The Development of Bayesian Theory and Its Applications in Business and Bioinformatics

Yifei Zhang¹

Art & Science School, Washington University in St.Louis
Campus Box 1143, One Brookings Drive, St. Louis, MO 63130, United States.
E-mail: 690272711@qq.com

Abstract. Bayesian Theory originated from an Essay of a British mathematician named Thomas Bayes in 1763, and after its development in 20th century, Bayesian Statistics has been taking a significant part in statistical study of all fields. Due to the recent breakthrough of high-dimensional integral, Bayesian Statistics has been improved and perfected, and now it can be used to solve problems that Classical Statistics failed to solve. This paper summarizes Bayesian Statistics’ history, concepts and applications, which are illustrated in five parts: the history of Bayesian Statistics, the weakness of Classical Statistics, Bayesian Theory and its development and applications. The first two parts make a comparison between Bayesian Statistics and Classical Statistics in a macroscopic aspect. And the last three parts focus on Bayesian Theory in specific -- from introducing some particular Bayesian Statistics’ concepts to listing their development and finally their applications.

1. Introduction
In the past few centuries, Classical Statistics’ theories have been unceasingly developed and improved with consistent effort of plenty of statisticians. These theories were accepted and adopted by many people in different fields and were used to solve a large amount of practical problems. However, in the last fifty years, some drawbacks of Classical Statistics have been exposed in practice. Bayesian Statistics provided the possibility of tackling those problems and therefore more and more statisticians started to pay attention to Bayesian Statistics.

Bayesian Statistics originated from Thomas Bayes’ essay in 1763 named An Essay Towards Solving a Problem in the Doctrine of Chances [1]. In the first half of his essay, Bayes proposed a list of concepts including ‘expectation’, ‘independent event’, ‘conditional probability’, etc. And Bayes brought up a question in the second half of his essay. He supposed to throw two balls (ball O and ball W) onto the square table ABCD, and he wanted to know the probability of Ball W falls onto a predetermined area (event M) after multiple independent trials. Differing from Classical Statistics, Bayes firstly predicted the probability interval of event M (Proposition 9, Bayes) and then used the result of his experiment to calculate the probability that his guess is correct (chance to be right). His experiment of ‘throwing balls’ introduced the idea of prior distribution.

The essay of Bayes was not published until two years after he died. It was first announced at UK’s Royal Society by a scholar named Price. But the short essay of Bayes did not arouse interests of statistician at first. Due to the consistent breakthrough of Classical Statistics, the main focus of statisticians was on Classic school, and Bayes’ experiment was overlooked at that time, until the turning
point emerged in mid-twentieth century when Classical Statistics experienced a bottleneck. Classical Statistics relies heavily on samples, and a precise theoretical study requires a large size of samples in most cases, yet it means high time cost of statistics workers especially those in application area. When statistician realized the limitation of Classical Statistics, they started to pay attention to Bayesian Statistics that is based on prior distribution and only needs a relatively small size of sample. Bayes’ essay was also republished on Biometrika [2].

In late-twentieth century, Bayesian Statistics reached its prime, and Bayesian statisticians solved a lot of practical problems through researches. Firstly, R. A. Fisher used induction to overcome difficulties in calculating likelihood ratio and maximum posterior estimation [3], and in 1961, Jeffrey proposed the general way of determining non-informative prior distribution, and his book is also authoritative. Then in 1971, Lindley offered inductions and explanations for some results of Classical Statistics by using Bayesian methods. And in 1985, J.O Berger published the book Statistical Decision Theory and Bayesian Analysis [4], which is considered a classic writing of Bayesian Statistics. The limitation of Classical Statistics and the great effort of Bayesian Statisticians are the two reasons why Bayesian Statistics got the same academic status as Classical Statistics. As a result, Bayesian Theory is widely implemented in various kinds of domains nowadays.

2. The shortcomings of Classical Statistics

1) There exists a huge discrepancy between Classical Statistics and Bayesian Statistics in both statistical and philosophical sense. Analyzing data and studying statistics itself are the processes of using samples to predict population. Classical Frequency Statistics uses samples information to predict population, but Bayesian Statistics proposed the concept of prior information and uses both sample information and prior information to predict population. The difference in thought patterns between the two schools is the fundamental distinction. Classical Statistics believes that prior information is subjective, and it is not based on samples itself. However, Bayesian Statistics argues that subjective probability is a concept widely used by people, and abandoning prior information which could have contributed to statistical study will bring unnecessary troubles in predicting population.

