The Bayesian Posterior Estimators under Six Loss Functions for Unrestricted and Restricted Parameter Spaces

Ying-Ying Zhang

Abstract

In this chapter, we have investigated six loss functions. In particular, the squared error loss function and the weighted squared error loss function that penalize overestimation and underestimation equally are recommended for the unrestricted parameter space $(-\infty, \infty)$; Stein's loss function and the power-power loss function, which penalize gross overestimation and gross underestimation equally, are recommended for the positive restricted parameter space $(0, \infty)$; the power-log loss function and Zhang's loss function, which penalize gross overestimation and gross underestimation equally, are recommended for $(0, 1)$. Among the six Bayesian estimators that minimize the corresponding posterior expected losses (PELs), there exist three strings of inequalities. However, a string of inequalities among the six smallest PELs does not exist. Moreover, we summarize three hierarchical models where the unknown parameter of interest belongs to $(0, \infty)$, that is, the hierarchical normal and inverse gamma model, the hierarchical Poisson and gamma model, and the hierarchical normal and normal-inverse-gamma model. In addition, we summarize two hierarchical models where the unknown parameter of interest belongs to $(0, 1)$, that is, the beta-binomial model and the beta-negative binomial model. For empirical Bayesian analysis of the unknown parameter of interest of the hierarchical models, we use two common methods to obtain the estimators of the hyperparameters, that is, the moment method and the maximum likelihood estimator (MLE) method.

1. Introduction

In Bayesian analysis, there are four basic elements: the data, the model, the prior, and the loss function. A Bayesian estimator minimizes some posterior expected loss (PEL) function. We confine our interests to six loss functions in this chapter: the squared error loss function (well known), the weighted squared error loss function ([1], p. 78), Stein's loss function [2–10], the power-power loss function [11], the power-log loss function [12], and Zhang's loss function [13]. It is worthy to note that among the six loss functions, the first and second loss functions are defined on $\Theta = (-\infty, \infty)$, and they penalize overestimation and

underestimation equally. The third and fourth loss functions are defined on $\Theta = (0, \infty)$, and they penalize gross overestimation and gross underestimation equally, that is, an action a will suffer an infinite loss when it tends to 0 or ∞. The fifth and sixth loss functions are defined on $\Theta = (0, 1)$, and they penalize gross overestimation and gross underestimation equally, that is, an action a will suffer an infinite loss when it tends to 0 or 1.

The squared error loss function and the weighted squared error loss function have been used by many authors for the problem of estimating the variance, σ^2, based on a random sample from a normal distribution with mean μ unknown (see, for instance, [14, 15]). As pointed out by [16], the two loss functions penalize equally for overestimation and underestimation, which is fine for the unrestricted parameter space $\Theta = (-\infty, \infty)$.

For $\Theta = (0, \infty)$, the positive restricted parameter space, where 0 is a natural lower bound and the estimation problem is not symmetric, we should not choose the squared error loss function and the weighted squared error loss function but choose a loss function which can penalize gross overestimation and gross underestimation equally, that is, an action a will suffer an infinite loss when it tends to 0 or ∞. Stein's loss function owns this property, and thus it is recommended for $\Theta = (0, \infty)$ by many researchers (e.g., see [2–10]). Moreover, [11] proposes the power-power loss function which not only penalizes gross overestimation and gross underestimation equally but also has balanced convergence rates or penalties for its argument too large and too small. Therefore, Stein's loss function and the power-power loss function are recommended for $\Theta = (0, \infty)$.

Analogously, for a restricted parameter space $\Theta = (0, 1)$, where 0 and 1 are two natural bounds and the estimation problem is not symmetric, we should not select the squared error loss function and the weighted squared error loss function but select a loss function which can penalize gross overestimation and gross underestimation equally, that is, an action a will suffer an infinite loss when it tends to 0 or 1. It is worthy to note that Stein's loss function and the power-power loss function are also not appropriate in this case. The power-log loss function proposed by [12] has this property. Moreover, they propose six properties for a good loss function on $\Theta = (0, 1)$. Specifically, the power-log loss function is convex in its argument, attains its global minimum at the true unknown parameter, and penalizes gross overestimation and gross underestimation equally. Apart from the six properties, [13] proposes the seventh property, that is, balanced convergence rates or penalties for the argument too large and too small, for a good loss function on $\Theta = (0, 1)$. Therefore, the power-log loss function and Zhang's loss function are recommended for $\Theta = (0, 1)$.

The rest of the chapter is organized as follows. In Section 2, we obtain two Bayesian estimators for $\theta \in \Theta = (-\infty, \infty)$ under the squared error loss function and the weighted squared error loss function. In Section 3, we obtain two Bayesian estimators for $\theta \in \Theta = (0, \infty)$ under Stein's loss function and the power-power loss function. In Section 4, we obtain two Bayesian estimators for $\theta \in \Theta = (0, 1)$ under the power-log loss function and Zhang's loss function. In Section 5, we summarize three strings of inequalities in a theorem. Some conclusions and discussions are provided in Section 6.

2. Bayesian estimation for $\theta \in (-\infty, \infty)$

There are two loss functions which are defined on $\Theta = (-\infty, \infty)$ and penalize overestimation and underestimation equally, that is, the squared error loss function (well known) and the weighted squared error loss function (see [1], p. 78).

2.1 Squared error loss function

The Bayesian estimator under the squared error loss function (well known), $\delta_2^\pi(\boldsymbol{x})$, minimizes the posterior expected squared error loss (PESEL), $\mathrm{E}[L_2(\theta, a)|\boldsymbol{x}]$, that is,

$$\delta_2^\pi(\boldsymbol{x}) = \arg\min_{a \in \mathcal{A}} \mathrm{E}[L_2(\theta, a)|\boldsymbol{x}], \tag{1}$$

where $\mathcal{A}\{a(\boldsymbol{x}) : a(\boldsymbol{x}) \in (-\infty, \infty)\}$ is the action space, $a = a(\boldsymbol{x}) \in (-\infty, \infty)$ is an action (estimator),

$$L_2(\theta, a) = (\theta - a)^2 \tag{2}$$

is the squared error loss function, and $\theta \in (-\infty, \infty)$ is the unknown parameter of interest. The PESEL is easy to obtain (see [16]):

$$PESEL(\pi, a|\boldsymbol{x}) = \mathrm{E}[L_2(\theta, a)|\boldsymbol{x}] = a^2 - 2a\mathrm{E}(\theta|\mathbf{x}) + \mathrm{E}(\theta^2|\boldsymbol{x}). \tag{3}$$

It is found in [16] that

$$\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x}) \tag{4}$$

by taking partial derivative of the PESEL with respect to a and setting it to 0.

