Edward Moradian

Professor John Guenther

Stat 6550: Bayesian Statistics

25 May 2018

A Short Recent History of Bayesian Statistics

1. Abstract

Bayesian statistics started from its nascent beginning in 1763 with Reverend Thomas Bayes’ seminal work, *An Essay Towards Solving a Problem in the Doctrine of Chances*, being published posthumously after his death in 1761. For the next 250 years and on, Bayesian statistics developed into three main distinct approaches – subjective Bayesianism, objective Bayesianism and empirical Bayesianism. Most recently, Bayesian statistics has seen an upsurge in Bayesian activity with the advent of computational power from computer technology. New algorithms, such as the Metropolis-Hastings and Gibbs sampling algorithms, have made intractable integral calculations on posterior distributions possible. Consequently, the scholarly literature on Bayesian statistics has dramatically increased in publications. Outside academia, Bayesian statistics has also gained widespread acceptance and practice among many different industries.

1. Introduction

There are three main steps in Bayesian statistics – model specification, calculation of the posterior distribution and model assessment. In model specification, a statistical model is selected for the statistical analysis. A complete probability model for given data requires two probability distributions – a probability distribution for the data given on unknown parameters (i.e., the likelihood distribution) and a probability distribution for the unknown parameters (i.e., the prior distribution). The prior distribution describes what we know about theta, an unknown parameter, before having examined the data. The likelihood distribution is the contribution of the data to information about theta. The change from the prior distribution to the posterior distribution is a reflection of the information provided by the data about theta. Either an informative prior or a noninformative prior must be selected depending on whether to include subjective prior information or objective prior information respectively. In the calculation of the posterior distribution, the two probability distributions, the prior distribution and the likelihood distribution, are combined to obtain an updated probability distribution. The posterior distribution denotes the most up-to-date information about theta that also can be updated in the future. From the posterior distribution, there are many different inferential conclusions that can be made. A point estimate or an interval estimate can be calculated to give a one-number or an interval of numbers summary for a parameter. Hypothesis testing is also conducted on the posterior distribution. In model assessment, the validity of the assumptions made in the model are assessed. Model validity can be assessed using sensitivity analysis, model checking and model comparison (Stern 373-376).

Bayesian statistics has a few key differences to frequentist statistics. In Bayesian statistics, parameters are treated as random variables, whereas in frequentist statistics, parameters are treated as fixed constants. In Bayesian statistics, the posterior distribution, that is calculated from combining the prior distribution with the likelihood distribution, is used to estimate the mean value and credible interval of a parameter. A credible interval is used in Bayesian statistics to state that a parameter is within a certain range by a 95% probability usually. This allows for a natural probabilistic interpretation of the plausible values of a parameter. In frequentist statistics, a parameter is given a confidence interval that is calculated using the sample mean and sample standard deviation. Parameters are estimated using only the data and a prior information external to the data is never used like it is in Bayesian statistics. Bayesian statistics allows for parameter distributions to be updated based on new data by having each posterior distribution be the subsequent prior distribution in a Bayesian calculation.

1. Recent History of Bayesian Statistics

Present-day Bayesian statistics was first developed in the time period around World War II. During this time, Alan Turing used Bayesian logic to solve the enigma code while others used Bayesian logic to locate the Axis U-boats to help the Allies protect commercial and military ships. During this period, Turing had also invented a Bayesian approach to sequential data analysis wherein he wanted use information arriving piecemeal to estimate the probability of a hypothesis with a prior probability (Fienberg 12). Following World War II, Bayesian statistics continued to be used the private sector. From 1950 to 1980, Bayesian statistics stagnated in academia primarily because of two main reasons. First, there was an argument among Bayesian statisticians of about how to approach prior distributions. Should subjective Bayesian statistics be used with an informative prior or should objective Bayesian statistics be used with a noninformative prior? Second, Bayesian statistics was restricted because of the limitations to computational technology that were not adequate to solve or approximate difficult integrals required in large-scale realistic Bayesian statistics (Lynch 380). Bayesian statistics experienced a dramatic upsurge in the late twentieth century with advances to computation. In particular, technological advances in computers and the use of computational general-purpose simulation algorithms made it possible to apply Bayesian statistics in contexts that were previously hard or impossible (Stern 373). For example, the Metropolis-Hastings algorithm and Gibbs sampling solved the problem in Bayesian statistics of approximating high-dimensional integrals. Instead of completely evaluating a posterior distribution, a sufficient number of draws from an approximate random sample from the posterior distribution can be made to empirically approximate a posterior distribution (Bolstad 23). Many disciplines benefited from Bayesian statistics where the frequentist approach was not amenable to. Computer science has used Bayesian statistics in machine learning and artificial intelligence. The social sciences have made use of Bayesian statistics in using a hierarchical structure. Hierarchical models are used where parameters are functions of other parameters. For example, spatial hierarchies can be created such as students nested within schools, communities and states. Bayesian statistics has experienced widespread adoption with a continuing increase in its use (Lynch 380-381). Ever since the mid-1990’s, developments in Bayesian statistics concentrated on the Markov Chain Monte Carlo, model evaluation and model comparison (Lynch 381).

1. Three Modern Approaches to Bayesian Analysis

In the recent history of Bayesian statistics, Bayesian statistics has developed to the point where there are many well-founded and continually developing different approaches to Bayesian statistics. Three main approaches to Bayesian statistics are objective Bayesian analysis, subjective Bayesian analysis and empirical Bayes.

