Revista Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 2 N°2 diciembre 2013 • ISSN 0121-747X • Research Paper
Páginas 38 a 42

LAPLACE PRIORI VERSUS JEFFREYS PRIORI: WHICH ONE?

A PRIORI DE LAPLACE VERSUS A PRIORI DE JEFFREYS: ¿CUÁL?

JUAN CARLOS CORREA MORALES [a], MARISOL VALENCIA CÁRDENAS [b]

Recibido 21-10-2013, aceptado 06-12-2013, versión final 12-12-2013.
Research Paper

ABSTRACT: A process to select a non-informative prior distribution makes the user to be faced to several possibilities and he/she thinks that only under some theoretical conditions one could be considered better than other, but in the applied problem this becomes irrelevant. We provide three basic examples that show that the choice is important and it depends on the underlying sample distributions, but also, on parameter values.

KEYWORDS: Priori distributions, Bayesian analysis.

RESUMEN: Un proceso para seleccionar una distribución a priori no informativa hace que el usuario se enfrente a varias posibilidades y que él/ella piense que sólo bajo ciertas condiciones teóricas una de estas podría ser considerada mejor que las otras, pero en el problema presentado ésto se vuelve irrelevante. Ofrecemos tres ejemplos básicos que muestran que la elección es importante y depende de las distribuciones muestrales subyacentes, pero también, en los valores de los parámetros.

PALABRAS CLAVE: Distribuciones a priori, análisis Bayesiano.

1. INTRODUCTION

Students or researchers in Bayesian topics are taught non-informative distributions in special Laplace and Jeffreys as alternatives to represent ignorance about the parameters. Many students assume that the choice of one or the other is irrelevant to the rest of the process and it is just a simple formality (Novick & Hall, 1969; Kass & Wasserman, 1996; Gill, 2002; Gelman, 2009; Zhu & Lu, 2004).
Tuyl *et al.* (2009) discuss the advantage of the Laplace prior under a binomial process as the predictive distribution. Berger & Bernardo (1992) provide a rigorous general definition, and give an explicit expression for the reference prior under very weak regularity conditions. We show that depending on the process that is being analyzed, it is possible to propose a method to choose an adequate non-informative prior, as a

[a] Ph. D. en Estadística, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.
[b] Ph. D.(c) en Ingeniería-Industria y Organizaciones, Universidad Nacional de Colombia, Sede Medellín.
email: mvalencia@unal.edu.co

pedagogical tool. This is done by using the mean integrated quadratic error (EMSE), measure used in many statistical techniques as an error quantification, but not too much in bayesian learning, we compare two of them in order to determine the less one.

2. BERNOULLI DISTRIBUTION

Le us assume that the interest is the estimation of an unknown π. We obtain a random sample from this population, say $x_1, x_2, \ldots, x_n$. The likelihood is

$$L(\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y},$$

where $y = \sum_{i=1}^n x_i$. If we select a Laplace prior, the posterior distribution is $Beta(y+1, n-y+1)$. If the choice is a Jeffreys prior, the posterior is $Beta(y+1/2, n-y+1/2)$.
In order to compare the two posteriors let us assume that the true population π is π_T. Then if we consider the expected mean square error (EMSE), where $\xi(\pi|y)$ is the posterior distribution of π given y.

$$EMSE = \sum_{y=0}^{n} \int_0^1 (\pi - \pi_T)^2 \xi(\pi|y) d\pi P(y|\pi_T)$$

$$EMSE = \sum_{y=0}^{n} (var(\pi|y) + (\mu(\pi|y) - \pi_T)^2) P(y|\pi_T).$$

For Laplace

$$EMSE_L = \sum_{y=0}^{n} \left(\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} + \left(\frac{y+1}{n+2} - \pi_T \right)^2 \right) \binom{n}{y} \pi_T^y (1-\pi_T)^{n-y}.$$

For Jeffreys

$$EMSE_J = \sum_{y=0}^{n} \left(\frac{(y+1/2)(n-y+1/2)}{(n+1)^2(n+2)} + \left(\frac{y+1/2}{n+1} - \pi_T \right)^2 \right) \binom{n}{y} \pi_T^y (1-\pi_T)^{n-y}.$$

Table 1 presents the difference between the Jeffreys expected mean square error and the Laplace expected mean square error for different true π's and different sample sizes. This difference changes, having the Jeffreys prior as a better choice than the Laplace prior only when the true parameter is less than 0.2 or greater than 0.8. Laplace prior is a better choice than the Jeffreys prior only when the true parameter is between (0.2, 0.8).