2.2 Weighted squared error loss function

The Bayesian estimator under the weighted squared error loss function, $\delta_{w2}^\pi(\boldsymbol{x})$, minimizes the posterior expected weighted squared error loss (PEWSEL) (see [1]), $\mathrm{E}[L_{w2}(\theta, a)|\boldsymbol{x}]$, that is,

$$\delta_{w2}^\pi(\boldsymbol{x}) = \arg\min_{a \in \mathcal{A}} \mathrm{E}[L_{w2}(\theta, a)|\boldsymbol{x}], \tag{5}$$

where $\mathcal{A}\{a(\boldsymbol{x}) : a(\boldsymbol{x}) \in (-\infty, \infty)\}$ is the action space, $a = a(\boldsymbol{x}) \in (-\infty, \infty)$ is an action (estimator),

$$L_{w2}(\theta, a) = \frac{1}{\theta^2}(\theta - a)^2 \tag{6}$$

is the weighted squared error loss function, and $\theta \in (-\infty, \infty)$ is the unknown parameter of interest. The PEWSEL is easy to obtain (see [1]):

$$PEWSEL(\pi, a|\boldsymbol{x}) = \mathrm{E}[L_{w2}(\theta, a)|\boldsymbol{x}] = a^2 \mathrm{E}\left(\frac{1}{\theta^2}|\boldsymbol{x}\right) - 2a\mathrm{E}\left(\frac{1}{\theta}|\boldsymbol{x}\right) + 1. \tag{7}$$

It is found in [1] that

$$\delta_{w2}^\pi(\boldsymbol{x}) = \frac{\mathrm{E}\left(\frac{1}{\theta}|\boldsymbol{x}\right)}{\mathrm{E}\left(\frac{1}{\theta^2}|\boldsymbol{x}\right)} \tag{8}$$

by taking partial derivative of the PEWSEL with respect to a and setting it to 0.

3. Bayesian estimation for $\theta \in (0,\infty)$

There are many hierarchical models where the parameter of interest is $\theta \in \Theta = (0, \infty)$. As pointed out in the introduction, we should calculate and use the

Bayesian estimator of the parameter θ under Stein's loss function or the power-power loss function because they penalize gross overestimation and gross underestimation equally. We list several such hierarchical models as follows.

Model (a) (hierarchical normal and inverse gamma model). This hierarchical model has been investigated by [10, 16, 17]. Suppose that we observe $X_1, X_2, ..., X_n$ from the hierarchical normal and inverse gamma model:

$$\begin{cases} X_i|\theta \overset{\text{iid}}{\sim} N(\mu, \theta), & i = 1, 2, ..., n, \\ \theta \sim IG(\alpha, \beta), \end{cases} \tag{9}$$

where $-\infty < \mu < \infty$, $\alpha > 0$, and $\beta > 0$ are known constants, θ is the unknown parameter of interest, $N(\mu, \theta)$ is the normal distribution, and $IG(\alpha, \beta)$ is the inverse gamma distribution. It is worthy to note that the problem of finding the Bayesian rule under a conjugate prior is a standard problem and the problem is treated in almost every text on mathematical statistics. The idea of selecting an appropriate prior from the conjugate family was put forward by [18]. Specifically, Bayesian estimation of θ under the prior $IG(\alpha, \beta)$ is studied in Example 4.2.5 (p. 236) of [17] and in Exercise 7.23 (p. 359) of [16]. However, they only calculate the Bayesian estimator with respect to $IG(\alpha, \beta)$ prior under the squared error loss, $\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x})$.

Model (b) (hierarchical Poisson and gamma model). This hierarchical model has been investigated by [1, 16, 19, 20]. Suppose that $X_1, X_2, ..., X_n$ are observed from the hierarchical Poisson and gamma model:

$$\begin{cases} X_i|\theta \overset{\text{iid}}{\sim} P(\theta), & i = 1, 2, ..., n, \\ \theta \sim G(\alpha, \beta), \end{cases} \tag{10}$$

where $\alpha > 0$ and $\beta > 0$ are hyperparameters to be determined, $P(\theta)$ is the Poisson distribution with an unknown mean $\theta > 0$, and $G(\alpha, \beta)$ is the gamma distribution with an unknown shape parameter α and an unknown rate parameter β. The gamma prior $G(\alpha, \beta)$ is a conjugate prior for the Poisson model, so that the posterior distribution of θ is also a gamma distribution. The hierarchical Poisson and gamma model (10) has been considered in Exercise 4.32 (p. 196) of [4]. It has been shown that the marginal distribution of X is a negative binomial distribution if α is a positive integer. The Bayesian estimation of θ under the gamma prior is studied in [19] and in Tables 3.3.1 (p. 121) and 4.2.1 (p. 176) of [1]. However, they only calculated the Bayesian posterior estimator of θ under the squared error loss function.

Model (c) (hierarchical normal and normal-inverse-gamma model). This hierarchical model has been investigated by [2, 21, 22]. Let the observations $X_1, X_2, ..., X_n$ be from the hierarchical normal and normal-inverse-gamma model:

$$\begin{cases} X_i|(\mu, \theta) \overset{\text{iid}}{\sim} N(\mu, \theta), & i = 1, 2, ..., n, \\ \mu|\theta \sim N(\mu_0, \theta/\kappa_0), \theta \sim IG(v_0/2, v_0\sigma_0^2/2), \end{cases} \tag{11}$$

where $-\infty < \mu_0 < \infty$, $\kappa_0 > 0$, $v_0 > 0$, and $\sigma_0 > 0$ are known hyperparameters, $N(\mu, \theta)$ is a normal distribution with an unknown mean μ and an unknown variance θ, $\mu|\theta$ is $N(\mu_0, \theta/\kappa_0)$ which is a normal distribution, and θ is $IG(v_0/2, v_0\sigma_0^2/2)$ which is an inverse gamma distribution. More specifically, with a joint conjugate prior $\pi(\mu, \theta) \sim N - IG(\mu_0, \kappa_0, v_0, \sigma_0^2)$, which is the normal-inverse-gamma distribution, the posterior distribution of θ was studied in Example 1.5.1 (p. 20) of [21] and Part I (pp. 69–70) of [22]. However, they did not provide any Bayesian posterior

estimator of θ. Moreover, the normal distribution with a normal-inverse-gamma prior which assumes that μ is unknown is more realistic than the normal distribution with an inverse gamma prior investigated by [10] which assumes that μ is known.

