IPS148: Statistical Literacy Across Disciplines

Uncertainty evaluation with the concepts of Bayesian Statistics

Richa Vatsa

vatsa.richa@gmail.com

Department of Statistics, Central University of South Bihar, Gaya, India

Abstract

Our daily life is regulated by two types of events: deterministic and uncertain. Uncertain situations are random and may result in different possible consequences when replicated. On the other hand, an uncertain scenario leaves us with different alternatives to choose from. Therefore, it requires proper planning and decision-making to choose the best alternative to live life securely and comfortably. Quantification of the uncertainty can make it easier to make the optimum decision. Probability theory has been the oldest concept to quantify uncertainty in real life; there are two approaches to quantify the uncertainty of an event- frequentist and Bayesian. The Bayesian approach is more suitable for real-life scenarios. For example, should I carry an umbrella during a trip if it rains in the coming days? What is the chance that a particular volcano shall erupt soon? Such real-life events cannot be replicated as laboratory experiments.

Moreover, the non-Bayesian approach may not explain them. Nevertheless, the uncertainty around these events may be quantified based on subjective or experts' opinions. We often update our beliefs or judgment in light of the available facts and act accordingly in our daily life. For example, suppose that a person is down with a fever. A doctor speculates it as viral fever. This expert's prior judgment is revised based on the appropriate medical diagnoses, and the medication is prescribed. Thus, prior knowledge is required in some life scenarios to make better decisions.

Theoretically, the uncertainty of an event is formulated in unknown model parameters to be estimated. In Bayesian statistics, the model parameters are assigned with probabilistic statements conditioned on the available observations. There are two types of uncertainty quantification in model parameters prior and posterior to observing data: prior density and posterior density. As a combination of prior and present knowledge formulation of a parameter, posterior density leads to further summarized uncertainty measurements in parameters essayed by Bayesian point estimates and credible or highest posterior density regions.

Uncertainty in a future event is predicted through predictive posterior densities given presently available observations and summing over the posterior uncertainty in parameters and summarized with credible regions or point estimates.

This article shall explain the key concepts and applications of Bayesian statistics to deal with uncertainty with a suitable example and R-codes.

Keywords: Bayesian estimation, Prediction, prior density, posterior density, Credible interval

Introduction

Our daily life is regulated by two types of events: deterministic and uncertain. Uncertain situations are random and may result in different possible consequences when replicated. On the other hand, an uncertain scenario leaves us with different alternatives to choose from. Therefore, it requires proper planning and decision-making to choose the best alternative to live life securely and comfortably.

Quantification of the uncertainty can make it easier to make the optimum decision. Probability theory has been the oldest concept to quantify uncertainty in real life. Let us consider a few examples to understand probability. If a student undergoes a test for Statistics subject, what is the chance of her passing the test? A lot consists of 10 useful articles, 4 with minor defects, and 2 with major defects. One article is chosen at random. What is the chance that it has no defects? Answers to all these problems of uncertainty can be found with the concept of probability.

Furthermore, there are two approaches to quantify the uncertainty of an event-frequentist and Bayesian. The Bayesian approach is more suitable for real-life scenarios. For example, should I carry an umbrella during a trip if it rains in the coming days? What is the chance that a particular volcano shall erupt soon? Such real-life events cannot be replicated as laboratory experiments. Moreover, the non-Bayesian approach may not explain them. Nevertheless, the uncertainty around these events may be quantified based on subjective or experts' opinions.

We often update our beliefs or judgment in light of facts available and act accordingly in our daily. For example, suppose that a person is down with fever and headache. A medical practitioner speculates it as viral fever. This expert's prior judgment is revised based on the appropriate medical diagnoses, and the medication is prescribed. Thus, prior knowledge is required in some life scenarios to make better decisions. On the other hand, suppose some prior knowledge, personal judgments, or experts' opinions are available about the events that could affect our decision in life. In that case, the inferences concerning uncertainty may be attributed to the theory of Bayesian statistics.