Objective Bayesian analysis is the objective approach to Bayesian statistics. A key way in which objectivism is strived for is by using noninformative or improper priors. Thomas Bayes and Pierre-Simon Laplace were objective Bayesians in the sense that they used a constant prior distribution for unknown priors. Some well-known noninformative priors are the Jeffrey’s prior, maximum entropy prior and reference prior (Berger 1271). When using a noninformative prior, the likelihood distribution is not changed and therefore the posterior distribution is based solely on the likelihood distribution. In other words, the only information used in the objective Bayesian analysis is internal to the data (i.e., the likelihood distribution). A primarily exercise in objective Bayesian analysis is to be parsimonious with the prior. That is, the prior should not contain more information than what is necessary (Efron, Why Isn't Everyone a Bayesian? 4).

In direct contrast to objective Bayesian analysis is subjective Bayesian analysis. Subjective Bayesian analysis uses subjective prior information and therefore utilizes informative priors (Berger 1270-1271). The prior distribution then changes the likelihood and in turn the posterior distribution is affected. In other words, the information used in the subjective Bayesian analysis is a combination of information both internal (i.e., the likelihood distribution) and external to the data (i.e., the prior distribution). Examples of information external to the data may be from the researcher’s intuition or the views of substantive experts. Two prominent statisticians who developed subjective Bayesian analysis are Bruni di Finetti and Leonard Jimmie Savage. Bruno di Finetti stated that the only objective view of probability is the subjective one, because it can be tested by the rules it must obey (Finetti).

Empirical Bayes tries to combine frequentist statistics and Bayesian statistics in a compromise between the two different approaches to statistics. It does this by the way of having a prior distribution that is estimated frequentistically that is data-dependent (Efron, Bayesians, Frequentists, and Scientists 3). In other words, the prior distribution is estimated by the data and therefore there are empirical priors in empirical Bayes. Estimations are usually done by the maximum likelihood. Once the prior distribution has been estimated the rest of the statistical calculation goes through the standard Bayesian paradigm with a posterior distribution calculated (Maritz and Lwin 13). In turn these prior distributions are objective which is a characteristic shared with objective Bayesian analysis. They are objective, however, because of the use of frequentist estimation methods which are objective. Empirical Bayes was first introduced by Richard von Mises in 1943 followed by Herbert Robbins in the 1950’s.

With three main approaches to Bayesian statistics, how does one choose which one to use? One view of the best practices for statisticians is to use the statistical analysis that is best for the given statistical task. This guideline includes whether to choose frequentist statistics or Bayesian statistics. According to the prominent statistician Carl N. Morris, collectively the three different Bayesian approaches should be viewed as three different tools to accomplish different statistical tasks. Among these three Bayesians approaches and frequentist statistics, Morris states that, “There can be no clear victory for any approach for all applications; rather, we should train statisticians for a frequency-Bayes compromise” (Morris 7-8). This view is also shared by another prominent statistician, Bradley Efron, who states that Bayesian statistics and frequentist statistics are, “more orthogonal than antithetical … practicing statisticians are free to use whichever methods seem better for the problem at hand – which is just what I do” (Efron, A 250-Year Argument: Belief, Behavior, and the Bootstrap). Objective Bayesian analysis can be used when objectivity is appropriate such as in scientific statistical analysis whereas subjective Bayesian analysis can be used when subjectivity is appropriate such as in business, law or political science statistical analysis. When a statistician wants to incorporate frequentist results into the prior distribution, empirical Bayesian analysis can be used.

1. Conclusion

Ever since 1990, there has been a sharp increase in Bayesian statistics. To illustrate this point further, here is a look at how academic literature on Bayesian statistics has sharply increased in recent decades.

In the first 200 years of Bayesian statistics, from 1769 to 1969, there were around 15 books written on Bayesian statistics. From 1970 to 1989, there were around 30 Bayesian statistics books produced. In 1990 through 1999, there were around 60 Bayesian statistics books produced. From 1986 to 1999, Bayesian methods in different disciplines such as archaeology, atmospheric sciences, economics, econometrics, education, epidemiology, engineering, genetics, hydrology, medicine, physical sciences and social sciences have been written (Berger 1270).

The academic literature of Bayesian statistics has also shown this increasing interest in doing Bayesian statistics. From 1971 to 1991, according to a Web of Science search, there were 2,160 papers involving Bayesian as a key word. From 1991 to 2011, in a Web of Science search, there were 40,242 papers involving Bayesian as a key word (Lynch 381).

Overall, activity in Bayesian statistics appears to be rapidly increasing.

Works Cited

Berger, James O. "Bayesian Analysis: A Look Today and Thoughts of Tomorrow." *Journal of the American Statistical Association* (2000): 1269-1276.

Bolstad, William M. *Introduction to Bayesian Statistics*. Hoboken: John Wiley & Sons, Inc., 2007.

Efron, Bradley. "A 250-Year Argument: Belief, Behavior, and the Bootstrap." *Bulletin of the American Mathematical Society* (2013): 129-146.

—. "Bayesians, Frequentists, and Scientists." *Journal of the American Statistical Association* (2005): 1-5.

—. "Why Isn't Everyone a Bayesian?" *The American Statistician* (1986): 1-5.

Fienberg, Stepen E. "When Did Bayesian Inference Become "Bayesian"?" *International Society for Bayesian Analysis* (2006): 1-40.

Finetti, Bruno di. "Theory of Probability." *London* (1974).

Lynch, Scott M. "Bayesian Theory, History, Applications, and Contemporary Directions." *International Encyclopedia of the Social and behavioral Sciences* (2015): 1056-1060.

Maritz, J.S. and T. Lwin. *Empirical Bayes Methods with Applications*. London: Chapman and Hall, 1989.

Morris, C. N. "Comment." *The American Statistician* (1986): 7-9.

Stern, Hal S. "Bayesian Statistics." *International Encyclopedia of the Social and Behavioral Sciences* (2015): 373-377.