3.1 Stein's loss function

3.1.1 One-dimensional case

The Bayesian estimator under Stein's loss function, $\delta_s^\pi(\boldsymbol{x})$, minimizes the posterior expected Stein's loss (PESL) (see [1, 10, 16]), $\mathrm{E}[L_s(\theta, a)|\boldsymbol{x}]$, that is,

$$\delta_s^\pi(\boldsymbol{x}) = \arg \min_{a \in \mathcal{A}} \mathrm{E}[L_s(\theta, a)|\boldsymbol{x}], \tag{12}$$

where $\mathcal{A}\{a(\boldsymbol{x}) : a(\boldsymbol{x}) > 0\}$ is the action space, $a = a(\boldsymbol{x}) > 0$ is an action (estimator),

$$L_s(\theta, a) = \frac{a}{\theta} - 1 - \log \frac{a}{\theta} \tag{13}$$

is Stein's loss function, and $\theta > 0$ is the unknown parameter of interest. The PESL is easy to obtain (see [10]):

$$PESL(\pi, a|\boldsymbol{x}) = \mathrm{E}[L_s(\theta, a)|\boldsymbol{x}] = a\mathrm{E}\left(\frac{1}{\theta}\Big|\boldsymbol{x}\right) - 1 - \log a + \mathrm{E}(\log \theta|\boldsymbol{x}). \tag{14}$$

It is found in [10] that

$$\delta_s^\pi(\boldsymbol{x}) = \frac{1}{\mathrm{E}\left(\frac{1}{\theta}|\boldsymbol{x}\right)} \tag{15}$$

by taking partial derivative of the PESL with respect to a and setting it to 0. The PESLs evaluated at the Bayesian estimators are (see [10])

$$\begin{aligned} PESL_s(\pi, \boldsymbol{x}) &= \mathrm{E}[L_s(\theta, a)|\boldsymbol{x}]\big|_{a=\delta_s^\pi(\boldsymbol{x})}, \\ PESL_2(\pi, \boldsymbol{x}) &= \mathrm{E}[L_s(\theta, a)|\boldsymbol{x}]\big|_{a=\delta_2^\pi(\boldsymbol{x})}, \end{aligned} \tag{16}$$

where $\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x})$ is the Bayesian estimator under the squared error loss function.

For the variance parameter θ of the hierarchical normal and inverse gamma model (9), [10] recommends and analytically calculates the Bayesian estimator:

$$\delta_s^\pi(\boldsymbol{x}) = \frac{1}{\alpha^* \beta^*}, \tag{17}$$

where

$$\alpha^* = \alpha + \frac{n}{2} \text{ and } \beta^* = \left[\frac{1}{\beta} + \frac{1}{2}\sum_{i=1}^n (x_i - \mu)^2\right]^{-1}, \tag{18}$$

with respect to $IG(\alpha, \beta)$ prior under Stein's loss function. This estimator minimizes the PESL. [10] also analytically calculates the Bayesian estimator,

$$\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x}) = \frac{1}{(\alpha^* - 1)\beta^*}, \tag{19}$$

with respect to $IG(\alpha, \beta)$ prior under the squared error loss, and the corresponding PESL. [10] notes that

$$\mathrm{E}(\log\theta|\boldsymbol{x}) = -\log\beta^* - \psi(\alpha^*), \tag{20}$$

which is essential for the calculation of

$$PESL_s(\pi, \boldsymbol{x}) = \log\alpha^* - \psi(\alpha^*) \tag{21}$$

and

$$PESL_2(\pi, \boldsymbol{x}) = \frac{1}{\alpha^* - 1} + \log(\alpha^* - 1) - \psi(\alpha^*), \tag{22}$$

depends on the digamma function $\psi(\cdot)$. Finally, the numerical simulations exemplify that $PESL_s(\pi, \boldsymbol{x})$ and $PESL_2(\pi, \boldsymbol{x})$ depend only on α and n and do not depend on μ, β, and $\boldsymbol{x}$; the estimators $\delta_2^\pi(\boldsymbol{x})$ are unanimously smaller than the estimators $\delta_s^\pi(\boldsymbol{x})$; and $PESL_s(\pi, \boldsymbol{x})$ are unanimously smaller than $PESL_2(\pi, \boldsymbol{x})$.

For the hierarchical Poisson and gamma model (43), [20] first calculates the posterior distribution of θ, $\pi(\theta|\boldsymbol{x})$, and the marginal pmf of x, $\pi(x)$, in Theorem 1 of their paper. [20] then calculates the Bayesian posterior estimators $\delta_s^\pi(\boldsymbol{x})$ and $\delta_2^\pi(\boldsymbol{x})$, and the PESLs $PESL_s(\pi, \boldsymbol{x})$ and $PESL_2(\pi, \boldsymbol{x})$, and they satisfy two inequalities. After that, the estimators of the hyperparameters of the model (10) by the moment method $\alpha_1(n)$ and $\beta_1(n)$ are summarized in Theorem 2 of their paper. Moreover, the estimators of the hyperparameters of the model (10) by the maximum likelihood estimator (MLE) method $\alpha_2(n)$ and $\beta_2(n)$ are summarized in Theorem 3 of their paper. Finally, the empirical Bayesian estimators of the parameter of the model (10) under Stein's loss function by the moment method and the MLE method are summarized in Theorem 4 of their paper. In numerical simulations of [20], they have illustrated the two inequalities of the Bayesian posterior estimators and the PESLs, the moment estimators and the MLEs are consistent estimators of the hyperparameters, and the goodness of fit of the model to the simulated data. The numerical results indicate that the MLEs are better than the moment estimators when estimating the hyperparameters. Finally, [20] exploits the attendance data on 314 high school juniors from two urban high schools to illustrate their theoretical studies.

