T E C H N I C A L
R E P O R T

0438

A NOTE ON THE CHOICE BETWEEN TWO LOSS FUNCTIONS IN BAYESIAN ANALYSIS

L. WANG

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

A Note on the Choice between Two Loss Functions in Bayesian Analysis

Lichun Wang *
Center for Statistics, Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium

Abstract In the present paper, we propose a criterion which tells us how to choose a loss function in Bayesian analysis.
Key words: Bayesian analysis, Loss function, Posterior risk
AMS Subject Classification: 62C10

1 Introduction

Bayesian analysis is an important method of modern statistics, it almost appears in all important areas of statistical research and has been very useful in many fields of applications (see [1]). A main different point between Bayesian analysis and some classical statistical methods is that we use not only the sample information but also some information about the parameter θ in Bayesian analysis. Essence in the Baysian approach is to regard the parameter θ as a value of some random variable $\bar{\Theta}$ with a known distribution. Usually, given the states of a random variable X, a conditional probability is attached to this variable, say $f(x|\theta)$, and a prior density of the parameter θ, say $p(\theta)$, is specified based on previous knowledge.

In Bayesian analysis, the prior density and the sample information are combined via Bayes theorem to obtain the posterior density of θ given the sample information X, $p(\theta|x)$. The posterior density of θ is given by

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int_\Theta f(x|\theta)p(\theta)d\theta}. \tag{1.1}$$

*E-mail: lichun.wang@luc.ac.be

From the Bayesian point of view, the posterior distribution summarizes all the information about θ in the light of the data. Hence, any decisions about θ should be made based on the posterior density $p(\theta|x)$.

2 Propose A Criterion

It is easy to obtain the Bayes point estimate (BPE) of the parameter θ under the squared error loss function defined by

$$L_1(\theta, d) = (\theta - d)^2 \tag{2.1}$$

as

$$\hat{\theta}_{BPE} = E(\theta|X) = \frac{\int_\Theta \theta f(x|\theta)p(\theta)d\theta}{\int_\Theta f(x|\theta)p(\theta)d\theta}, \tag{2.2}$$

which minimizes the posterior expected loss defined by

$$E_{\theta|X}[L_1(\theta, d)] = \int_\Theta (\theta - d)^2 p(\theta|x)d\theta. \tag{2.3}$$

In fact, the posterior expected loss of $\hat{\theta}_{BPE}$ in this case is just the posterior variance. Before obtaining the sample X, a reasonable estimate of the parameter θ with respect to the prior density $p(\theta)$ is obviously $\hat{\theta} = E\theta$, which minimizes $E(\theta - d)^2$. Now, if we denote $E(\theta - \hat{\theta})^2$ and $E_{\theta|X}[L_1(\theta, \hat{\theta}_{BPE})]$, respectively, by $V_1(\theta)$ and $V_1(\theta|X)$, then, similar to [2], we can use the quantity $\omega_1(X) = V_1(\theta) - V_1(\theta|X)$ as a measure of the information provided by X about the parameter θ.

Since $\omega_1(X)$ is a random variable, we take its expectation

$$E[\omega_1(X)] = V_1(\theta) - E(V_1(\theta|X)) = V_1(E(\theta|X)) = V_1(\hat{\theta}_{BPE}) \tag{2.4}$$

as an average measure.

In above discussion, if we adopt the following weighted square loss function

$$L_2(\theta, d) = (\theta - d)^2/\theta^2, \tag{2.5}$$

then we first can get a estimator

$$\widetilde{\theta} = \frac{E\theta^{-1}}{E\theta^{-2}}, \tag{2.6}$$

which minimizes the risk $E[(\theta - d)^2\theta^{-2}]$. After obtaining the sample information X, under (2.5), the BPE of the parameter θ is

$$\widetilde{\theta}_{BPE} = \frac{E(\theta^{-1}|X)}{E(\theta^{-2}|X)}. \tag{2.7}$$

Hence, similar to (2.4), we can define an average measure of the information in X about θ as follows,

$$\begin{aligned} E[\omega_2(X)] &= E\{E[(\theta - \widetilde{\theta})^2\theta^{-2}] - E[(\theta - \widetilde{\theta}_{BPE})^2\theta^{-2}|X]\} \\ &= E[E^2(\theta^{-1}|X)E^{-1}(\theta^{-2}|X)] - E^2\theta^{-1}E^{-1}\theta^{-2}. \end{aligned} \tag{2.8}$$

For the above Bayesian analysis problem, which one we should take between the squared error loss and the weighted square loss function? Naturally, it is much better to adopt the loss function which can make the most use of the sample information X.

Thus, for the two loss functions (2.1) and (2.5), we follow the following steps: if

$$\frac{E[\omega_2(X)]}{E[(\theta - \widetilde{\theta})^2\theta^{-2}]} - \frac{E[\omega_1(X)]}{E(\theta - \hat{\theta})^2} \geq 0, \tag{2.9}$$

we adopt the weighted square loss function, otherwise, we take the squared loss.