2) Bayesian School doubts the probability-based frequency interpretations (interval estimation, point estimation, hypothesis test, etc.) of Classical Statistics. Bayesian School argues the precision and degree of reliability of the interpretations are the average of large amount of repetitive experiments, and they are pre-decided before drawing the sample, which are called ‘ex ante precision’ and ‘ex ante accuracy’. Bayesian scholars believe that precision and reliability should be related to conditions of samples, and they should be called as ‘ex post precision’ and ‘ex post accuracy’.

3) Data analysis of Classical Statistics is based on large samples, and as a result it has limitations when the sample size is small, drawing sample is difficult or costs of sampling is high, for example, in the case of quality testing of industrial assemble lines. Moreover, in some statistical applications, event may only appear once, and the event is nearly impossible to repeat under same conditions, for instance, forecast and prediction of natural calamities. Under these circumstances, Bayesian Statistics which does not rely on sample size is more effective and applicable.

3. Bayesian Statistical Theory

3.1 The concepts of Bayesian Statistics

Bayesian Statistic theory was first brought up by Thomas Bayes in mid-eighteenth century [5]. He proposed that combining prior distribution of the parameter with likelihood function will get posterior distribution of the parameter, which is the famous Bayes formula. The posterior distribution from using this formula is the fundament of statistical description and deduction of Bayesian Statistics. Therefore, compared with Classical Statistics, Bayesian Statistics not only makes use of population information and sample information, but also utilizes prior information. The essay of Thomas Bayes creates a new way of thinking in statistics deduction. [6]
3.2 Bayes formula
Bayes formula is the most basic rule of Bayesian statistics. It is written as [7]:

\[ P(A) = P \left( \sum_{i=1}^{n} AB_i \right) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \]

The concise formula contains a lot of information. It is a mathematical formula, but it also conveys a philosophical ideology that does not need to be proved in a mathematical way. By inferring characteristics of people from their behavior, or inferring essence of things from their appearances, the process is subjective deduction. Like the concept of ‘chance to be right’, which has been mentioned in the introduction, Bayes formula estimates the probability of true essence of things by using the observed counts of events that are related to the essence.

4. The development of Bayesian Theory

4.1 Bayesian Network
Bayesian network is also called belief network, which is a kind of probabilistic network used to describe statistical relationship among multiple variables [5]. Bayesian network is a graphical network built on the basis of probability theory, and the network gives an intuitive presentation of the complex multivariable relationship. Bayesian network was first proposed to solve non-qualitative problems, and it is favored by scholars because it can express the large-scale multiple-relational problems in a simple and straightforward way. Now Bayesian network has received attention from research works in many fields, such as data mining, artificial intelligence and pattern recognition. The structure of Bayesian network is shown in the figure 1 [8].

![Bayesian network’s structure](image)

In this directed acyclic graph S, every node represents a variable, attribute, state, object, proposition or other entity. Arcs between nodes represent relationship between variables, and all the nodes pointed to node X are called parent nodes of node X. Bayesian network is a systematics expression based on probabilistic rules. In addition to Bayesian network, this structure is also called causal network, probabilistic causal network and belief network.

4.2 Bayesian Decision Theory
Statistical decision is a process of making choices under given information and uncertainty (noted by \( \theta \)). Classical Statistics deduces a conclusion from sample information, but Bayesian Statistics uses sample information, loss function and prior information to make a decision. The earliest book that gives a systematic explanation of Bayesian decision theory is Statistical Decision Theory and Bayesian Analysis by J. O. Berger in 1985. In his book, Berger proposes concepts of posterior expected loss, Bayes risk and Bayes solution, which established the basis for Bayesian decision theory’s study.

After its emergence, Bayesian decision theory was first widely used in biological study. In 1996, it was used to explain perceptual bias in psychophysical [9]. The scene of visual perception is prior
information, a likelihood function explains how the scene generates the image, and a decision rule determines the scene interpretation. In 2006, Bayesian decision theory was found similar to human’s sensorimotor control [10]. Biological signals received by sensory perceptual system and motor system are corrupted by variability or noise, and nervous system solves such estimation and decision problem in a way close to that predicted by Bayesian decision theory.

Beyond which, Bayesian decision theory is also used in market analysis and management. Including game theory, customer selection and arrangement of work schedule.

5. Applications of Bayesian Theory

5.1 Applications in business administration
Cost and return are two of the most important factors in enterprise operation, and statistical theory is crucial in cost savings. In staff recruitment, department of human resources is required to consider economic costs and time costs of both screening and training. Therefore, making prediction of job seeker’s performance in future is very important. In some cases, employers use minimum error Bayes decision theory in screening of labor force. Employers take advantage of the low cost of psychological tests and the correlation psychological tests and performance assessments, to maximize profits [11].

