Bayesian Methods with Application in Risk Analysis

Dongyang DING

Center for Anti-Corruption Studies, Nanchang
University, Nanchang, China
Email: derakding@hotmail.com

Xiyang LIU

Yingtan Central Sub-Branch, The People's Bank
of China, Yingtan, China

Abstract — In risk analysis, Bayesian methods are more adaptability and flexibility than traditional methods when be used to construct decision framework, estimate risk distribution and parameterize model, but has shortcomings at the same time. Robust methods make up some limitations of Bayesian methods, the analysis of uncertainty indicate that robust Bayesian methods can produce more reliable inference in the absence of comprehensive statistical information.

Keywords-risk; uncertainty; Bayesian; robust

I. Introduction

Risk analysis refers to a set of methods address various issues that arise from the uncertainty, including risk identification, risk estimation, risk control and management. The concept of risk comes from economics, and the concept of uncertainty comes from the statistical decision theory. The economist Knight defined risk is reliability of the probability estimation and the possibility of the treatment as the cost of an insurance, Knight uncertainty means difficult to predict and quantitative analysis the results of events by existing theory or experience(Knight, 1921)[1]. The concept of "uncertainty" in Risk analysis includes not only the Knight uncertainty and the risk. Computational problems in risk analysis have three distinctive features: some variables lack of empirical data; the adoption of the analyst's judgment or expert opinion is necessary; the mathematical model exist uncertainty.

The Bayesian approach is the theory used to examine the uncertainty of probability model parameters, was recognized as the proper way to make use of expert opinion and subjective information, and thus has a distinct advantage in addressing the problem of risk analysis. The Bayesian approach has unified theory about inference and decision, but in fact there are significant differences within its methods. For example, the vast majority of Bayesian advocates believe subjective probability and make it as the basic principles in the framework of inference, but there are also some "objective Bayesians" opposed to the subjective probability theory; majority of Bayesians must simplify the parameter distribution to a single value, but the others use the entire distribution to describe the uncertainty of parameters; most of Bayesians believe that can have a precise probability model, but the other analysts who advocating robust Bayesian methods suggest to use a series of probability models.

Standard Bayesian approach has unique position in the analysis of risk and probabilistic uncertainty, but is not robust under the conditions of realities contrary to the own assumptions. The standard approach requires that the analyst can provide more information than one can obtain in the normal risk analysis which lack of data, at the same time its standard assumptions are not accurate due to underestimate the tail risk. Robust Bayesian method is preferred to draw firm conclusions in the absence of comprehensive statistical information, and it is more applicable to the issues considered in risk analysis. This paper primarily describes the standard Bayesian method and robust Bayesian approach can be applied to risk analysis.

II. Bayesian methods in risk analysis

Application of Bayesian methods in risk analysis has three ways: the first is to build decision framework, take full control of the analysis and decision process, provide a comprehensive program for inference and decision; the second is to estimate the distribution of risk which is a core part of the risk analysis, the quantitative characteristics and distribution drawn from the analysis are the most important information to decision; the third is to parameterize the model, estimate the distribution of the exogenous variables of the model rather than directly conclude, which is technical support in the process of decision(Scott, 2004)[2].

A Advantages

Within the range of applied Statistics in risk analysis, general concern is the advantages of Bayesian statistics to the traditional frequency statistics. The frequency statistics represented by Fisher, Neyman, Pearson and Gosset, still occupy a mainstream position, but Bayesian grew up at an alarming rate in the past quarter-century. The following highlights several aspects of its advantages in order to stimulate interest of risk analysts in Bayesian methods.

1) Effectively address uncertainty.