Obviously, we can extend the above method to how to make a decison between any two loss functions such as $L^1(\theta, d)$ and $L^2(\theta, d)$.

Define

$$\omega_L(X) = E_\theta[L(\theta, \bar{\theta})] - E_{\theta|X}[L(\theta, \bar{\theta}_{BPE})], \tag{2.10}$$

where $\bar{\theta}$ is an estimator of θ which minimizes the loss $E_\theta[L(\theta, d)]$ with respect to a specified prior density of θ, and $\bar{\theta}_{BPE}$ is the Bayes point estimate under the loss function $L(\theta, d)$.

For any two loss functions $L^1(\theta, d)$ and $L^2(\theta, d)$, adopt $L^2(\theta, d)$ if

$$\frac{E[\omega_{L^2}(X)]}{E_\theta[L^2(\theta, \bar{\theta})]} - \frac{E[\omega_{L^1}(X)]}{E_\theta[L^1(\theta, \bar{\theta})]} \geq 0. \tag{2.11}$$

3 Two Examples

Example (I)

In this example, we make a comparison between the squared error loss function (2.1) and the weighted square loss function (2.5). Consider a random variable X with, X given θ, being distributed according to the following one-parameter scale exponential family

$$f(x|\theta) = \frac{1}{\Gamma(r)} x^{r-1}\theta^{-r}\exp(-x/\theta)I(x>0), \tag{3.1}$$

where $r > 0$ and $I(A)$ denotes the indicator of the set A.

Taking the prior density of the parameter θ as follows

$$p(\theta) = \frac{1}{\Gamma(b-2)}(\frac{1}{\theta})^{b-1}\exp(-1/\theta)I(b>2), \quad \theta > 0, \tag{3.2}$$

then, we can easily obtain the marginal density of X

$$\begin{aligned} f(x) &= \int_0^\infty f(x|\theta)p(\theta)d\theta = \frac{x^{r-1}}{\Gamma(r)\Gamma(b-2)}\int_0^\infty (\frac{1}{\theta})^{b+r-1}\exp(-(x+1)/\theta)d\theta \\ &= \frac{\Gamma(b+r-2)}{\Gamma(r)\Gamma(b-2)}\frac{x^{r-1}}{(x+1)^{b+r-2}}, \end{aligned} \tag{3.3}$$

and

$$E(\theta|X) = \frac{\int_0^\infty \theta f(x|\theta)p(\theta)d\theta}{f(x)} = \frac{x+1}{b+r-3}. \tag{3.4}$$

Hence, for $b > 4$, we have

$$\begin{aligned} E[\omega_1(X)] &= V_1\left(\frac{X+1}{b+r-3}\right) = E\left(\frac{X+1}{b+r-3}\right)^2 - \left[E\left(\frac{X+1}{b+r-3}\right)\right]^2 \\ &= \frac{\Gamma(b+r-2)\Gamma(b-4)}{\Gamma(b-2)\Gamma(b-4+r)(b+r-3)^2} - \left[\frac{\Gamma(b+r-2)}{\Gamma(b-2)(b+r-3)}\right]^2\left[\frac{\Gamma(b-3)}{\Gamma(b-3+r)}\right]^2 \\ &= \frac{b+r-4}{(b-3)(b-4)(b+r-3)} - \frac{1}{(b-3)^2}. \end{aligned} \tag{3.5}$$

On the other hand, using the following facts

$$
\begin{aligned}
E(\theta^{-1}) &= b-2, \quad E(\theta^{-2}) = (b-1)(b-2); && (3.6)\\
E(\theta^{-k}|X) &= \frac{\int_0^\infty \theta^{-k} f(x|\theta)p(\theta)d\theta}{f(x)} = \frac{\Gamma(b+k+r-2)}{\Gamma(b+r-2)(x+1)^k}, \quad k=1,2; && (3.7)
\end{aligned}
$$

we have

$$
E[\omega_2(X)] = \frac{(b+r-2)}{(b+r-1)} - \frac{b-2}{b-1}. \tag{3.8}
$$

Following the above discussion, by $E\theta = (b-3)^{-1}$ and $E\theta^2 = 1/[(b-3)(b-4)]$ and simple computation, we know, for $b > 4$ and any $r > 0$

$$
\frac{E[\omega_2(X)]}{E[(\theta-\widetilde{\theta})^2\theta^{-2}]} - \frac{E[\omega_1(X)]}{E(\theta-\hat{\theta})^2} = \frac{r}{b+r-1} - \frac{r}{b+r-3} = \frac{-2r}{(b+r-3)(b+r-1)} < 0. \tag{3.9}
$$

Hence, it is reasonable to take the squared loss function for the scale exponential family (3.1) and the prior distribution (3.2).