Theoretically, the uncertainty of an event is formulated in unknown model parameters to be estimated. In Bayesian statistics, the model parameters are considered random and quantified with probabilistic statements conditioned on the available observations. Thus, there are two types of uncertainty quantification in model parameters prior and posterior to observing data: prior density and posterior density. As a combination of prior and present knowledge formulation of a parameter, posterior density leads to further summarized uncertainty measurements in parameters essayed by Bayesian point estimates and credible or highest posterior density regions. Uncertainty in a future event is predicted through predictive posterior densities given presently available observations and summing over the posterior uncertainty in parameters and summarized with credible regions or point estimates.

In Bayesian statistics, the uncertainty statements are straightforward and may consider both subjective and objective probability approaches. The prior knowledge based on belief or judgment is called subjective. In contrast, the objective notion provides probability measurements based on the model formulation of available facts.

Bayesian statistics has found increasing popularity in applied sciences and social science to deal with uncertain real-life scenarios. Therefore, this article aims to present the fundamental concepts of Bayesian estimation of uncertainty with a suitable example.

Methodology

Let us consider a simple example of electoral polls. Suppose candidate A has got 54 votes, and B received 46 votes in a 100 people poll. Then, what are the chances of A winning the election? Further, what is the probability that a new voter in poll shall vote for candidate B?

Let x denote the number of votes received by A. Then the uncertainty in behaviour of voters can be modelled as a Bernoulli variate y with θ probability of wining of A; y=1: vote to A, y=0: vote to B. Then x as the sum of values of y shall follow a binomial density with parameters as total number of voters n and probability of wining of A, θ.

In general, let a r.v. X follows the distribution $f(x|\theta), \theta \in \Theta$. Then, considering θ as a random quantity, the conditional probability density of θ given x may be computed by the Bayes' theorem as, $P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$. The term $P(\theta)$ stands for the prior uncertainty measurement (density) of θ over Θ reflecting the knowledge around θ gathered before observing data x. Further, the term $P(\theta|x)$ refers to the posterior uncertainty measurement (density) of θ given data x comprising of the updated knowledge of θ in light of data. Nevertheless, $P(x)$ is called marginal likelihood as a function of data x only. It can be computed by integrating/summing the joint density, $P(\theta, x) = P(x|\theta)P(\theta)$, over the uncertainty of θ; however, its computation may be analytically amenable, which may lead to further approximations to the posterior densities.

Why is θ random?

As a critical concept in Bayesian statistics, the parameter $\underline{\theta}$ is considered random. The subjective notion of probability underlies the Bayesian framework. Under alternative hypotheses, an unknown (parameter) may have distinct values, accounting for its uncertainty. This uncertainty can be measured using expert assessment, personal belief, or previous research on the unknown. The question of assigning a prior density to an unknown parameter has long been a source of contention among Bayesian statisticians. As a result, prior elicitation emerges as a branch of Bayesian statistics research.

Choosing a prior:

The initial step in calculating posterior uncertainty is to choose an appropriate prior by assessing knowledge, no-knowledge, or little knowledge about θ before obtaining data. A prior density plays a vital role in posterior inference about θ in the situation of weak or insufficient evidence. Prior density, on the other hand, only has a significant effect on a posteriori inference given sufficient and strong datasets. As a result, choosing a prior density for a meaningful inference about θ should be done with caution.

Prior elicitation is the process of quantifying a prior density around θ depending on the experts' knowledge or judgement. The *subjective approach* of selecting a prior allows experts' opinions to be translated into probabilistic statements. According to Berger [1985], the subjective approach can be divided into different categories based on the subjective knowledge accessible regarding $\theta \in \Theta$- relative likelihood approach, histogram approach, matching of a given functional form, and CDF determination. To fit no or little prior information about θ *objective priors* are also known as vague, non-informative, and weak priors. Laplace's prior, Jeffreys priors, reference prior, and locally uniform prior are the most commonly utilised priors in this category. *Conjugate priors* can be classified either informative or non-informative. The conjugate priors are the most useful as they enable posterior densities easy to compute, at least for lower dimensional situations. Berger [1985], Lee [2012], Gelman et al.[2006], and Gosh et al.[2006] may be explored for more information on choosing a prior.

