

It is defined that A_m is the event that a coin is chosen randomly and flipped m times, and heads come. HP_1 - the two headed coin, and HP_2 - the usual coin is the fair hypotheses. Therefore, $P(HP_1) = \frac{1}{N}$ and $P(HP_2) = \frac{N-1}{N}$. So, $P(A_m|HP_1) = 1$ and $P(A_m|HP_2) = 1/2^m$ are the conditional probabilities for any m.

By using the total probability formula

$$P(A_m) = \frac{2^m + N - 1}{2^m N},$$

and

$$P(HP_1|A_m) = \frac{2^m}{2^m + N - 1}$$

This example is about the electronic devices produced in a factory by a machine. The statistical data shows that most of the time the machine works properly 95% and produces 97% correct parts. Sometimes the machine is broken and produces 73% correct parts. During the 8 days, the manager expects the machine works in the following results:

$$C, B, C, C, C, C, C, B.$$

The correct and bad components are illustrated by C and B, respectively. Find the probability that the machine is working properly.

In this example, there exist two different models, the first one is that the machine is working correctly, and another one is that the machine is broken. Based on the given data, the prior probabilities are 0.95 and 0.05 for the cases of working and broken, respectively. The data are the result of the inspection records during the eight days. It calculates the sampling probabilities, and it means the probability of each data results for each case to understand the relationship between the cases and the data.

In the case of working correctly, the probability of correct (C) part is 0.97 and for the bad (B) case is 0.03. So, the conditional probabilities for the two cases are:

$$P(C|Work) = 0.97, \qquad P(B|Work) = 0.03.$$

On the other hand, if the machine is not working or broken, the probabilities are:

$$P(C|Break) = 0.73, \qquad P(B|Break) = 0.27.$$

The Bayes's rule can be used for the set of inspection:

$$Data = \{C, B, C, C, C, C, C, B\}.$$

The probabilities of this data for each of the two different models are the likelihood. Assume that the outcomes are independent from each other:

$$Likelihood(work) = P(\{C, B, C, C, C, C, C, B\}|work)$$

$$= P(C|Work) \times \ldots \times P(B|Work)$$

$$= 0.97 \times 0.03 \times 0.97 \times \ldots \times 0.03$$

$$= 0.0007497.$$

In the similar way, the likelihood of the broken model is

$$Likelihood(Break) = P(\{C, B, C, C, C, C, C, B\}|Break)$$

$$= P(C|Break) \times \ldots \times P(B|Break)$$

$$= 0.73 \times 0.27 \times 0.73 \times \ldots \times 0.27$$

$$= 0.0110323.$$

Now, posterior probability is calculated by likelihood multiplied by prior probability, which is demonstrated in the table 1.

Table 1. Results of posterior probabilities after eight parts

Model	**Prior**	**Likelihood**	**Product**	**Posterior**
1.Working	0.95	0.0007497	0.0007122	0.5635
2.Broken	0.05	0.0110323	0.0005516	0.4365

The calculation shows that the posterior probability of the broken machine is over 0.43.

When there exists sequential data, there is another way of implementing the Bayes' rule. The inspection probabilities of working and broken of the machine are respectively 0.95 and 0.05 before collecting any data. The manager observed the quality of the first session and he/she can update his probability by Bayes' rule. The table 2 shows the results of computation of posterior probabilities after observing the first session.

Table 2. Posterior probabilities after the first session

Model	**Prior**	**Likelihood**	**Product**	**Posterior**
1.Working	0.95	0.97	0.9215	0.9619
2.Broken	0.05	0. 73	0.0365	0.0381

By single observation, it is noticed that the probability is more that 96%. The table 3 shows the data that is related to the observation after the second session.

Table 3. Posterior probabilities after the second session

Model	**Prior**	**Likelihood**	**Product**	**Posterior**
1.Working	0.9619	0.03	0.0288	0.7366
2.Broken	0.0381	0.27	0.0103	0.2634

Now, the probability of working machine is 0.7366.

Continue in this way for eight sessions, and in each session the prior probability is the posterior of the previous session. The table 4 demonstrates the posterior probability in each step after all eight sessions are done.