Example (II)

We assume that

$$
X|\theta \sim N(\theta, \sigma_1^2) \tag{3.10}
$$

and the prior distribution of the parameter θ is $N(\theta_0, \sigma_2^2)$, where σ_1^2, σ_2^2 and θ_0 are known constants.

Suppose that there are two certain loss fuctions: the squared error loss (2.1) and the absolute error loss $L(\theta, d) = |\theta - d|$.

It is easy to see that the marginal distribution of X is

$$
N(\theta_0, \sigma_1^2 + \sigma_2^2), \tag{3.11}
$$

and the posterior distribution of θ given $X = x$ is

$$
N\left(\frac{\sigma_1^2\theta_0 + \sigma_2^2 x}{\sigma_1^2+\sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right). \tag{3.12}
$$

First, under the squared error loss, by (2.10) and (2.11), since $\bar{\theta} = E\theta = \theta_0$ and $\bar{\theta}_{BPE} = \frac{\sigma_1^2\theta_0+\sigma_2^2 x}{\sigma_1^2+\sigma_2^2}$, one has

$$\frac{E[\omega_L(X)]}{E_\theta[L(\theta,\bar{\theta})]} = 1 - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}. \tag{3.13}$$

On the other hand, for the absolute error loss $L(\theta, d) = |\theta - d|$, the Bayes point estimate will be given by a median of the posterior distribution. This corresponds to the fact that, for a random variable X with $E|X| < \infty$: $E|X - a|$ is minimum when a is any median of X. Thus, we have

$$\frac{E[\omega_L(X)]}{E_\theta[L(\theta,\bar{\theta})]} = E\left[E_\theta|\theta-\theta_0| - E_{\theta|X}\left|\theta - \frac{\sigma_1^2\theta_0+\sigma_2^2 X}{\sigma_1^2+\sigma_2^2}\right|\right](E_\theta|\theta-\theta_0|)^{-1}. \tag{3.14}$$

Note that

$$E_\theta|\theta-\theta_0| = 2\int_0^\infty \frac{t}{\sqrt{2\pi\sigma_2^2}}\exp(-\frac{t^2}{2\sigma_2^2})dt = \frac{\sqrt{2}\sigma_2}{\sqrt{\pi}}, \tag{3.15}$$

and, similarly

$$E_{\theta|X}\left|\theta - \frac{\sigma_1^2\theta_0+\sigma_2^2 X}{\sigma_1^2+\sigma_2^2}\right| = \frac{\sqrt{2}}{\sqrt{\pi}} \times \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}, \tag{3.16}$$

we get, under the absolute error loss,

$$\frac{E[\omega_L(X)]}{E_\theta[L(\theta,\bar{\theta})]} = 1 - \sqrt{\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}}. \tag{3.17}$$

Comparing (3.13) with (3.17), we find it is much better to choice the squared error loss since it can better make use of the sample information. Only when $\sigma_2^2 = 0$, i.e., the parameter θ is a constant, it is the same to select the squared error loss or the absolute error loss, otherwise, we always should choice the squared error loss in Baysian analysis when confronted with the case (3.10).

Usually, in Bayesian analysis, it is not difficult to deal with the formula (2.11) and get some useful information to decide how to make a seletion between two loss functions.

4 Conclusion

In this paper, we define a criterion to describe the information provided by X about the parameter of interest θ. In Bayesian analysis, we can use the criterion to know which loss function make the most use of the sample information X.

It should be noted that our method can also be applied to many other distributions such as uniform distribution, exponential distribution, and so on, moreover, although many people like to adopt the squared error loss, it is easy to present one sample to show that it is not always better to choice the squared error loss in Bayesian analysis.

Finally, the readers are referred to literature [3], [4] and [5] and the references cited there for Bayesian analysis of the exponential distribution.

Acknowledgement Financial support from the IAP research network (# P5/24) of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

References

[1] Berger, J. O., Statistical Decision Theory and Bayesian Analysis, first edition, Chinese Staistical Publishing House, Beijing, 1998.

[2] Fraiwan Al-Saleh, M. and Masoud, Fawaz A., A note on the posterior expected loss as a measure of accuracy in Bayesian methods, *Applied Mathematics and Computation*, **134** (2003), 507-514.

[3] Anoop Chaturvedi et.al., Bayesian analysis of disturbances variance in the linear regression model under asymmetric loss functions, *Applied Mathematics and Computation*, **114** (2000), 149-153.

[4] Kubokawa, T. et. al., Bayes, minimax and nonnegative estimators of variance components under Kullback-Leibler loss, *J. Statist. Plann. Inference*, **86** (2000) , 201-214.

[5] Zellner, A., Bayesian estimation and prediction using asymmetric loss function, *Journal of American Satistical Association* , **81** (1986), 446-451.