Bayesian interval estimation:

Rather than simply placing their uncertainty measures, it is often more convenient to deal with the most likely values of the unknowns (probabilities). Given a sample of data on the IQ of students of various ages, there is a 95% chance that the average IQ will fall between the range of 100-120. A summary estimate of the uncertainty in an unknown is provided by an interval estimate. The interval estimate of θ is relatively simple in the Bayesian technique. Credible intervals mainly need to find subsets with a certain posterior probability. A $100(1-\alpha)\%$ credible interval $(a, b) \in \Theta$ is defined such that, $\int_a^b P(\theta|x)\, d\theta \geq (1-\alpha),\ if\ \theta$ is continuous. For discrete θ, integration sign is replaced by summation. There can be numerous choices of $100\,(1-\alpha)\%$ credible intervals $(a, b) \in \Theta$ following its properties. The idea is to select the shortest interval with the most likely values of θ as its members. The shortest possible credible intervals (or regions) comprising the most probable values of θ are termed as highest posterior density intervals, or HPD intervals (or regions). HPD intervals are equivalent to equal tail credible intervals when $P(\theta|x)$ is symmetric and unimodal.

Predictive posterior density:

Quantifying uncertainty in future events is often desirable for decision making for government administration, public sectors, private sectors, industrialists, and even for individuals for maximum gain or minimum loss. For example, prediction for the unemployment rate next year would help the government open a number of new jobs. Prediction of a future observation is also subjected to uncertainty measurement in Bayesian inference. The posterior uncertainty measurement in a future or a new observation X_{new} given the current observations x is termed as a predictive posterior density. It is found as, $P(X_{new}|x) = \int_\Theta P(X_{new}|\theta, x)P(\theta|x) = \int_\Theta P(X_{new}|\theta)P(\theta|x)\, d\theta$. As per the assumption of independently and identically distributed observations given θ we may set $P(X_{new}|\theta, x) = P(X_{new}|\theta)$ to compute $P(X_{new}|x)$; however, the posterior density $P(\theta|\underline{x})$ must be found in advance.

Similarly to unknown parameters, uncertainty in future observations can be summarised with credible intervals or HPD intervals.

Application and Results

Let us consider the abovementioned example of electoral polls to show the effect of prior choices and different sample sizes on posterior probabilities.

There can be three steps to find the posterior probabilities-1. Set a prior density. 2. Define the likelihood. 3. Compute the posterior density. **Step 1. Prior density:** (a) For a no prior information on winning chances of candidates, θ, let us assign a uniform prior, i.e.

$\theta \sim U(0,1)$. (b) Assign prior density as $Beta(\theta; a = 4, b = 2.5)$ for candidate A having more chances of winning, i.e., for θ > 0.5 a priori. (c) Consider $Beta(\theta; a = 2.5, b = 4)$ as the prior density for B as more preferential candidate, i.e., θ < 0.5. **Step 2: Likelihood:** Let X denote the number of votes to candidate A, then $P(X = 54|\theta) = \quad {}^{100}C_{54}\, \theta^{54}(1-\theta)^{46}, 0 < \theta < 1,$ is the likelihood of θ. The term $(1-\theta)$ is the probability of winning of candidate B. **Step 3: Posterior density**: By the Bayes' law, the posterior density of θ is computed as, $P(\theta|X = 54) = \frac{L(\theta|X=54)P(\theta)}{\int_\theta L(\theta|X=54)P(\theta)\, d\theta}$. Analytical solution to the computation of the posterior density if found under different prior densities, as

(a) $P(\theta|X = 54) = \frac{\theta^{54}(1-\theta)^{46}}{\int_\theta \theta^{54}(1-\theta)^{46}\, d\theta} = Beta(\theta; 55, 47),$

(b) $P(\theta|X = 54) = \frac{\theta^{54}(1-\theta)^{46}\theta^{4}(1-\theta)^{2.5}}{\int_\theta \theta^{54}(1-\theta)^{46}\, \theta^{4}(1-\theta)^{2.5} d\theta} = Beta(\theta; 59, 49.5),$