Probability in Bayesian theory is a subjective quantity, rather than the traditional measure of the limited frequency, so the Bayesian method can calculate a credible interval to describe the uncertainty of parameter estimation. Compared with the traditional Neyman-Pearson confidence interval, credible interval is more realistic and easy to handle for feasible calculus in the latter part of calculation. Parameter in the model can be expressed by using the probability distribution in Bayesian theory, in contrast, the frequency method does not allow distribution as a model constant, and thus often can not use distribution to express the model parameters, only choose the probability of random data. Bayesian methods continue updating during the data accumulated, do not need to pre-set sample size, can stop collecting data at any time, calculation and interpretation of the results does not depend on the experimental design of the commencement program. This means that the Bayesian

method has greater adaptability and flexibility, and can give a more accurate inference in extreme events may be encountered in risk analysis.

2) Comfortably data mining.

Another important advantage of the Bayesian approach is to allow a general understanding of the data in advance. Traditional Neyman-Pearson school insists that data mining is not scientific, and the statistical analysis should be gradual implemented, preparation of the programs first, and then select the data set, and calculate P-value of a single assumption, the final conclusion. And the number of parameters should not be too much when only have a very small sample, also can not estimate too many parameters in one time. Analysts do not and should not be the bound of these blames in practice since data are often in short supply, but also the analysts inevitably see the overview of data in the collection process. Only continue to receive new data, it is possible to clearly understand the problem and correctly infer the model. It is quite normal to have prior knowledge of the data in the Bayesian approach, and can estimate the large number of parameters do not have to worry about a small sample.

3) Rationally making decision.

The Bayesian method is very suitable for decision analysis, can ensure decision-making in line with consistency by manifesting all the uncertainty with probability, and thus make the analysis and decision-makers to build a consistent set of policies on risk assessment and management. Traditionally, the only conclusion that can be given by the frequency method is whether to reject the null hypothesis depending on the data set, the method only controls the first type error which rejects the proper null hypothesis and did not balance the cost of two type error. It is necessary to balance two types of costs to give a reasoned decision in actual decision-making. Bayesian approach can fully express the meaning of risk management work as the basic principles of decision analysis. Although the risk analysts do not require mastering the Bayesian theory, basic knowledge structure of its analytical work is indeed the scope of the Bayesian method.

4) Orderly using subjective information.

Bayesian approach formalize the way to direct use subjective information to a certain extent, make it possible to use subjective information in the objective equation, the information may be the analyst's personal judgment or the views of experts. The experience of analysts or expert advice is very useful in risk analysis; it is very profligate to ignore these messages for the risk analysis which is lack of information. Frequentist also use subjective information such as to modify the model structure or change the parameters, however their use is ambiguous and can not illustrate the effect of this information on the calculated result. Therefore, the Bayesian direct using subjective information is better than those fuzzy of frequentist.

B   Limitations

There are also some limitations of Bayesian methods in risk analysis, mainly due to the lack of empirical data makes the set of prior distribution give more effect on inference

accuracy than normal, and must face calculation difficult in the quantitative analysis, which are the challenges when using Bayesian methods in risk analysis.

1) Prior distribution.

Most important challenges Bayesian faced is how to choose the prior distribution, the Bayes rule can not draw any conclusions especially for the event does not believe in advance. The posterior distribution will change as changing the prior probability; it can be any possible probability through changes in prior probability. Although the differences come from prior will soon be overcome and eliminate with the accumulation of data, but it is often lack of data in the risk analysis.

2) Zero preservation.

Zero preservation means no matter how much the likelihood function value, and regardless of any new data, a prior probability is set to zero will lead to a posterior probability is zero. Similarly, regardless of prior probability, the likelihood function is zero, the posterior probability will be zero. This means that if analysts deny the events may occur, then any contrary information can not affect the conclusion.

In addition, the Bayesian model ignores the difference between uncertainty and probability. Furthermore, Bayes factor is used as the weight of model to constitute mixed model and distribution in Bayesian averaging method to describe the model uncertainty, whose drawback is that the analysts are able to list all the possible models, which in reality is very difficult to overcome. Even if all possible models can be listed, another problem is that require analyst to specify the prior probability of each model, which may result in underestimated risk.