For the variance parameter θ of the normal distribution with a normal-inverse-gamma prior (11), [23] recommends and analytically calculates the Bayesian posterior estimator, $\delta_s^\pi(\boldsymbol{x})$, with respect to a conjugate prior $\mu|\theta \sim N(\mu_0, \theta/\kappa_0)$, and $\theta \sim IG(v_0/2, v_0\sigma_0^2/2)$ under Stein's loss function which penalizes gross overestimation and gross underestimation equally. This estimator minimizes the PESL. As comparisons, the Bayesian posterior estimator, $\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x})$, with respect to the same conjugate prior under the squared error loss function, and the PESL at $\delta_2^\pi(\boldsymbol{x})$, are calculated. The calculations of $\delta_s^\pi(\boldsymbol{x})$, $\delta_2^\pi(\boldsymbol{x})$, $PESL_s(\pi, \boldsymbol{x})$, and $PESL_2(\pi, \boldsymbol{x})$ depend only on $\mathrm{E}(\theta|\boldsymbol{x})$, $\mathrm{E}(\theta^{-1}|\boldsymbol{x})$, and $\mathrm{E}(\log\theta|\boldsymbol{x})$. The numerical simulations exemplify their theoretical studies that the PESLs depend only on v_0 and n, but do not depend on μ_0, κ_0, σ_0, and especially $\boldsymbol{x}$. The estimators $\delta_2^\pi(\boldsymbol{x})$ are unanimously larger than the estimators $\delta_s^\pi(\boldsymbol{x})$, and $PESL_2(\pi, \boldsymbol{x})$ are unanimously larger

than $PESL_s(\pi, \boldsymbol{x})$. Finally, [23] calculates the Bayesian posterior estimators and the PESLs of the monthly simple returns of the Shanghai Stock Exchange (SSE) Composite Index, which also exemplify the theoretical studies of the two inequalities of the Bayesian posterior estimators and the PESLs.

3.1.2 Multidimensional case

For estimating a covariance matrix which is assumed to be positive definite, many researchers exploit the multidimensional Stein's loss function (e.g., see [2, 8, 24–31]). The multidimensional Stein's loss function (see [2]) is originally defined to estimate the $p \times p$ unknown covariance matrix Σ by $\hat{\Sigma}$ with the loss function:

$$L(\Sigma, \hat{\Sigma}) = \mathrm{tr}\Sigma^{-1}\hat{\Sigma} - \log\det\Sigma^{-1}\hat{\Sigma} - p. \tag{23}$$

When $p = 1$, the multidimensional Stein's loss function reduces to

$$L_s(\sigma^2, a) = \frac{a}{\sigma^2} - \log\frac{a}{\sigma^2} - 1, \tag{24}$$

which is in the form of (13), the one-dimensional Stein's loss function.

3.2 Power-power loss function

The Bayesian estimator under the power-power loss function, $\delta_p^\pi(\boldsymbol{x})$, minimizes the posterior expected power-power loss (PEPL) (see [11]), $\mathrm{E}[L_p(\theta, a)|\boldsymbol{x}]$, that is,

$$\delta_p^\pi(\boldsymbol{x}) = \arg\min_{a \in \mathcal{A}} \mathrm{E}[L_p(\theta, a)|\boldsymbol{x}], \tag{25}$$

where $\mathcal{A}\{a(\boldsymbol{x}) : a(\boldsymbol{x}) > 0\}$ is the action space, $a = a(\boldsymbol{x}) > 0$ is an action (estimator),

$$L_p(\theta, a) = \frac{a}{\theta} + \frac{\theta}{a} - 2 \tag{26}$$

is the power-power loss function, and $\theta > 0$ is the unknown parameter of interest. The PEPL is easy to obtain (see [11]):

$$PEPL(\pi, a|\boldsymbol{x}) = \mathrm{E}[L_p(\theta, a)|\boldsymbol{x}] = a\mathrm{E}\left(\frac{1}{\theta}|\boldsymbol{x}\right) + \frac{1}{a}\mathrm{E}(\theta|\boldsymbol{x}) - 2. \tag{27}$$

It is found in [11] that

$$\delta_p^\pi(\boldsymbol{x}) = \sqrt{\frac{\mathrm{E}(\theta|\boldsymbol{x})}{\mathrm{E}\left(\frac{1}{\theta}|\boldsymbol{x}\right)}} \tag{28}$$

by taking partial derivative of the PEPL with respect to a and setting it to 0. The PEPLs evaluated at the Bayesian estimators are (see [11])

$$PEPL_p(\pi, \boldsymbol{x}) = \mathrm{E}[L_p(\theta, a)|\boldsymbol{x}]\big|_{a=\delta_p^\pi(\boldsymbol{x})},$$
$$PEPL_2(\pi, \boldsymbol{x}) = \mathrm{E}[L_p(\theta, a)|\boldsymbol{x}]\big|_{a=\delta_2^\pi(\boldsymbol{x})}. \tag{29}$$

The power-power loss function is proposed in [11], and it has all the seven properties proposed in his paper. More specifically, it penalizes gross overestimation and gross underestimation equally, is convex in its argument, and has balanced convergence rates or penalties for its argument too large and too small. Therefore, it is recommended for the positive restricted parameter space $\Theta = (0, \infty)$.

4. Bayesian estimation for $\theta \in (0, 1)$

There are some hierarchical models where the unknown parameter of interest is $\theta \in \Theta = (0, 1)$. As pointed out in the introduction, we should calculate and use the Bayesian estimator of the parameter θ under the power-log loss function or Zhang's loss function because they penalize gross overestimation and gross underestimation equally. We list two such hierarchical models as follows.

Model (d) (beta-binomial model). This hierarchical model has been investigated by [1, 12, 13, 16, 32, 33]. Suppose that $X_1, X_2, ..., X_n$ are from the beta-binomial model:

$$\begin{cases} X_i|\theta \overset{\text{iid}}{\sim} Bin(m, \theta), & i = 1, 2, ..., n, \\ \theta \sim Be(\alpha, \beta), \end{cases} \tag{30}$$

where $\alpha > 0$ and $\beta > 0$ are known constants, m is a known positive integer, $\theta \in (0, 1)$ is the unknown parameter of interest, $Be(\alpha, \beta)$ is the beta distribution, and $Bin(m, \theta)$ is the binomial distribution. Specifically, Bayesian estimation of θ under the prior $Be(\alpha, \beta)$ is studied in Example 7.2.14 (p. 324) of [16] and in Tables 3.3.1 (p. 121) and 4.2.1 (p. 176) of [1]. However, they only calculate the Bayesian estimator with respect to $Be(\alpha, \beta)$ prior under the squared error loss, $\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x})$. Moreover, they only consider one observation. The beta-binomial model has been investigated recently. For instance, [32] uses the beta-binomial to draw the random removals in progressive censoring; [12, 13] use the beta-binomial to model some magazine exposure data for the monthly magazine *Signature*; [33] develops estimation procedure for the parameters of a zero-inflated overdispersed binomial model in the presence of missing responses.