Bayesian theory is also extensively applied in enterprises with assemble lines. Sample size of quality inspection of assemble lines is directly related to inspection costs. Bayesian theory is used in calculating process capability index due to its low requirement of sample size.

5.2 Applications in image recognition
Face recognition is an important technique used in government agencies and social security. Accuracy and speed are the keys of the technique. In the face recognition process, face database serves as sample, and ‘face’, ‘non-face’ are the two results of decision. Making decisions based on minimum error Bayes decision theory will effectively increase accuracy and speed of face recognition. During the decision process, sample is not drawn from data base, but from one-direction vectors that are filtered by BP neural network. According to the classification of Bayesian decision theory, researchers can make a more accurate estimation of threshold value of discriminating face/ non-face. In the meantime of increasing accuracy, decision time can also be lowered by setting up hidden nodes [12].

5.3 Applications in biophosphorylation
The study of phosphoproteome is a significant topic in the field of bioinformatics. Human genome has 518 genes of coding protein kinases [13], therefore predicting the kind of protein kinase that a particular amino acid residue gets phosphorylated is a typical multi-class classification problem. Xue et al. introduced Bayesian decision theory into kinase-specific phosphorylation site prediction studies [14]. The amino acid frequencies of each pair of positive and negative data sets were counted according to the conservativeness of the sequences. The amino acid frequencies of 9 amino acids were calculated by selecting 8 amino acids in the neighbor phosphorylation sites. They calculated the probability distribution for each position and each amino acid in each sample.

6. Conclusion
Bayesian Statistics is a hot topic at present. This essay introduces fundamental theory of Bayesian Statistics and Bayesian Theory’s development and applications, as well as the history of Bayes Statistics, and compares the differences between Classical Statistics and Bayesian Statistics. The essay also illustrates some concepts of Bayesian Statistics and their developments.

Bayesian Theory shows advantages under conditions involving small samples or high sample costs, such as on industrial assemble lines and disaster warning. Bayesian decision theory is used to simulate human behaviors in researches on psychophysics and cognitive sciences. Though Bayes Theory has been applied to various fields including industry, business and biology, it still has potential in other fields and worth being studied. For example, data of virtual property is hard to be drawn due to the interest of
related operator, and behavior of users is highly irregular and dependent. This could be a field where Bayesian Statistics is more convenient, and protecting consumers on virtual property will be more and more important as electronic industry grows.

Reference
[1] T Bayes. An essay towards solving a problem in the doctrine of chances [J]. Resonance, 2003, 8 (4): 80-88.
[2] T Bayes. An essay towards solving a problem in the doctrine of chances [J]. Biometrica, 1958, 46: 293-298 (reprinted from an original paper of 1763).
[3] RA Fisher, Theory of Statistical Estimation [M], Institute of Mathematics, University of Tsukuba, 2008, 1 (1): 51-65.
[4] JO Berger, Statistical Decision Theory and Bayesian Analysis [J], Springer, 2002, 83 (401): 266.
[5] RT Bayes. An essay toward solving a problem in the doctrine of chances [J]. Philosophical ransactions of the Royal Society of London, 1763, 53: 370-418.
[6] DR Cox, DV Hinkley. Theoretical Statistics [M]. Chapman & Hall: London, 1974.
[7] P Congdon. Bayesian statistical modeling [M]. John Wiley & Sons, Ltd, Chichester, 2006.
[8] MM Zhu, W Liu, YL Yang. Construction algorithm of MPD-JT for Bayesian networks based on full conditional independence [J]. Systems Engineering and Electronics, 2010, 32 (6): 8-11.
[9] AL Yuille, HH Bülthoff, Bayesian Decision Theory and Psychophysics [J]. Perception as Bayesian Inference, 1996, 11(4): 123-161.
[10] KP Körding, DM Wolpert, Bayesian decision theory in sensorimotor control [J]. Trends in Cognitive Sciences, 2006, 10 (7): 319.
[11] LW Tang, BQ Diao, XB Ma, MQ Fan, XF Lv. The recruitment decision in enterprises based on the Bayes decision of minimum error ratio [J]. Operations Research and Management Science, 2005, 2 (14): 1.
[12] P Li, GH Geng, MQ Zhou. Face detection method based on neural network and Bayesian decision [J]. Application Research of Computers, 2007, 24 (8): 198-200.
[13] G Manning, DB Whyte, R Martinez, et al. The protein kinase complement of the human genome [J]. Science, 2002, 298: 1912-1934.
[14] Y Xue, A Li, L Wang, et al. PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory [J]. BMC Bioinformatics, 2006, 7: 163.