Table 4. Results of Posterior probability in each step for all eight sessions

	Observation	**P(Work)**	**P(Break)**
1	Prior	0.95	0.05
2	C	0.9619	0.0381
3	B	0.7366	0.2634
4	C	0.7879	0.2121
5	C	0.8316	0.1684
6	C	0.8677	0.1323
7	C	0.8971	0.1029
8	C	0.9206	0.0794
9	C	0.5633	0.4367

The next example is about two competitions in a TV game including a series of various questions. The game ends when the player answers the question correctly. Let's define two players E and F and the probabilities that E and F answer the

question correctly are α and β in the same order. Calculate the probability that E is a winner if:

(a) E answers the first question;

(b) F answers the first one.

In order to solve this example, first of all, we should define:

A set of all feasible infinite sequence of answers which is shown by Ω;

Event E - means E answers the question number one;

Event M - means TV game finishes after the question number one;

Event N - means E wins the game.

The aim of this example is to find the

$$P(N|E) \quad \text{and} \quad P(N|E').$$

By using the total probability theorem, and the partition of $\{M, M'\}$

$$P(N|E) = P(N|E \cap M)P(M|E) + P(N|E \cap M')P(M'|E).$$

Obviously, $P(M|E) = P(E \text{ answers the question number one correctly }) = \alpha$, $P(M'|E) = 1 - \alpha$, and $P(N|E \cap M) = 1$, on the other hand $P(N|E \cap M') = P(N|E')$, therefore

$$P(N|E) = (1 \times \alpha) + \big(P(N|E') \times (1 - \alpha)\big) = \alpha + P(N|E')(1 - \alpha) \qquad (3)$$

Similarly,

$$P(N|E') = P(N|E' \cap M)P(M|E') + P(N|E' \cap M')P(M'|E').$$

So, $P(M|E') = P(F \text{ answers the question number one correctly}) = \beta$, $P(M'|E) = 1 - \beta$, but $P(N|E' \cap M) = 0$. Finally, $P(N|E' \cap M') = P(N|E)$, so

$$P(N|E') = (0 \times \beta) + \big(P(N|E) \times (1 - \beta)\big) = P(N|E)(1 - \beta) \qquad (4)$$

Solving the equations (3) and (4) simultaneously for parts (a) and (b)

$$P(N|E) = \frac{\alpha}{1 - (1 - \alpha)(1 - \beta)}, \quad P(N|E') = \frac{(1 - \beta)\alpha}{1 - (1 - \alpha)(1 - \beta)}.$$

Notice that for any events A_1 and A_2

$$P(A_1'|A_2) = 1 - P(A_1|A_2),$$

But not necessarily that

$$P(A_1|A_2') = 1 - P(A_1|A_2).$$

The next example is for a subset of the survey including data on salary and education for a sample of females over 30 years old. Assume that $S = \{A_1, A_2, A_3, A_4\}$ is a set of events that randomly selected woman from S, each of the A_i has 25% in terms of salary. Therefore, by definition

$$\{P(A_1), P(A_2), P(A_3), P(A_4)\} = \{0.25,0.25,0.25,0.25\}$$

The set S is a partition, so as a result, the summation of these probabilities must be equal to one. Now, suppose that one chooses a sample randomly from the survey of college education and this event is shown by F. Based on the survey,

$$\{P(F|A_1), P(F|A_2), P(F|A_3), P(F|A_4)\} = \{0.11,0.19,0.31,0.53\}.$$

These probabilities are demonstrated the proportions of college-educated people in the four different salary subpopulations A_1, A_2, A_3 and A_4. By using the Bayes' rule, it is able to calculate the salary distribution of the college-educated population.

$$\{P(A_1|F), P(A_2|F), P(A_3|F), P(A_4|F)\} = \{0.09,0.17,0.24,0.47\}$$

It is illustrated that the salary distribution for persons in the college degree population is different from $\{0.25,0.25,0.25,0.25\}$, the distribution for the general

population. Moreover, the total summation of conditional probabilities of the event in the partition equals to one.

Notice that in Bayesian inference, the set $\{A_1, A_2, A_3, \dots, A_k\}$ often refers to the state of nature or disjoint hypothesis, and F shows the result of the survey, experiment or study. By calculating the following ratio, hypotheses can be compared post-experimentally

$$\frac{P(A_i|F)}{P(A_j|F)} = \frac{P(F|A_i)P(A_i)/P(F)}{P(F|A_j)P(A_j)/P(F)}$$

$$= \frac{P(F|A_i)P(A_i)}{P(F|A_j)P(A_j)}$$

$$= \frac{P(F|A_i)}{P(F|A_j)} \times \frac{P(A_i)}{P(A_j)}$$

$$= Bayes\ factor \times Prior\ beliefs$$

This calculation demonstrates that Bayes' rule says how the first beliefs change after observing data, and it does not mention anything about what individual's beliefs should be after observing the data.