Model (e) (beta-negative binomial model). This hierarchical model has been investigated by [1, 34]. Suppose that $X_1, X_2, ..., X_n$ are from the beta-negative binomial model:

$$\begin{cases} X_i|\theta \overset{\text{iid}}{\sim} NB(m, \theta), & i = 1, 2, ..., n, \\ \theta \sim Be(\alpha, \beta), \end{cases} \tag{31}$$

where $\alpha > 0$ and $\beta > 0$ are known constants, m is a known positive integer, $\theta \in (0, 1)$ is the unknown parameter of interest, $Be(\alpha, \beta)$ is the beta distribution, and $NB(m, \theta)$ is the negative binomial distribution. Specifically, Bayesian estimation of θ under the prior $Be(\alpha, \beta)$ is studied in Tables 3.3.1 (p. 121) and 4.2.1 (p. 176) of [1]. However, he only calculates the Bayesian estimator with respect to $Be(\alpha, \beta)$ prior under the squared error loss function, $\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x})$. Moreover, he only considers one observation.

4.1 Power-log loss function

A good loss function $L(\theta, a) = L(a|\theta) = L(x)|_{x=a/\theta}$ for $\Theta = (0, 1)$ should have the six properties summarized in **Table 1** (see **Table 1** in [12]).

In **Table 1**, property (a) means that any action a of the parameter θ should incur a nonnegative loss. Property (b) means that when $x = a/\theta = 1$, or $a = \theta$, that is, a correctly estimates θ, the loss is 0. Property (c) means that when $x = a/\theta \to (1/\theta)^-$, that is, a is moving away from θ and tends to 1^-, it will incur an infinite loss. Property (d) means that when $x = a/\theta \to 0^+$, that is, a is moving away from θ and tends to 0^+, it will also incur an infinite loss. Properties (c) and (d) mean that the loss function will penalize gross overestimation and gross underestimation equally. Property (e) is useful in the proofs of some propositions of the minimaxity and the admissibility of the Bayesian estimator (see [1]). Property (f) means that 1 and θ are the local extrema of $L(x)$ and $L(a|\theta)$, respectively. Property (f) also implies that $L(\theta + \Delta a|\theta) = o(\Delta a)$, that is, the loss incurred by an action $a = \theta + \Delta a$ near θ ($\Delta a \approx 0$), is very small compared to Δa.

Let

$$g_{pl}(x) = \frac{\left(\frac{1}{\theta} - 1\right)^2}{\frac{1}{\theta} - x} - \log x \text{ and } g_{pl}(1) = \frac{1}{\theta} - 1. \tag{32}$$

Define

$$L_{pl}(x) = g_{pl}(x) - g_{pl}(1) = \frac{\left(\frac{1}{\theta} - 1\right)^2}{\frac{1}{\theta} - x} - \log x - \left(\frac{1}{\theta} - 1\right). \tag{33}$$

Thus

$$L_{pl}(\theta, a) = L_{pl}(a|\theta) = L_{pl}(x)\big|_{x=a/\theta} = \frac{\left(\frac{1}{\theta} - 1\right)^2}{\frac{1}{\theta} - \frac{a}{\theta}} - \log \frac{a}{\theta} - \left(\frac{1}{\theta} - 1\right)$$

$$= \frac{\theta\left(\frac{1}{\theta} - 1\right)^2}{1 - a} - \log a + \log \theta - \left(\frac{1}{\theta} - 1\right). \tag{34}$$

It is easy to check (see the supplement of [12]) that $L_{pl}(\theta, a) = L_{pl}(a|\theta) = L_{pl}(x)\big|_{x=a/\theta}$, which is called the power-log loss function, satisfies all the six properties listed in **Table 1**. Consequently, the power-log loss function is a good loss function for $\Theta = (0, 1)$, and thus it is recommended for $\Theta = (0, 1)$.

We remark that the power-log loss function on $\Theta = (0, 1)$ is an analog of the power-log loss function on $\Theta = (0, \infty)$, which is the popular Stein's loss function.

Properties	$L(x)$	$L(a|\theta)$
(a)	$L(x) \geq 0$ for all $0 < x < \frac{1}{\theta}$	$L(a|\theta) \geq 0$ for all $0 < a < 1$
(b)	$L(1) = 0$	$L(\theta|\theta) = L(a|\theta)\big|_{a=\theta} = 0$
(c)	$L\left(\left(\frac{1}{\theta}\right)^-\right) = \lim_{x \to \left(\frac{1}{\theta}\right)^-} L(x) = \infty$	$L(1^-|\theta) = \lim_{a \to 1^-} L(a|\theta) = \infty$
(d)	$L(0^+) = \lim_{x \to 0^+} L(x) = \infty$	$L(0^+|\theta) = \lim_{a \to 0^+} L(a|\theta) = \infty$
(e)	Convex in x for all $0 < x < \frac{1}{\theta}$	Convex in a for all $0 < a < 1$
(f)	$L'(1) = \frac{dL(x)}{dx}\big|_{x=1} = 0$	$\left[\frac{\partial}{\partial a} L(a|\theta)\right]\big|_{a=\theta} = 0$

Table 1.
(Table 1 in [12]) The six properties of a good loss function for $\Theta = (0, 1)$. $0 < \theta < 1$ is fixed.

The Bayesian estimator under the power-log loss function, $\delta_{pl}^{\pi}(\boldsymbol{x})$, minimizes the posterior expected power-log loss (PEPLL) (see [12]), $\mathrm{E}\left[L_{pl}(\theta, a)|\boldsymbol{x}\right]$, that is,

$$\delta_{pl}^{\pi}(\boldsymbol{x}) = \arg\min_{a \in \mathcal{A}} \mathrm{E}\left[L_{pl}(\theta, a)|\boldsymbol{x}\right], \tag{35}$$

where $\mathcal{A}\{a(\boldsymbol{x}) : a(\boldsymbol{x}) \in (0, 1)\}$ is the action space, $a = a(\boldsymbol{x}) \in (0, 1)$ is an action (estimator), $L_{pl}(\theta, a)$ given by (34) is the power-log loss function, and $\theta \in (0, 1)$ is the unknown parameter of interest. The PEPLL is easy to obtain (see [12]):

$$PEPLL(\pi, a|\boldsymbol{x}) = \mathrm{E}\left[L_{pl}(\theta, a)|\boldsymbol{x}\right] = \frac{E_1(\boldsymbol{x})}{1 - a} - \log a + E_2(\boldsymbol{x}) - E_3(\boldsymbol{x}) + 1, \tag{36}$$

where

$$E_1(\boldsymbol{x}) = \mathrm{E}\left[\theta^{-1}(1 - \theta)^2|\boldsymbol{x}\right] > 0,$$
$$E_2(\boldsymbol{x}) = \mathrm{E}[\log \theta|\boldsymbol{x}] < 0, \tag{37}$$
$$E_3(\boldsymbol{x}) = \mathrm{E}\left[\theta^{-1}|\boldsymbol{x}\right] > 0.$$

