Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol.9, No.1, February 2021, pp.156-161

Performance different uninformative efficiency of priors for binomial model

Inaam Rikan Hassan
University of Information Technology and Communications, Baghdad, Iraq
drinrh@uoitc.edu.iq

ABSTRACT

The current paper studies the performance efficiency of two uninformative priors, namely Bayes-Laplace (Uniform) prior and Jeffrey's prior for Binomial model. Several performance measures, such as the Bayes estimators under different loss functions, the posterior distribution skewness coefficient, the Bayesian point estimates, and the posterior variance, are used for comparison. Using these two uninformative priors, we conducted numerical simulation which showed that they perform extreme similarly.

Keywords: Uninformative prior, Skewness, Loss function, Bayes estimator, Posterior distribution.

Corresponding Author:

Dr. Inaam Rikan Hassan,
University of Information Technology and Communications, Baghdad, Iraq
E-mail:drinrh@uoitc.edu.iq

1. Introduction

The great development in the technologies used in our daily life dependson the sciences of mathematics and statistics for being the basis in many different fields and applications [1]–[6]. Bayesian inference contains several key parts. The Prior distribution represents one of them. It stands for the information among the uncertain parameter p. Besides, the posterior distribution is generated by combining the probability distribution with the prior distribution. Thus, the prior distribution is employed for future inference and decision p[7], [8].
A prior may be informative or uninformative. If the minimum effect of the previous distribution has a minimum effect on the parameter's subsequent distribution, then it is uninformative. Ingeneral, flat prior, diffuse, and vagueare additional names for the uninformative prior. In the prior distribution, if the researcher have confirmed beliefs about the hyper-parameters, then using informative prior, which reflects these beliefs will be a wise choice. In contrast, the researcher may have just unclear information about the interesting distribution of parameters ahead of observing the data. Therefore, he has to choose uninformative priors instead. There may be more than uninformative prior for a given problem. However, for more details, [7], [9] a review of several methods was introduced for deriving uninformative prior.
The posterior distributions of the Binomial model parameter, utilizing conjugate Beta prior, Jeffrey's prior, and uniform prior, are given in Section 2. Section 3 contains some simulated data. Section 4 introduces some numerical comparisons under various performance measures, such as posterior variance, coefficient of skewness, etc. Finally, Section 5 presents some concluding remarks.

2. The posterior distribution

The correspondings subsections present the Binomial model of subsequent distribution under conjugate Beta prior, uniform prior, and Jeffrey's prior.

2.1. Binomial distribution and conjugate Beta prior

The distribution of the number of successes x in m Bernoulli trials follows the Binomial distribution. Therefore, the posterior mass function (pmf) of the Binomial distribution for a random variable x with parameter p is:

$$p(x) = \binom{m}{x}(1-p)^{m-x} \ , x = 0, 1, 2, \ldots., m$$

$$0 < p < 1$$

For a simple casual sample of size n; x1, x2,, xn, the likelihood function is given by:

$$L(x_1, x_2, \ldots x_n|p) = \prod_{i=1}^{n}\binom{m}{xi}p^{xi}(1-p)^{m-xi}(2.1)$$

$$\propto p^{y}(1-p)^{nm-y} \text{ with } y = \sum_{i=1}^{n} xi(2.2)$$

The parameter p is unknown.
If the prior and posterior distributions are part of the same family, then the prior will be a conjugate prior for the distributions family, i.e., the posterior form has the same prior distributional form. The Binomial likelihood (2.2) has a conjugate Beta prior with probability density functions.

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}(2.3)$$

$$\alpha p^{\alpha-1}(1-p)^{\beta-1}\alpha > 0\ , \beta > 0$$

Using Bayes rule, the posterior distribution becomes as

$$\pi(p|y) \propto \pi(p).L(x_1, x_2, \ldots x_n|p)$$

So,

$$\pi(p|y) \propto p^{\alpha+y-1}(1-p)^{\beta+mn-y-1}(2.4)$$

This is the kernel of another Beta density

$$\pi(p|y) = \frac{\lceil(\alpha+\beta+mn)}{\lceil(\alpha+y)\lceil(\beta+mn)-y}p^{\alpha+y-1}(1-p)^{\beta+mn-y-1}$$

Or

$$\pi(p|y) = Beta(\alpha+y, \beta+mn-y)(2.5)$$

For this posterior distribution, the posterior mean is

$E(p|y) = \frac{\alpha+y}{\alpha+\beta+mn}(2.6) = \lambda\frac{\alpha}{\alpha+\beta} + (1-\lambda)\frac{y}{mn}$, with $\lambda = \frac{\alpha+\beta}{\alpha+\beta+mn}(2.7)$

When $\frac{\alpha}{\alpha+\beta}$ is the prior mean of p and $\frac{y}{mn}$ is the maximum likelihood estimates of p.