Let's consider another example. Suppose the table 5 illustrates the joint distribution of occupational categories of son and father.

Table 5. Joint distribution of occupational categories

	Son's occupation				
Fathers' occupation	Accountant	Architecture	Manager	Nurse	Teacher
Accountant	0.054	0.105	0.093	0.024	0.054
Architecture	0.006	0.347	0.198	0.099	0.213
Manager	0.002	0.138	0.197	0.067	0.176
Nurse	0.002	0.036	0.038	0.020	0.102
Teacher	0.003	0.095	0.105	0.141	0.429

Now suppose X_1 be the fathers' job and X_2 is the son's job. Then

$$P(X_2 = teacher | X_1 = accountant) = \frac{P(X_2 = teacher \cap X_1 = accountant)}{P(X_1 = accountant)}$$

$$= \frac{0.054}{0.054 + 0.105 + 0.093 + 0.024 + 0.054}$$

$$= 0.164$$

In the next example it is assumed that a box includes b blue beads and g green beads. It is needed to calculate the probability of picking one blue bead in one selection only. Assume a person does not have any information about the color of the first

bead. Without replacing it, the person picks another one. We calculate the probability of picking a blue bead in the second selection.

The certain event is that two beads are picked and each of them can be either blue or green. Therefore, $\Omega = \{0,1\}^2$ where the jth component of $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2) \in \Omega$ is zero if the jth picked is blue and 1 if the jth picked is green. So, the event of the first bead is blue, and is associated with

$$B_1 = \{(0,0), (0,1)\}$$

Thus $P(B_1) = b/b + g$. In fact, when the first bead is "randomly" selected, the probability of success is $b/b + g$. Similarly, if the first selected bead is green, the event is associated with

$$G_1 = \{(1,0), (1,1)\} = \Omega \backslash B_1$$

Therefore $P(G_1) = 1 - P(B_1) = g/g + b$. Now calculate that the probability of the second bead is blue, and this event is associated with

$$B_2 = \{(0,0), (1,0)\}$$

After picking the first beads, two cases may happen. Assume the first one is blue, then there exist $b - 1$ blue beads and g green beads in the box,

$$P(B_2|B_1) = \frac{b-1}{b+g-1}$$

and similarly

$$P(B_2|G_1) = \frac{b}{b+g-1}.$$

Then the law of total probability shows that

$$\begin{aligned} P(B_2) &= P(B_2 \cap B_1) + P(B_2 \cap G_1) \\ &= P(B_2|B_1)P(B_1) + P(B_2|G_1)P(G_1) \\ &= \frac{b-1}{b+g-1}\,\frac{b}{b+g} + \frac{b}{b+g-1}\,\frac{g}{b+g} = \frac{b}{b+g}. \end{aligned}$$

Without any information about the color of the first selected bead, the probability of selecting the bead with the same color in the second selection remains the same.

Chapter 4

SUBJECTIVE AND OBJECTIVE PROBABILITIES

4.1 Properties of Subjective and Objective Probabilities

The terms subjective and objective are used in several ways; however, there is another interpretation of these terms in probability theory. There exists a distinction between beliefs of scientists about a phenomenon before collecting the data, and beliefs of scientists after the data have been gathered and analyzed. The specific things are defined to be only a subjective reality, and these cases are formed depending on beliefs and views of the human mind. From another point of view, the objective fact is used in some cases that are outside the minds of the human being, and also it is defined without considering of whether a person perceives it or not. For instance, the view and opinion of a person about a subject such as a social issue and political are just a personal belief that has only a subjective reality. However, the starting point considering external reality is that the sun and the moon exist without considering of human being perceived them. As a simple example, it is to state that an existence of all laws and rules of nature in the world is not depending on human belief. Human intuition and sense are involved in anything that is seen or can be measured by human beings. After observing or perceiving entities by a human in the objective word, either by measuring instrument or even by the senses and feelings, the consequence of measurements, which is called data, are described by a human in

a subjective way that demonstrates the personals' experience, comprehending, and preliminary feeling and also beliefs about the entities, the things, the objects and phenomenon, or make being measured. The comprehension of these data is influenced by the human's state of sense. The term "subjective" is used to refer to preexisting beliefs or views about entities. Broadly speaking the term "subjective" is used to indicate a human's belief, and intuition. The views about the hypothesis that a person held prior may be different across individuals, this view or belief is called subjective.