It is found in [12] that

$$\delta_{pl}^{\pi}(\boldsymbol{x}) = \frac{2 + E_1(\boldsymbol{x}) - \sqrt{E_1(\boldsymbol{x})(E_1(\boldsymbol{x}) + 4)}}{2} \tag{38}$$

by taking partial derivative of the PEPLL with respect to a and setting it to 0. The PEPLLs evaluated at the Bayesian estimators are (see [12])

$$PEPLL_{pl}(\pi, \boldsymbol{x}) = \mathrm{E}\left[L_{pl}(\theta, a)|\boldsymbol{x}\right]\Big|_{a = \delta_{pl}^{\pi}(\boldsymbol{x})},$$
$$\tag{39}$$
$$PEPLL_2(\pi, \boldsymbol{x}) = \mathrm{E}\left[L_{pl}(\theta, a)|\boldsymbol{x}\right]\Big|_{a = \delta_2^{\pi}(\boldsymbol{x})}.$$

Finally, the numerical simulations and a real data example of some monthly magazine exposure data (see [35]) exemplify the theoretical studies of two size relationships about the Bayesian estimators and the PEPLLs in [12].

4.2 Zhang's loss function

Zhang et al. [12] proposed six properties for a good loss function $L(\theta, a) = L(a|\theta) = L(x)|_{x=a/\theta}$ on $\Theta = (0, 1)$. Apart from the six properties, [13] proposes the seventh property (balanced convergence rates or penalties for the argument too large and too small) for a good loss function on $\Theta = (0, 1)$. Moreover, the seven properties for a good loss function on $\Theta = (0, 1)$ are summarized in **Table 1** of [13]. The explanations of the first six properties in **Table 1** of [13] can be found in the previous subsection (see also [12]). In **Table 1** of [13], property (g) (the seventh property) means that $L\left(k_1(\theta)\frac{1}{n}\right)$ and $L\left(\frac{1}{\theta}\left(1 - \frac{1}{n}\right)\right)$ tend to ∞ at the same rate and $L\left(k_2(\theta)\frac{1}{n}|\theta\right)$ and $L\left(1 - \frac{1}{n}|\theta\right)$ tend to ∞ at the same rate. In other words,

$$\lim_{n \to \infty} \frac{L\left(k_1(\theta)\frac{1}{n}\right)}{L\left(\frac{1}{\theta}\left(1 - \frac{1}{n}\right)\right)} = 1 \text{ and } \lim_{n \to \infty} \frac{L\left(k_2(\theta)\frac{1}{n}|\theta\right)}{L\left(1 - \frac{1}{n}|\theta\right)} = 1. \tag{40}$$

And they say that $L\left(k_1(\theta)\frac{1}{n}\right)$ and $L\left(\frac{1}{\theta}\left(1-\frac{1}{n}\right)\right)$ are asymptotically equivalent. Similarly, $L\left(k_2(\theta)\frac{1}{n}|\theta\right)$ and $L\left(1-\frac{1}{n}|\theta\right)$ are said to be asymptotically equivalent. They also say that $L(x)$ $(L(a|\theta))$ has balanced convergence rates or penalties for x (a) too large and too small. It is worthy to note that $k_1(\theta)\frac{1}{n}\to 0$ and $\frac{1}{\theta}\left(1-\frac{1}{n}\right)\to\frac{1}{\theta}$ at the same order $O\left(\frac{1}{n}\right)$. Analogously, $k_2(\theta)\frac{1}{n}\to 0$ and $1-\frac{1}{n}\to 1$ at the same order $O\left(\frac{1}{n}\right)$. Finally, only when properties (c) and (d) hold, property (g) may hold.

Let

$$g_z(x) = \frac{1}{\left(\frac{1}{\theta}-1\right)^2 x} + \frac{1}{\frac{1}{\theta}-x} \text{ and } g_z(1) = \frac{1}{\theta\left(\frac{1}{\theta}-1\right)^2}. \tag{41}$$

Let

$$L_z(x) = g_z(x) - g_z(1) = \frac{1}{\left(\frac{1}{\theta}-1\right)^2 x} + \frac{1}{\frac{1}{\theta}-x} - \frac{1}{\theta\left(\frac{1}{\theta}-1\right)^2}. \tag{42}$$

Thus

$$L_z(\theta,a) = L_z(a|\theta) = L_z(x)|_{x=a/\theta} = \frac{1}{\left(\frac{1}{\theta}-1\right)^2 \frac{a}{\theta}} + \frac{1}{\frac{1}{\theta}-\frac{a}{\theta}} - \frac{1}{\theta\left(\frac{1}{\theta}-1\right)^2}$$

$$= \frac{\theta}{\left(\frac{1}{\theta}-1\right)^2 a} + \frac{\theta}{1-a} - \frac{1}{\theta\left(\frac{1}{\theta}-1\right)^2}. \tag{43}$$

It is easy to check (see the supplement of [13]) that $L_z(\theta,a) = L_z(a|\theta) = L_z(x)|_{x=a/\theta}$, which is called Zhang's loss function, satisfies all the seven properties listed in **Table 1** of [13]. Consequently, Zhang's loss function is a good loss function, and thus it is recommended for $\Theta = (0,1)$.