2.2. Bayes-Laplace (Uniform) prior [10], [11]

It is a particular situation of the Beta distribution, where U (0, 1) ≡ $Beta(1, 1)$. Thus, on the parameter space, the uniform prior of p is selected to be the constant

$\pi(p)$=1. The density kernel is:

$\pi(p) \propto 1 \quad , 0 < p < 1(2.8)$

The posterior distribution produced with a uniform U (0, 1) prior and a Binomial likelihood is:

$\pi(p|y) \propto p^{y}(1-p)^{mn-y}(2.9)$

Which is the Beta distribution density kernel with parameters $(y+1)\ and\ (mn-y+1)$.

Thus, the posterior distribution of p given data is:

$Beta(y+1, mn-y+1)$

2.3. Jeffrey's prior [12], [13]

Jeffrey's prior is an extremely useful prior. It achieves the local uniformity property, which means that a prior doesn't too much vary throughout the region, as well, the likelihood is important and doesn't presume big values out of the range. Moreover, it is founded on the matrix of fisher information. The definition of Jeffrey prior is:

$$\pi(p) \propto \left| I(p) \right|^{1/2} \quad (2.10)$$

Where $\left| \ \right|$ is the determinant, and $I(p)$ is the matrix of fisher knowledge based on the likelihood function$p(y|p)$:

$$I(p) = -E[\frac{\partial^2 \log p(y|p)}{\partial p^2} \quad (2.11)$$

From (2.2), we get

$$\log L = y \log p + (mn - y) \log(1-p) + \text{constant}$$

$$\frac{\partial^2 \log L}{\partial p^2} = \frac{-y}{p^2} - \frac{mn-y}{(1-p)^2}$$

So

$I(p) = \frac{nmp}{p^2} + \frac{mn-mnp}{(1-p)^2} = \frac{mn}{p(1-p)}$Where

$E(y) = nmp \quad (2.12)$

Taking the square root and removing the constantmn,

Gives $\pi(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} \quad (2.13)$

This is $Beta(1/2, 1/2)$ which is a special case of $Beta\ (\alpha\ , \beta)$ with $\alpha\ =\ 1/2$and$\beta = 1/2$.

The posterior distribution produced with Jeffery's prior and a binomial likelihood is

$$P\left(p\middle| y\right) \propto p^{y-1/2}(1-\ p)^{mn-y-1/2} \quad (2.14)$$

That represents the density kernel of the Beta distribution with parameter

$y + \frac{1}{2}, mn - y\ \ + \frac{1}{2}$.

Thus, the posterior distribution of p given data is

$$Beta\ (y + \frac{1}{2}, mn - y + \frac{1}{2}\ )$$

3. Simulated data

The following data of size $n = 5$ is generated from Binomial distribution with parameters $= 20, p = \frac{1}{2}$: 14, 9,12,10,12 using a routine written in C^{++} language (i.e$n = 5, \sum_{i=1}^{5} xi = 57$)

a. Under uniform prior
The posterior distribution of the parameter p for the given data$\underline{\times}= (x1, \ldots, x5)$, using (2.9) is a the distribution of Beta with parameters $\alpha = 58$ and $\beta = 44$, i.e $Beta(58,44)$.

b. Under Jeffrey's
The posterior distribution of the parameter p for the given data, using (2.14) is the Beta distribution with parameters $\alpha = 57.5$and$\beta = 43.5$, i.e $Beta(57.5,43.5)$.

4. Numerical comparisons

This section contains some numerical comparisons of the efficiency of both priors, under the above data, using the following performance measures:

4.1. Posterior variance

The posterior variance of parameter p with $Beta\ (\alpha, \beta)$ prior

$$\pi(p|\underset{\sim}{\overset{Var}{\times}})(p) = \frac{(\alpha + y)(mn - y + \beta)}{(mn + \alpha + \beta)^2(mn + \alpha + \beta + 1)}$$

a. Using uniform prior, where $\alpha = 1, \beta = 1$, the posterior variance is $Var(p) = 0.00238$.

b. Using Jeffery's prior, where $\alpha = {}^1/_2$, $\beta = {}^1/_2$, the posterior variance is $Var(p) = 0.00240$.

We notice from (a) and (b) that $Var(p)$ utilizing uniform and Jeffrey's prior which are approximately equal. We conclude that the uniform and Jeffrey's prior have approximately similar efficiency. However, the uniform prior is preferred for its simplicity.

4.2. Coefficient of skewness

The skewness coefficient of the posterior distribution is given by

$$\text{Coefficient of skewness } = \frac{2(mn - 2y)}{mn + \alpha + \beta + 2}\sqrt{\frac{mn + \alpha + \beta + 1}{(y + \alpha)(mn - y + \beta)}}$$

a. Coefficient of skewness for posterior distribution using a uniform prior $Beta(1,1)$ is -0.0538.

b. Coefficient of skewness for posterior distribution utilizing Jeffrey's prior $Beta(\frac{1}{2}, \frac{1}{2})$ is -0.0546.

From (a) and (b), we note that the coefficients of skewness are negative. They both are very slightly negatively and almost equally skewed. However, the uniform prior may be preferred to the Jeffrey's prior for its simplicity.