Let's try to clarify the definition "objective" in an experimental way to understand how scientists describe this definition. By this definition, both scientists and science are objectives in the following sense. A hypothesis is evolved, and the specialists design a study to test the hypothesis. The data might be gathered by performing observation carefully in a non-experimental setting or even by designed experiment. After collecting the data, the scientists evaluate the results, consequences, and the implications for the hypothesis. The scientists describe the study of publication if the outcomes support the hypothesis. On the other hand, the scientist rejects the hypothesis and either reviews the hypothesis considering the new finding and repeats everything, or continues to other concerns [41].

The terms "subjective" and "objective" can be also used in another sense. It is merely defined that subjective probability refers to the human's degree of belief individually about the event to happen. However, objective probability refers to the numerical probability, mathematical chance of some event happening. The definition of

objectivity that is used by some philosophers needs to be testable by anyone. It is believed that such consistency of explanation is rare. Different people have different ideas because of their preexisting views, and induce them to analyze the data in different forms.

There is a quote of Ian Hacking who mentioned that one of the most intriguing aspects of the subjectivist theory is the use of a point about Bayes' theorem. Hacking believed that observers have different interpretation of exactly same data; however, eventually with an adequately large number of experiments or trials, the differing prior views about the same data by the different observers will mostly disappear [42].

It will be enlightening to commence the discussion of the term subjectivity in scientific methodology with mentioning an example that demonstrates how various observers unwittingly bring the personal beliefs and ideas.

During the time, it is always interested to find the probability of some events in future observations such as it will be raining during the next hour, or the probability that a candidate will win in the next election. These kinds of events will happen only once; therefore, it is needed to expand the definition of probability in order to cover all such problems and situations.

Finally, probability theory is defined to be a personal degree of individual belief about any unknown quantity or some proposition. The term subjective probability is defined as individual beliefs or personal probability. Additionally the subjective

probability implies that this concept is based on a personal belief. The belief might be that the mathematical probability or the long-range relative frequency.

In order to clarify the definition of subjective probability, let's assume the proposition "William Henry Harrison was the eighth president of United States" - this is either true of false. If a person does not know about this proposition, so for the person it will be uncertain, but this person has a degree of belief about its correctness. This person may think there is 50% chance that it is correct, on the other hand, based on some extra information such as the history of United States this person may think that there is 80% chance that this proposition is true. Personal belief on 80% chance of correctness of the proposition is equivalent to the situation that someone randomly selects the black ball out of a box containing 80 black balls and 20 white balls (William Henry Harrison was ninth president of the United States). The similar arguments about any unknown quantity can be done like the hair color of the next person that might be seen accidentally in the street. In the first step, scientists might have a belief that the probability in a special theory has 50% chance to be true. After that, the scientist may collect related observational data and add more evidence about the correctness of the theory. In the last step, the scientist might say that the chance of correctness of probability has increased to 75%. In fact, the scientist cannot be certain 100% about the hypothesis [41].

The probabilities in these examples such as the probability of raining in the next one hour, the probability of winning of a candidate in an election, or the probability of winning of a driver in a race can be considered to be subjective or at least include a

subjective component. However, there is another probability which seems, at least at first sight, to be totally objective. For instance, in rolling the dice the probability of each side is 1/6, and in such case this is completely objective regardless of an individual belief or subjectivity. Another example is to calculate the probability of an isotope of uranium disintegrating in one year. This calculation is completely based on the quantities of books related to physics, not the personal belief. Subjective theory cannot support these kinds of examples [43].

In 1763, Thomas Bayes defined a method for making statistical inferences that broaden earlier comprehending and understanding of the phenomenon. This method joins the earlier comprehending with currently measured data to update the scientific belief or subjective probability of the experiment. The previous experiment and comprehending are identified as prior belief and the updated prior belief, which is given by combining the prior belief with a new observation, is identified as posterior belief. To clarify, posterior belief can be defined as a belief that is held after collecting the current data and also having examined those data considering how well they confirm the preliminary data. This inferential process which is suggested by Thomas Bayes is called Bayesian inference. To find subjective probability for some events, unknown quantity, or proposition, based on this method, is needed to multiply prior belief by an appropriate summary of the observational data. Therefore, Bayes indicated that all scientific inferences include two parts: one of them based on the subjective understanding and information (prior knowledge), and another part based on scientific experiment and observation [41].