(c) $P(\theta|X = 54) = \frac{\theta^{54}(1-\theta)^{46}\theta^{2.5}(1-\theta)^{4}}{\int_\theta \theta^{54}(1-\theta)^{46}\, \theta^{2.5}(1-\theta)^{4} d\theta} = Beta(\theta; 57.5, 51).$

An equivalent R-code to compute the above posterior densities with numerical integration method (discretization of integration), without solving it manually is as follows.

```
n= 100; r = 54; step = 0.01; theta = seq (0,1, by=step)
# prior density under (a), (b) and (c) cases
ptheta_a = dunif (theta, 0, 1); ptheta_b = dbeta(theta, 4, 2.5); ptheta_c= dbeta(theta, 2.5,4)
### likelihood
likeli = thetaˆr * (1−theta )ˆ(n−r)
# posterior density
joint_a= likeli *ptheta_a; ml = sum(joint_a *step); post_a = joint1/ml
```

Shown in the left side of Figure 1, we can compare these results using the posterior probability plots as obtained with the R-codes and. We can see from the figure that the posterior densities are almost overlapped by each other. That is, the effect of different prior densities on the resultant posterior is near negligible. It happens in the case of a strong likelihood, with a large data set. The R-syntax to produce the graph is as follows.

```
matplot (theta, cbind( post_a, post_b, post_c ), type= " l " , ylab = “Posterior density ” )
legend (“topright” , legend = c (“posterior 1” , “posterior 2” , “posterior 3” ), col = 1:3, lty = 1:3)
```

Let us consider a total of 10 total voters only, 7 to A and 3 to B. Right hand side of Figure 1 reflects that the choices of different priors have a strong impact on the resulting posterior densities given small sample sizes; i.e. the nature of the posterior density is sensitive to prior density choices. Therefore, for small sample sizes or weak data, it is advisable to use informative prior density as a strong prior information where possible.

Now, candidate A shall win if $\theta > 0.5$. Therefore, posterior probability of him wining is equal to P[$\theta > 0.5|r = 54, n = 100$] = 1- P[$\theta \leq 0.5|r = 54, n = 100$]. With r-code, 1-pbeta(0.5, 57.5,51) = 0.7345439

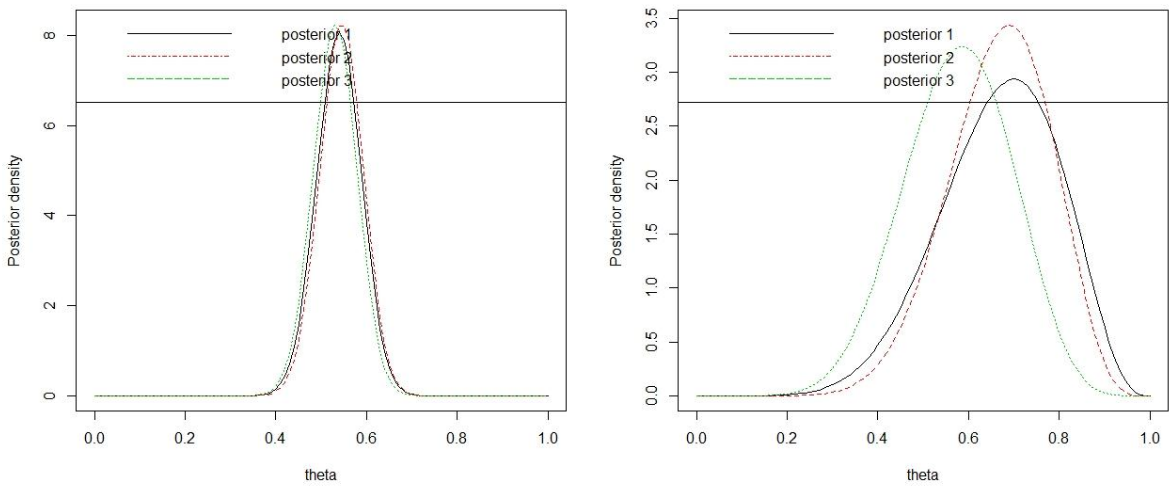

Figure 1: Comparison of posterior densities for different choices of priors with total number of voters and number of votes to candidate A, left–100 and 54, right–10 and 7, respectively.