4.3. Bayesian point estimates

If $Beta(\alpha, \beta)$ is the prior, then the posterior mode is $\frac{\alpha+y-1}{mn+\alpha+\beta-2}$and the posterior mean is $\frac{\alpha+y}{mn+\alpha+\beta}$

a. If the prior is uniform prior, Beta(1,1), then the posterior mode is $\frac{y}{mn} = 0.5700$and the posterior mean is $\frac{y}{mn+2} = 0.5686$.

b. If the prior is Jeffrey's prior, Beta$\left(\frac{1}{2}, \frac{1}{2}\right)$, then the posterior mode is$\frac{y-\frac{1}{2}}{mn-1} = 0.5707$ and the posterior mean $\frac{y+\frac{1}{2}}{mn+1} = 0.5693$.

From these values, we notice that the posterior mode and posterior mean using the two priors are nearly the same as the maximum likelihood estimate, which is equal$\frac{y}{mn} = 0.5700$.

4.4. Bayes estimator (using loss function)

Using the loss function, the Bayes decision (the best decision), is the decision $(d*)$ that reduces the posterior expected loss function. If we consider the decision is an estimator choice, then Bayes decision is the Bayes estimator. In the following table, we present the Baye's estimator based on the uniform and Jeffrey's prior for different loss functions.

From the table, it can be seen that the Bayes estimator for the two-loss functions, using the two priors is almost equal to the maximum likelihood estimate (MLE=0.5700).

Table1. Bayes estimator for the two loss functions

Loss function $\boldsymbol{L(p,d)}$	Quadratic (squared error) $\boldsymbol{L_1(p,d) = (d-p)2}$	Relative quadratic squared error $\boldsymbol{L_2(p,d) = \frac{(d-p)^2}{p(1-p)}}$
Bayes Estimator $\boldsymbol{p_{\hat{B}} = d^*}$ With $\boldsymbol{Beta(\alpha,\beta)}$ prior	$d^* = \frac{\propto + y}{mn + \propto + \beta}$	$d^* = \frac{\propto + y - 1}{\propto + \beta + mn - 2}$
Jeffrey's prior $\boldsymbol{Beta}\,(\frac{1}{2},\frac{1}{2})$	$d^* = 0.693$	$d^* = 0.5707$
Uniform prior $\boldsymbol{Beta(1,1)}$	$d^* = 0.5686$	$d^* = 0.5700$

5. Conclusion

In this paper, we study the efficiency of two uninformative priors, namely Jeffrey's prior and uniform prior using several performance measures.
From previous sections, we note that

I. The two priors have nearly the same efficiency under all performance measures.
II. Uniform prior may be preferred because it is simpler than Jeffrey's prior. However, Jeffrey's prior has the invariant property.
III. The Bayes estimators are almost equal to the maximum likelihood estimator.

References

[1] A. H. Omran and Y. M. Abid, "Design of smart chess board that can predict the next position based on FPGA."

[2] A. Aljuboori, A. Hamza, and Y. Alasady, "Novel Intelligent Traffic Light System Using PSO and ANN," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, pp. 1528–1539, 2019.

[3] D. M. S. et al. Alaa Hamza Omran, "A Novel Intelligent Detection Schema of Series Arc Fault in Photovoltaic (PV) System Based Convolutional Neural Network," *Period. Eng. Nat. Sci.*, 2020.

[4] A. H. Omran, Y. M. Abid, and H. Kadhim, "Design of artificial neural networks system for intelligent chessboard," in *2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, Nov. 2017, Pp. 1–7, doi: 10.1109/ICETAS.2017.8277882.

[5] R. J. Yaser M. Abid, Alaa Hamza Omran, "Supervised feed forward neural networks for Smart chessboard based on FPGA," *J. Eng. Appl. Sci.*, vol. 13, no. 11, pp. 4093–4098, 2018.

[6] A. H. Omran, Y. M. Abid, A. S. Ahmed, H. Kadhim, and R. Jwad, "Maximizing the power of solar cells by using intelligent solar tracking system based on FPGA," in *2018 Advances in Science and Engineering Technology International Conferences (ASET)*, Feb. 2018, pp. 1–5, doi: 10.1109/ICASET.2018.8376786.

[7] J. O. Berger, *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[8] W. M. Bolstad and J. M. Curran, *Introduction to Bayesian statistics*. John Wiley & Sons, 2016.

[9] R. E. Kass and L. Wasserman, "The selection of prior distributions by formal rules," *J. Am. Stat. Assoc.*,

vol. 91, no. 435, pp. 1343–1370, 1996.

[10] A. Downey, *Think Bayes: Bayesian statistics in python*. " O'Reilly Media, Inc.," 2013.

[11] N. Sehgal and K. K. Pandey, "Artificial intelligence methods for oil price forecasting: a review and evaluation," *Energy Syst.*, vol. 6, no. 4, pp. 479–506, 2015.

[12] C. Gandrud, "simPH: an R package for showing estimates from Cox proportional hazard models including for interactive and nonlinear Effects," *J. Stat. Softw.*, vol. 65, no. 3, pp. 1–20, 2015.

[13] P. M. Todd and G. E. Gigerenzer, *Ecological rationality: Intelligence in the world.* Oxford University Press, 2012.