Probability that is used in all statistical methods is subjective in the sense of depending on mathematical idealizations of the world. Bayesian analysis especially mentions that the term subjective due to its dependence on the prior distribution, however, in some problems, scientific belief and judgment is fundamental in order to define the "likelihood" and "prior" parts of the model. Here is a general principle: when there is duplication, there is a scope to calculate a probability distribution and therefore constructing the analysis in more objective form. If we are dealing with a replication of the whole experiment several times, the parameter of the prior distribution could be calculated from data. However, some elements requiring scientific idea still remain, remarkably the selection of data in the analysis, for the distribution the parametric forms is assumed [36].

In the logical explanation, the probability of F given E is identified with a reasonable degree of belief which a person who had evidence E would give to F. The reasonable level of belief is supposed to be the same for all rational individuals. The subjective explanation of probability rejects the hypothesis of rationality going to consensus. Based on the definition of the subjective theory, although different individuals such as person A, person B, and person C, all perfectly logical and having the same evidence E, might have various degrees of belief in F yet. Thus, probability is defined as the degree of opinion and the belief of an individual [43].

One of the pioneers of the subjective theory is de Finetti whose first writing was in French and Italian at the beginning of 1930. The idea of probability via expectation

was introduced by de Finetti that was called “prevision” or “subjective” probability [44].

De Finetti’s treatise on probability theory started with the statement “Probability does not exist”, which means the probability exists only subjectively rather than in sense of objectively. Moreover, he thought that objective probability does not exist without depending on the human mind and belief. According to this view, probabilities are built as degree of personal beliefs. The theory of subjective probability is mostly attributed to three mathematicians de Finetti, Ramsey, and Savage. They proposed the behavior in the definition of probability by mentioning the example of betting rates. Betting rate shows the personal probability or individual belief of the human being that is the only probability that really exists. In de Finetti’s theory, he believed that bets are just for money therefore an individual probability of events directly affects the money that the person is willing to win. He introduced the notation of “Pr” which is interchangeable by Probability, Prevision, and Price [45].

The objective view of probability includes three major principles mentioned below:

1) Probability: An individual’s degree of belief ought to be represented by probabilities. Therefore, for instance, a personal degree of belief that a certain candidate will win in the next election and will not win should sum to 1;

2) Calibration: An individual’s degree of belief has to be properly constrained by empirical evidence. If the person thinks that today is rainy with probability between

80% and 90%, then the person's degree of belief about tomorrow should also lie in the interval [0.8, 0.9];

3) Equivocation: An individual's degree of belief has to be as middling as these restrictions permit. In the previous example, it should be equivocated as far as possible between rainy weather or not. It should be believed that tomorrow is rainy with the probability degree of 80%.

These days objective theory of probability is a popular topic in statistics, physics, engineering and also artificial intelligence, especially in machine learning and language processing [46].

4.2 Generalized Bayes' Theorem

In the use of probability theory, Bayesian rule is a key point for the diagnosis process. Assume two spaces, Θ and X represent space of diseases and symptoms, respectively.

Given the conditional probability shown by $P_{\mathrm{X}}(x|\theta_i)$ of observing x which is a subset of X in each disease class θ_i that belongs to Θ, and a prior probability P_{Θ} over Θ, calculates a posterior probability $P_{\Theta}(\theta|x)$ over Θ that the ill person is in a disease class in θ given the symptom x has been seen. By Bayes rule,

$$P_{\Theta}(\theta_i|x) = \frac{P_{\mathrm{X}}(x|\theta_i)P_{\Theta}(\theta_i)}{\sum_j P_{\mathrm{X}}(x|\theta_j)P_{\Theta}(\theta_j)}$$

In generalized Bayes' theorem, it is assumed that each disease with class θ_i has a belief function to be shown by $bel_X(.|\theta_i)$ over X and it represents the personal belief about if the person has a disease that symptom can be observed. Let's define the a priori belief by bel_Θ which represents a personal belief about the disease class to which the ill person belongs. Assume bel_Θ means that there is no a priori personal belief about the disease. The formula (5) shows a posterior plausibility function over Θ:

$$pl_\Theta(\theta|x) = 1 - \prod_{\theta_i \in \theta}\big(1 - pl_X(x|\theta_i)\big), where\ x \subseteq X \qquad (5)$$