The Bayesian estimator under Zhang's loss function, $\delta_z^\pi(\boldsymbol{x})$, minimizes the posterior expected Zhang's loss (PEZL) (see [13]), $E[L_z(\theta,a)|\boldsymbol{x}]$, that is,

$$\delta_z^\pi(\boldsymbol{x}) = \arg\min_{a\in\mathcal{A}} E[L_z(\theta,a)|\boldsymbol{x}], \tag{44}$$

where $\mathcal{A}\{a(\boldsymbol{x}) : a(\boldsymbol{x})\in(0,1)\}$ is the action space, $a = a(\boldsymbol{x})\in(0,1)$ is an action (estimator), $L_z(\theta,a)$ given by (43) is Zhang's loss function, and $\theta\in(0,1)$ is the unknown parameter of interest. The PEZL is easy to obtain (see [13]):

$$PEZL(\pi,a|\boldsymbol{x}) = E[L_z(\theta,a)|\boldsymbol{x}] = \frac{E_1(\boldsymbol{x})}{a} + \frac{E_2(\boldsymbol{x})}{1-a} - E_3(\boldsymbol{x}), \tag{45}$$

where

$$E_1(\boldsymbol{x}) = E\left[\frac{\theta^3}{(1-\theta)^2}|\boldsymbol{x}\right],$$

$$E_2(\boldsymbol{x}) = E(\theta|\boldsymbol{x}), \tag{46}$$

$$E_3(\boldsymbol{x}) = E\left[\frac{\theta}{(1-\theta)^2}|\boldsymbol{x}\right].$$

It is found in [13] that

$$\delta_z^\pi(\boldsymbol{x}) = \frac{\sqrt{E_1(\boldsymbol{x})}}{\sqrt{E_1(\boldsymbol{x})} + \sqrt{E_2(\boldsymbol{x})}} \tag{47}$$

by taking partial derivative of the PEZL with respect to a and setting it to 0. The PEZLs evaluated at the Bayesian estimators are (see [13])

$$PEZL_z(\pi, \boldsymbol{x}) = \mathrm{E}[L_z(\theta, a)|\boldsymbol{x}]|_{a=\delta_z^\pi(\boldsymbol{x})},$$

$$PEZL_2(\pi, \boldsymbol{x}) = \mathrm{E}[L_z(\theta, a)|\boldsymbol{x}]|_{a=\delta_2^\pi(\boldsymbol{x})}. \tag{48}$$

Zhang et al. [13] considers an example of some magazine exposure data for the monthly magazine *Signature* (see [12, 35]) and compares the numerical results with those of [12].

For the probability parameter θ of the beta-negative binomial model (31), [34] recommends and analytically calculates the Bayesian estimator $\delta_z^\pi(\boldsymbol{x})$, with respect to $Be(\alpha, \beta)$ prior under Zhang's loss function which penalizes gross overestimation and gross underestimation equally. This estimator minimizes the PEZL. They also calculate the usual Bayesian estimator $\delta_2^\pi(\boldsymbol{x}) = \mathrm{E}(\theta|\boldsymbol{x})$ which minimizes the PESEL. Moreover, they also obtain the PEZLs evaluated at the two Bayesian estimators, $PEZL_z(\pi, \boldsymbol{x})$ and $PEZL_2(\pi, \boldsymbol{x})$. After that, they show two theorems about the estimators of the hyperparameters of the beta-negative binomial model (31) when m is known or unknown by the moment method (Theorem 1 in [34]) and the MLE method (Theorem 2 in [34]). Finally, the empirical Bayesian estimator of the probability parameter θ under Zhang's loss function is obtained with the hyperparameters estimated by the moment method or the MLE method from the two theorems.

In the numerical simulations of [34], they have illustrated three things: the two inequalities of the Bayesian posterior estimators and the PEZLs, the moment estimators and the MLEs, which are consistent estimators of the hyper-parameters, and the goodness of fit of the beta-negative binomial model to the simulated data. Numerical simulations show that the MLEs are better than the moment estimators when estimating the hyperparameters in terms of the good-ness of fit of the model to the simulated data. However, the MLEs are very sensitive to the initial estimators, and the moment estimators are usually proved to be good initial estimators.

In the real data section of [34], they consider an example of some insurance claim data, which are assumed from the beta-negative binomial model (31). They consider four cases to fit the real data. In the first case, they assume that $m = 6$ is known for illustrating purpose (of course, one can assume another known m value). In the other three cases, they assume that m is unknown, and they provide three approaches to handle this scenario. The first two approaches consider a range of m values, for instance, $m = 1, 2, ..., 20$. The first approach is to maximize the log-likelihood function. The second approach is to maximize the p-value of the goodness of fit of the model (31) to the real data. The third approach is to determine the hyperparameters α, β, and m from Theorems 1 and 2 in [34] by the moment method and the MLE method, respectively, when m is unknown. Four tables which show the number of claims, the observed frequencies, the expected probabilities, and the expected frequencies of the insurance claims data are provided to illustrate the four cases.

5. Inequalities among Bayesian posterior estimators

For the six loss functions, we have the corresponding six Bayesian estimators $\delta_{w2}^{\pi}(\boldsymbol{x})$, $\delta_{pl}^{\pi}(\boldsymbol{x})$, $\delta_s^{\pi}(\boldsymbol{x})$, $\delta_p^{\pi}(\boldsymbol{x})$, $\delta_2^{\pi}(\boldsymbol{x})$, and $\delta_z^{\pi}(\boldsymbol{x})$. Interestingly, for the six Bayesian estimators, we discover three strings of inequalities which are summarized in Theorem 1 (see Theorem 1 in [36]). To our surprise, an order between the two Bayesian estimators $\delta_{w2}^{\pi}(\boldsymbol{x})$ and $\delta_{pl}^{\pi}(\boldsymbol{x})$ on $\Theta = (0,1)$ does not exist. It is worthy to note that the three strings of inequalities only depend on the loss functions. Moreover, the inequalities are independent of the chosen models, and the used priors provided the Bayesian estimators exist, and thus they exist in a general setting which makes them quite interesting.

In this section, we compare the six Bayesian estimators $\delta_{w2}^{\pi}(\boldsymbol{x})$, $\delta_{pl}^{\pi}(\boldsymbol{x})$, $\delta_s^{\pi}(\boldsymbol{x})$, $\delta_p^{\pi}(\boldsymbol{x})$, $\delta_2^{\pi}(\boldsymbol{x})$, and $\delta_z^{\pi}(\boldsymbol{x})$. The domains of the loss functions, the six Bayesian estimators, the PELs, and the smallest PELs are summarized in **Table 2** (see **Table 1** in [36]). The six PELs are PEWSEL, PEPLL, PESL, PEPL, PESEL, and PEZL. In **Table 2**, each Bayesian estimator minimizes some corresponding PEL. Furthermore, the smallest PEL is the PEL evaluated at the corresponding Bayesian estimator.