For further results on application of Bayesian estimation, let us consider posterior density with prior density with option (c) and data as n=100, r=54. The posterior mode of the density is found 0.53 with the following R-code.

```
### finding posterior mode
findmode <- function(post, theta) { index <- which.max(post); mode<-theta[index]; print(mode)
}
```

Thus, candidate A is most likely to win given the poll results.

The 95% HPD interval in which the value of winning chance of candidate is 95% times likely to lie is found as [0.44, 0.62] with the following R-codes.

```
###  95% HPD interval
credMass = 0.95;  sortedPost = sort(post3 , decreasing=T)
HDIheightIdx = min(which(cumsum(sortedPost*step)>=credMass))
HDIheight = sortedPost[HDIheightIdx] # posterior probabilities at index
indices = which(post3>= HDIheight); HPD_Lni = min(theta[indices])
HPD_Uni = max(theta[indices]); show(data.frame(HPD_Lni, HPD_Uni))
```

We also found equal-tail 95% credible interval with the following R-codes as [0.4361343, 0.6227271].

```
###Equal-tail 95% credible interval
q_L = qbeta(0.025, 57.5, 51); q_U = qbeta(0.975, 57.5, 51); print(data.frame(q_L,q_U))
```

We can see that the 95% HPD interval and the 95% credible interval are the same, which is the case for symmetric posterior densities.

We also found the predictive posterior probability of a voter choosing for candidate A and B, with the following R-code, as 0.5305164 and 0.4694836, respectively. Thus, there are more chances of candidate A to be voted by a new voter.

```
###predictive posterior density
density_A = dbinom(1,1,theta); density_B = 1-density_A;
predictive_A = sum(density_A*post3*step); predictive_B = 1-predictive_A;
print(data.frame(predictive_A,predictive_B))
```

When applied to problems with multidimensional unknown parameters, Bayesian computation (evaluation of posterior densities, posterior estimates, posterior predictive distribution) can become analytically intractable. In solving Bayesian computation problems, simulation-based methods, such as MCMC techniques, and functional approximations such as Gaussian approximation, Laplace approximation, INLA, and variational Bayes approximations are widely used. However, the article does not include descriptions of these techniques.

Conclusion

Non-Bayesians frequently strive with the Bayesian concepts of uncertainty. The credible sources on these concepts either provide extensive theory or are primarily applied in nature. This article combines simple theoretical and applied approaches to understanding uncertainty using basic Bayesian concepts and R-codes. We hope that the article shall prove to be a productive yet straightforward platform for readers to deal with uncertainty in real-life scenarios with the notions of Bayesian statistics.

References

1. Berger, J. O. (1985). Statistical decision theory and Bayesian analysis; 2nd ed. Springer Series in Statistics. New York: Springer.
2. Fienberg, S. E. (2006). When did Bayesian inference become Bayesian? Bayesian Analysis 1 (1), 1–40.
3. Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2006). Bayesian Data Analysis. Chapman and Hall/CRC.
4. Gosh, J., Delampady, M. and Samanta, T. (2006). An Introduction to Bayesian Analysis:Theory and Methods (1 ed.). Springer–Verlag New York.
5. Lee, P. M. (2012). Bayesian Statistics: An Introduction (4th ed.). Wiley Publishing.
6. Oakley, J. E. and O'Hagan, A. (2007). Uncertainty in Prior Elicitations: A Nonparametric Approach. Biometrika 94 (2).
7. Stigler, S. M. (1982). Thomas Bayes' Bayesian inference. Journal of RSS 145 (2), 250–258.
8. Stigler, S. M. (1986). Laplace's 1774 memoir on inverse probability. Statistical Science 1 (3), 359–363.