One of the significant properties of generalized Bayes' theorem is dealing with the case when there exist two different independent observations. To clarify this property, assume two symptoms spaces such as X and Y. Therefore, bel_X and bel_Y are represented an individual's belief on X and Y. Based on this property, it is assumed that the symptoms are not depending on each other within each disease class. Being independence indicates that if a person had knowledge about which disease is related to the observation of a symptom could not affect the personal belief of other symptoms. The meaning of the independence property is that the conditional joint belief over the space X × Y given θ_i is

$$pl_{X\times Y}(x \cap y|\theta_i) = pl_X(x|\theta_i)pl_Y(y|\theta_i)$$

There are observations about the symptoms $x \subseteq X$ and $y \subseteq Y$, and afterwards $bel_\Theta(.|x)$ and $bel_\Theta(.|y)$ by using the generalized Bayes' theorem are defined. In

order to find the personal belief $bel_{\Theta}(.|x,y)$ about Θ given both of symptoms x and y and combine these two beliefs together, Dempster's rule should be used. On the other hand, by using the generalized Bayes' theorem on $bel_{X\times Y}(.|\theta_i)$, it is possible to calculate $bel_{\Theta}(.|x,y)$. Using the Dempster's rule and generalized Bayes' theorem, the same results are reached, as it should be.

By using the generalized Bayes' theorem, it is allowed to enlarge the disease domain Θ. It is obvious that, in that class, the individual's belief about the symptoms is vacuous. However, by using generalized Bayes' theorem, it is allowed to calculate a posterior belief that whether the patient belongs to a new class or not. The personal belief about the "discovery", which is impossible by using the probabilistic framework, is computed.

Suppose $\Theta = \{\theta_1, \theta_2, \theta_\omega\}$ is a set of diseases that θ_1 and θ_2 are two well-known diseases which mean that there exist some beliefs about which symptoms reveal when θ_1 or θ_2 holds. The compliment of $\{\theta_1, \theta_2\}$ is defined by θ_ω that means relative to all possible diseases plus the other diseases that are unknown yet. Therefore, in this example the personal belief on the symptoms can be vacuous [47].

Chapter 5

CONCLUSION

In this thesis, Bayes probability is considered that studies the relation between probability and uncertainty. Probability is the science of uncertainty that prepare mathematical rules to comprehend and analyze the uncertain situation. Probability cannot tell about next week's stock prices or even tomorrow's weather, instead it gives framework to work with imprecise information to make a sensible decision based on previous knowledge and information. Uncertainty has been with human forever, however, the mathematical theory of probability commenced in the seventeenth century. Thomas Bayes is one of the mathematicians who introduced the probability theory. Bayes theory is a significant method in probability theory for the aim of comprehending the prepared observations, what is actually known, and some other information. To calculate the probability of an unknown hypothesis, the Bayesian probability identifies and recognizes some prior probabilities that will be updated in next steps, in the light of related data.

Probability theory plays a noticeable role in various applications of science and technology. In some cases, there exists the risk such as risk of losing money. If the problem people encounter, individually or collectively, become more sophisticated, people need to improve and evolve their rough understanding of the idea of

probability to form an exact and precise approach. This is one of the reasons that probability theory has been altered to a mathematical and scientific subjects. In fact, probability theory is prepared a precise comprehending of uncertainty which is helpful for making prediction, optimal decision, and estimating risk in everyday life. Probability is dealing with quantifying or measuring uncertainty.

In this thesis, the definition of probability starts with a sample space. The sample space can be any set that includes all possible outcomes of some unknown situations or experiments such as $\{rainy, snowy, sunny, cloudy\}$ for predicting the future weather. A probability model needs a collection of events, which are subsets of the sample space to which probabilities can be allocated. Finally, the probability model needs a probability measure which is essential in the model and represented by P. Another important point is the conditional probability.

The differences between the terms “subjective” and “objective” in probability theory are presented. The term subjective probability mentions the human’s degree of belief is individual about the chance of some events happen. On the other hand, objective probability is about the numerical probability, mathematical chance of some events happen. In the Bayesian analysis, the term “subjective” is independent from the prior distribution, however, in some problems, human belief is required for specifying the likelihood and prior parts of the model. The first scientist who mentioned the term of subjectivity was de Fenetti, who believed that the probability does not exist in the sense of objectivity. On the other hand, nowadays objective theory of probability is a

popular topic in various types of majors such as statistics, physics, artificial intelligence etc.