	True Population proportion							
n	0.05	0.1	0.15	0.2	0.5	0.8	0.9	0.95
5	-0.0122	-0.0071	-0.0027	0.0012	0.0119	0.0012	-0.0071	-0.0122
15	-0.0026	-0.0016	-0.0007	0.0001	0.0023	0.0001	-0.0016	-0.0026
25	-0.0011	-0.0007	-0.0003	0.0001	0.0010	0.0001	-0.0007	-0.0011
35	-0.0006	-0.0004	-0.0002	0.0000	0.0005	0.0000	-0.0004	-0.0006
45	-0.0004	-0.0002	-0.0001	0.0000	0.0003	0.0000	-0.0002	-0.0004
55	-0.0003	-0.0002	-0.0001	0.0000	0.0002	0.0000	-0.0002	-0.0003
65	-0.0002	-0.0001	-0.0000	0.0000	0.0002	0.0000	-0.0001	-0.0002
75	-0.0001	-0.0001	-0.0000	0.0000	0.0001	0.0000	-0.0001	-0.0001
85	-0.0001	-0.0001	-0.0000	0.0000	0.0001	0.0000	-0.0001	-0.0001
95	-0.0001	-0.0001	-0.0000	0.0000	0.0001	0.0000	-0.0001	-0.0001

Table 1: Jeffreys vs Laplace

3. POISSON DISTRIBUTION

If we are interested in the parameter of a Poisson, λ, and if we have a sample $x_1, \ldots, x_n$, the likelihood function is given by

$$L(\lambda) = \left(\prod_{i=1}^{n} x_i!\right)^{-1} \lambda^y \exp(-n\lambda),$$

where $y = \sum_{i=1}^{n} x_i$.

For the Laplace prior, the posterior distribution is $Gamma(y+1, n)$. If the choice is a Jeffreys prior, the posterior is $Gamma(y+1/2, n)$.

The true parameter of y is λ_T, then the $EMSE$ is defined as

$$EMSE = \sum_{y=0}^{\infty} \int_0^{\infty} (\lambda - \lambda_T)^2 \xi(\lambda|y)\, d\lambda\, P(y|\lambda_T).$$

For Laplace

$$EMSE_L = \sum_{y=0}^{\infty} \left(\frac{(y+1)}{n^2} + \left(\frac{y+1}{n} - \lambda_T \right)^2 \right) \frac{\lambda_T^y}{y!} \exp(-\lambda_T).$$

For Jeffreys

$$EMSE_J = \sum_{y=0}^{\infty} \left(\frac{(y+1/2)}{n^2} + \left(\frac{y+1/2}{n} - \lambda_T \right)^2 \right) \frac{\lambda_T^y}{y!} \exp(-\lambda_T).$$

The difference between these two EMSE's is

$$EMSE_L - EMSE_J = n^{-2}(5/4 + \lambda(1-n)).$$

In this case the Laplace prior is a better choice than the Jeffreys prior if $\lambda > 5/4$; if $\lambda < \frac{5/4}{n-1}$ and $n > \frac{5/4}{\lambda} + \lambda$, the conclusion is the same, as it is shown on Table 2; when $\lambda < \frac{5/4}{n-1}$, and $n < \frac{5/4}{\lambda} + \lambda$, Jeffreys is better than Laplace, as positive results show. But if $\lambda = \frac{5/4}{n-1}$, both have same EMSE.

	True λ							
n	1.25	0.625	0.417	0.313	0.25	0.208	0.1786	0.1563
2	0	0.156	0.208	0.234	0.25	0.26	0.2679	0.2734
3	-0.139	0	0.046	0.069	0.083	0.093	0.0992	0.1042
4	-0.156	-0.039	0	0.02	0.031	0.039	0.0446	0.0488
5	-0.15	-0.05	-0.017	0	0.01	0.017	0.0214	0.025
6	-0.139	-0.052	-0.023	-0.009	0	0.006	0.0099	0.013
7	-0.128	-0.051	-0.026	-0.013	-0.005	0	0.0036	0.0064
8	-0.117	-0.049	-0.026	-0.015	-0.008	-0.003	0	0.0024
9	-0.108	-0.046	-0.026	-0.015	-0.009	-0.005	-0.0022	0

Table 2: Jeffreys vs Laplace

In case of the exponential distribution $f(x) = \lambda_T \exp(-\lambda_T)$, the difference between these two EMSE's is

$$EMSE_L - EMSE_J = \lambda_T \frac{2}{n-1}\left(\frac{n+1}{n-2} - 1\right) > 0.$$

In this case the Jeffreys prior is a better choice than the Laplace prior.

4. CONCLUSIONS

We have shown that under non-informative distributions such as Laplace and Jeffreys the best choice depends upon the sampling distribution but also, on values of real parameter, which change the final result. The examples shown here are easy to implement in a Bayesian course.

References

Berger, M.; Bernardo, A. Y. (1992), On the development of reference priors. *Bayesian Statistics*, 4, 35–60.

Gelman, A. (2009), Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics. *Statistical Science*, 24(2), 176–178.

Gill, J. (2002), Bayesian Methods- A social and Behavioral Sciences Approach. Chapman and Hall/CRC, U.S.A., 459 p.

Kass, R. E.; Wasserman, L. (1996), The Selection Prior Distributions by Formal Rules. *Journal of the American Statistical Association*, 91(435), 1343–1370.

Novick, M. R.; Hall, W. J. (1969), A bayesian indifference procedure. *Journal of the American Statistical Association*, 60(312), 1104–1117.

Tuyl, F.; Gerlach, R.; Mengersen, K. (2009), Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters. *Bayesian Analysis*, 4(1), 151–158.

Zhu, M.; Lu, A. Y. (2004), The Counter-intuitive Non-informative Prior for the Bernoulli Family. *Journal of Statistics Education*, 12(2), 1–10.