III. Robust Bayesian methods and risk analysis

Robust Bayesian methods, also known as Bayesian sensitivity analysis, is used to study the robustness of the conclusions with uncertainty (Geweke, 2005)[3]. If the conclusion is not sensitive depends on the assumptions and the exogenous variables, which indicates conclusion is robust. Robust Bayesian methods believe that it is difficult to specify a precise prior distribution in many cases, and the likelihood function for a particular issue is also doubtful. In robust Bayesian method, all possible combination of the prior distribution and the likelihood function are through standard Bayesian methods, which means a series of posterior distribution will be received by all possible combination. Robust Bayesian method also uses a similar strategy, pair a class of probability model and a class of utility functions together to infer the appropriate decision, and uncertainty will be described for the best probability model and utility function. In both cases, if the conclusions are approximately the same given in each combination, which indicates the robustness; if the conclusion is significantly different, then the credibility will be given from the analysis. In standard Bayesian method, the uncertainty is measured by a single additional probability, personal attitudes and values are represented by the precise utility function, which are all for convenience, mainly due to cost and time constraints not allowed to spend too much energy to explore the precise function, the robust Bayesian

approach apparently is not consistent with these Bayesian thinking. Walley (1991) considered robust methods distinguished the class of variability from the uncertainty, is an extension of the traditional Bayesian approach.

Robust Bayesian approach involves some important pioneering concept from other areas, whose ideas often can be applied to Bayesian analysis without modification. Robust Bayesian methods can be constructed in several ways, including the use of: *a) conjugate parameter family; b) non-conjugate parameter family; c) the relative density (the bounded density distribution); d) ε -pollution, mixture, quartile et al.; e) bound of the cumulative distribution.* Although computational problems in Robust Bayesian are profound, sometime is very straightforward in special circumstances.

A   Interval estimation

First define the following interval formula:

$$x + y = [x_1, x_2] + [y_1, y_2] = [x_1 + y_1, x_2 + y_2]$$
$$x - y = [x_1, x_2] - [y_1, y_2] = [x_1 - y_2, x_2 - y_1]$$
$$x \times y = [x_1, x_2] \times [y_1, y_2] = [x_1 \times y_1, x_2 \times y_2]$$
$$x \div y = [x_1, x_2] + [y_1, y_2] = [x_1 \div y_2, x_2 \div y_1]$$

(1)

The x and y in (1) represent the probability of event, if x or y is the accurate value s ，then the interval can be expressed as $[s, s]$ ，which can also be used in (1). Gelman (2003)[4]pointed out that only use once the interval expression for each uncertain variable to compute would get a better result.   However, there is duplication in the interval expression,for example：

$$\frac{P(\text{x} \mid y) \times P(y)}{P(\text{x} \mid y) \times P(y) + (1 - P(y)) \times p(x \mid \overline{y})}$$

(2)

Both $p(y)$ and $p(x \mid y)$ in (2) appeared many times, the normal interval estimation can not make a reliable analysis of (2). If the variable has uncertainty, then use no repeat parameter expression is a better choice:

$$1 / (1 + ((1 / p(y) \text{-}1) \times p(x \mid \overline{y})) / p(x \mid y))$$

(3)

The formula (3) is equivalent to the original cases (2), but when the input value is the interval form, will get more accurate results. The probability inferred from the expression without repeated parameters will be the smallest possible interval.

B   P-boxes

The P-boxes refers to the upper and lower limits of the cumulative distribution function, by which can make qualitative analysis on the uncertainty of the prior distribution. If the uncertainty of the likelihood function can be made qualitative analysis through the upper and lower limit of the integral function, then the likelihood function also has a P-boxes as long as whose limits can be standardized. In robust Bayesian analysis, the limits of cumulative distribution function as posterior distribution most likely to be a rectangle P-boxes when the P-boxes of prior distribution and likelihood

function are combined, that is their intersection.