A general framework for reasoning with uncertainty that is using belief function is the transferable belief model. In the particular topic of interest, the generalized Bayes' theorem is an expansion of Bayes' theorem that is probability measures are replaced by belief functions and there exist no prior experiment. Lately, applications of the generalized Bayes' theorem have been limited, mostly due to the lack of methods for making belief functions from observation data. However, these days by using this method and also combination rules to merge partly overlapping item of evidence expand the application of the transferable belief model.

REFERENCES

[1] Ehsan Goodarzi, Mina Ziaei, & Lee Teang Shui. (2013). Introduction to Risk and Uncertainty in Hydrosystem Engineering, *Volume 22.*

[2] Douglas W. Hubbard. (2010). How to Measure Anything: Finding the Value of Intangibles in Business, 2nd Edition.

[3] SAS Institute Inc. (2009). *SAS/STAT ® 9.2User's Guide, Second Edition*. Cary, NC: SAS Institute Inc.

[4] Malin Adolfson, Stefan Laséen**,** Jesper Lindé, & Mattias Villani. (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics 72,* pp. 481-511.

[5] Thomas A. Lubik, & Frank Schorfheide. (2007). Do central banks respond to exchange rate movements? A structural investigation. *Journal of Monetary Economics 54,* pp. 1069-1087.

[6] Jesus Fernandez Villaverde, & Juan F. R. Ramirez. (2004). Comparing dynamic equilibrium models to data: a Bayesian approach. *Journal of Econometrics 123,* pp. 153-187.

[7] Subal C. Kumbhakar, & Efthymios G. Tsionas. (2005). Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach. *Journal of Econometrics 126,* pp. 355-384.

[8] Renna Jiang, Puneet Manchanda, & Peter E. Rossi. (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics 149,* pp. 136-148.

[9] Wayne S. DeSarbo, Youngchan Kim, & Duncan Fong. (1999). A Bayesian multidimensional scaling procedure for the spatial analysis of revealed choice data. *Journal of Econometrics 89,* pp. 79-108.

[10] Kai Li. (1999). Exchange rate target zone model: A Bayesian evaluation. *Journal of Applied Econometricsics, 14,* pp. 461-490.

[11] Gary Koop, Dale J. Poirier, & Justin Tobias. (2005). Semiparametric Bayesian inference in multiple equation models. *Journal of Applied Econometrics, 20,* pp. 723-747.

[12] Frank Smets, & Raf Wouters. (2005). Comparing shocks and frictions in US and euro area business cycles: A Bayesian DSGE approach. *Journal of Applied Econometrics, 20,* pp. 161-183.

[13] Theo S. Eicher, Chris Papageorgiou, & Adrian E. Raftery. (2011). Default priors and performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics, 26,* pp. 30-55.

[14] Brendan Kline, & Justin L. Tobias. (2008). The wages of BMI: Bayesian analysis of a skewed treatment-response model with nonparametric endogeneity. *Journal of Applied Econometrics, 23,* pp. 767-793.

[15] Chun Liu, & John M. Maheu. (2009). Forecasting realized volatility: A Bayesian model-averaging approach. *Journal of Applied Econometrics, 24,* pp. 709-733.

[16] Miguel A. Jua´rez, & Mark F. J. Steel. (2010). Non-Gaussian dynamic Bayesian modeling for panel data. *Journal of Applied Econometrics, 25,* pp. 1128-1154.

[17] Van H. K. Dijk, & Kloek. (1978). Bayesian estimation of equation system parameters: An application of integration by Monte Carlo. *Econometrica, Vol. 46, No. 1,* pp. 1-19.

[18] Joel Huber. (2001). On the similarity of classical and Bayesian estimates of individual mean partworhs. *Marketing letters 12:3,* pp. 259-269.

[19] Dries F. Benoit, & Dirk Van Den Poel. (2012). Binary quantile regression: A Bayesian approach based on the asymmetric Laplace distribution. *Journal of Applied Econometrics, 27,* pp. 1174–1188.

[20] Zhihai Wang, & Geoffrey I. Webb. (2001). Comparison of Lazy Bayesian Rule and Tree-Augmented Bayesian Learning. *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002),* pp. 490-497.