Degradation of posterior class is well understood, such as in zero preservation as mentioned earlier, whether prior or likelihood function is 0, the posterior must be 0. The P-boxes of prior distribution allow the prior cumulative distribution is relatively flat in the range of values, when the prior and likelihood function are qualitative description of the border, can not do more inferred except for knowing the boundary of posterior distribution. The shape of the two P-boxes has no effect on the inference, only the intersection of them is useful.

C   Parameter class

Another way to define distribution family in Robust Bayesian methods is to specify their parameter class. Suppose there are two types of distribution, one type has the same variance and different mean, and the other has the same mean and different variance. The difference of parameters indicates the different distribution, such as normal distribution family, it could be other shapes drawn by the conjugate, or may not come from in the conjugate family. The parameter class means that analyst is uncertain of the prior, different parameters also shows the uncertainty. In a sense, robust Bayesian parametric methods just have the opposite problem with the P-boxes, in which the class is too large to make important inference about the posterior, and in parameter method the class is too small to represent the uncertainty of the prior and likelihood function.

D   Bounded density class

The prior and the likelihood function can also be bound by using the probability density instead of the cumulative probability distribution. The limitation of posterior density function can easily draw from the product of their two boundaries. For any value of θ ，the prior and the likelihood function can always be bound by different interval, and can make product of two interval using interval estimation algorithm. The product makes up the upper and lower of θ in non-standardized posterior distribution. If the distribution class from P-boxes method is too loose and too large, the distribution from parameter class method limit too much and is too small, bounded density method may be suitable for practical robust Bayesian analysis.

Should be noted that setting an a prior and likelihood function, would not be bound by data lacking in Robust Bayesian methods due to the use of the distribution class, but Robust Bayesian method still have some problem which is the same as in traditional Bayesian approach, such as zero preservation.

IV.   Conclusion

Probability theory is very successful in the field of analysis due to create the way to address the divergence in the real world with the "probability" idea. Based on conflict of pluralism and universality, some Bayesians who brought all probability for their own use, show a lot of the advantages compared to traditional methods in risk analysis. Jaynes (2003)[5] strongly advocate the probability theory is not the mathematical science of frequency, but calculation methods on creditability about certain events relative to the other

events, and creditability should not be on the primary position in risk analysis. In order to carry out risk analysis, a mathematical science of frequency and a calculation method that allows calculating the distribution is necessary, so that the weights of all possible data collected can be visual interpreted. At the same time, this calculation method can take advantage of the subjective information, and should not garble the concept of variability and uncertainty to some extent. Still has the fear of calculate difficulties, robust Bayesian approach amends many troubling problems in the traditional Bayesian approach, which also reflects the gradual improvement of probability and mathematical sciences in risk analysis.

As the statistical profession matures in recent years, the discussion of Bayesian and Frequentist become more inclusive, both sides have a clear understanding of their own advantages and disadvantages. However, areas outside the statistical profession, debate in risk analysis and other mathematical or quantitative analysis, is increasingly fierce, which shows risk analysis is a young thriving areas. The focal point of the divergence is difficult to choose some methods, different ideas bring different concepts, which may never be rationalized since this is the performance of not fully understood the complexity of human decision-making in the world. But with the confusion be cleared and misunderstanding be corrected, different concepts will complement each other, in order to expand and deepen the understanding of risk, and will continue to improve the field of risk analysis.

References

[1]   Knight, F.H. Risk Uncertainty and Profit[M]. Boston: Houghton Mifflin, 1921. 2-12.
[2]   Scott F. Bayesian Methods in Risk Assessment[R]. Working Paper. Head of Waste & Storage Unit, 2004. 5-20.
[3]   Geweke, J. Contemporary Bayesian Econometrics and Statistics[M]. New York: Wiley, 2005. 120-130.
[4]   Gelman., A. et.al. Bayesian Data Analysis, 2nd ed.[M]. London: Chapman and Hall, 2003. 220-228.
[5]   Jaynes, E.T. Probability Theory[M]. Cambridge: Cambridge University Press, 2003. 95-108.