[21] John A. Stankovic. (1985). An Application of Bayesian Decision Theory to Decentralized Control of Job Scheduling. *IEEE Transactions on Computers, Vol. C-34, Issue 2,* pp. 117-130.

[22] Ashay Kadam, & Peter Lenk. (2008). Bayesian inference for issuer heterogeneity in credit ratings migration. *Journal of Banking & Finance 32,* pp. 2267-2274.

[23] Alon Brav. (2000). Inference in Long-Horizon Event Studies: A Bayesian Approach with Application to Initial Public Offerings. *The Journal of Finance, Vol. Lv, No. 5*.

[24] Wayne S. De Sarbo, Youngchan Kim, Michel Wedel, & Duncan K. H. Fong. (1998). A Bayesian approach to the spatial representation of market structure from consumer choice data. *European Journal of Operational Research 111,* pp. 285-305.

[25] Dalla L. Valle, & Giudici. (2008). A Bayesian approach to estimate the marginal loss distributions in operational risk management. *Computational Statistics & Data Analysis 52,* pp. 3107-3127.

[26] Pavel V. Shevchenko, & Mario V. Wüthrich. (2006). The Structural Modelling of Operational Risk via Bayesian inference: Combining Loss Data with Expert Opinions. *The Journal of Operational Risk 1(3),* pp. 3-26.

[27] Catalina Stefanescu, Radu Tunaru, & Stuart Turnbull. (2009). The credit rating process and estimation of transition probabilities: A Bayesian approach. *Journal of Empirical Finance 16,* pp. 216-234.

[28] Linhua Li, Li Zhu, & Daniel Z. Sui. (2007). A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography 15,* pp. 274-285.

[29] Maria Anna Di Lucca, Alessandra Guglielmi, Peter Muller, & Fernando A. Quintana. (2013). A Simple Class of Bayesian Nonparametric Autoregression Models. *Bayesian Analysis 8, Number 1,* pp. 63-88.

[30] John R. Bryant, & Patrick J. Graham. (2013). Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources. *Bayesian Analysis 8, Number 3,* pp. 591-622.

[31] Mehran Sahami, Susan Dumais, David Heckerman, & Eric Horvitz. A Bayesian approach to filtering junk E-mail, Stanford University.

[32] Richard O. Duda, Peter E. Hart, & Nils J. Nilsson. Subjective Bayesian methods for rule-based inference systems. Stanford Research Institute Menlo Park, California.

[33] Chung ki Min, & Arnold Zellner. (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics, Volume 56, Issues 1-2,* pp. 89-118.

[34] Arnold Zelner, Chansik Hong, & Chung Ki Min. (1991). Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics, Volume 49, Issues 1-2,* pp. 275-304.

[35] Peter D. Hoff. (2009). A First Course in Bayesian Statistical Methods. Springer, p. 270.

[36] Andrew Gelman, John B. Carlin, Hal S. Stern, & Donald B. Rubin. (2013). Bayesian Data Analysis. Second Edition, Chapman & Hall/CRC Texts in Statistical Science, p. 668.

[37] Paulos John Allen. (2011). The Mathematics of Changing Your Mind.

[38] Christian P. Robert. (2008). The Bayesian Choice: From Decision Theoretic Foundations to Computational Implementation. Springer, p. 602.

[39] Giuseppe Modica, & Laura Poggiolini. (2013). A First Course in Probability and Markov Chains. John Wiley & Sons, p. 334.

[40] Allan Gut. (2009). An Intermediate Course in Probability, Second Edition. Springer, p. 302.

[41] James S. Press, & Judith M. Tanur. (2001). The Subjectivity of Scientists and the Bayesian Approach. John Wiley & Sons, p. 274.

[42] Ian Hacking. (1976). Logic of statistical inference. Cambridge University Press.

[43] Donald Gillies. (2000). Philosophical theories of Probabilities. Psychology Press, p. 223.

[44] Bruno de Finetti. (1975). Theory of probability. John Wiley & Sons, Volume II.

[45] Robert F. Nau. (2001). De Finetti was right: Probability does not exist. *Theory and Decision 51*, pp. 89-124.

[46] Jon Williamson. (2011). Objective Bayesianism, Bayesian conditionalization and voluntarism. *Synthese, Volume 178, Issue 1,* pp. 67-85.

[47] Dov M. Gabbay, & Philippe Smets. (1998). Quantified Representation of Uncertainty and Imprecision. Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1, Springer, p. 477.