It is easy to see that all the six loss functions are well defined on $\Theta = (0,1)$, and thus all the six Bayesian estimators are well defined on $\Theta = (0,1)$. There are only four loss functions defined on $\Theta = (0,\infty)$, since the power-log loss function and Zhang's loss function are only defined on $\Theta = (0,1)$. Hence, only four Bayesian estimators are well defined on $\Theta = (0,\infty)$. Moreover, only the weighted squared error loss function and the squared error loss function are defined on $\Theta = (-\infty,\infty)$, and therefore only two Bayesian estimators are well defined on $\Theta = (-\infty,\infty)$. Among the six Bayesian estimators, there exist three strings of inequalities which are summarized in the following theorem.

Theorem 1 (Theorem 1 in [36]). *Assume the prior satisfies some regularity conditions so that the posterior expectations involved in the definitions of the six Bayesian estimators exist. Then for $\Theta = (0,1)$, there exists a string of inequalities among the six Bayesian estimators*:

$$\max\left(\delta_{w2}^{\pi}(\boldsymbol{x}), \delta_{pl}^{\pi}(\boldsymbol{x})\right) \leq \delta_s^{\pi}(\boldsymbol{x}) \leq \delta_p^{\pi}(\boldsymbol{x}) \leq \delta_2^{\pi}(\boldsymbol{x}) \leq \delta_z^{\pi}(\boldsymbol{x}). \tag{49}$$

Moreover, for $\Theta = (0,\infty)$, there exists a string of inequalities among the four Bayesian estimators:

$$\delta_{w2}^{\pi}(\boldsymbol{x}) \leq \delta_s^{\pi}(\boldsymbol{x}) \leq \delta_p^{\pi}(\boldsymbol{x}) \leq \delta_2^{\pi}(\boldsymbol{x}). \tag{50}$$

Finally, for $\Theta = (-\infty,\infty)$, there exists an inequality between the two Bayesian estimators:

$$\delta_{w2}^{\pi}(\boldsymbol{x}) \leq \delta_2^{\pi}(\boldsymbol{x}). \tag{51}$$

The proof of Theorem 1 exploits a key, important, and unified tool, the covariance inequality (see Theorem 4.7.9 (p. 192) in [16]), and the proof can be found in the supplement of [36].

It is worthy to note that the six Bayesian estimators and the six smallest PELs are all functions of π, $\boldsymbol{x}$, and the loss function. Because there exists three strings of inequalities among the six Bayesian estimators, we would wonder whether there exists a string of inequalities among the six smallest PELs, in other words, $PEWSEL_{w2}(\pi,\boldsymbol{x})$, $PEPLL_{pl}(\pi,\boldsymbol{x})$, $PESL_s(\pi,\boldsymbol{x})$, $PEPL_p(\pi,\boldsymbol{x})$, $PESEL_2(\pi,\boldsymbol{x})$, and $PEZL_z(\pi,\boldsymbol{x})$. The answer to this question is no! The numerical simulations of the smallest PELs exemplify this fact (see [36]).

Domain	Bayesian estimators	PELs	Smallest PELs
$\Theta = (-\infty, \infty)$	$\delta_{w2}^\pi(\boldsymbol{x}) = \dfrac{E\left(\frac{1}{\theta}\middle|\boldsymbol{x}\right)}{E\left(\frac{1}{\theta^2}\middle|\boldsymbol{x}\right)}$	$\begin{aligned} PEWSEL(\pi, a|\boldsymbol{x}) &= E[L_{w2}(\theta, a)|\boldsymbol{x}] \\ &= E\left[\frac{1}{\theta^2}(\theta - a)^2\middle|\boldsymbol{x}\right] \end{aligned}$	$PEWSEL_{w2}(\pi, \boldsymbol{x}) = PEWSEL(\pi, a|\boldsymbol{x})\big|_{a=\delta_{w2}^\pi(\boldsymbol{x})}$
$\Theta = (0, 1)$	$\delta_{pl}^\pi(\boldsymbol{x})$ $= \dfrac{2 + E_1^{pl}(\boldsymbol{x}) - \sqrt{E_1^{pl}(\boldsymbol{x})\left(E_1^{pl}(\boldsymbol{x}) + 4\right)}}{2}$ with $E_1^{pl}(\boldsymbol{x}) = E\left[\frac{(1-\theta)^2}{\theta}\middle|\boldsymbol{x}\right] > 0$	$\begin{aligned} PEPLL(\pi, a|\boldsymbol{x}) &= E[L_{pl}(\theta, a)|\boldsymbol{x}] \\ &= E\left[\frac{\theta\left(\frac{1}{\theta}-1\right)^2}{1-a} - \log a + \log \theta - \frac{1}{\theta} + 1\middle|\boldsymbol{x}\right] \end{aligned}$	$PEPLL_{pl}(\pi, \boldsymbol{x}) = PEPLL(\pi, a|\boldsymbol{x})\big|_{a=\delta_{pl}^\pi(\boldsymbol{x})}$
$\Theta = (0, \infty)$	$\delta_s^\pi(\boldsymbol{x}) = \dfrac{1}{E\left(\frac{1}{\theta}\middle|\boldsymbol{x}\right)}$	$\begin{aligned} PESL(\pi, a|\boldsymbol{x}) &= E[L_s(\theta, a)|\boldsymbol{x}] \\ &= E\left[\frac{a}{\theta} - \log \frac{a}{\theta} - 1\middle|\boldsymbol{x}\right] \end{aligned}$	$PESL_s(\pi, \boldsymbol{x}) = PESL(\pi, a|\boldsymbol{x})\big|_{a=\delta_s^\pi(\boldsymbol{x})}$
$\Theta = (0, \infty)$	$\delta_p^\pi(\boldsymbol{x}) = \sqrt{\dfrac{E(\theta|\boldsymbol{x})}{E\left(\frac{1}{\theta}\middle|\boldsymbol{x}\right)}}$	$\begin{aligned} PEPL(\pi, a|\boldsymbol{x}) &= E[L_p(\theta, a)|\boldsymbol{x}] \\ &= E\left[\frac{a}{\theta} + \frac{\theta}{a} - 2\middle|\boldsymbol{x}\right] \end{aligned}$	$PEPL_p(\pi, \boldsymbol{x}) = PEPL(\pi, a|\boldsymbol{x})\big|_{a=\delta_p^\pi(\boldsymbol{x})}$
$\Theta = (-\infty, \infty)$	$\delta_2^\pi(\boldsymbol{x}) = E(\theta|\boldsymbol{x})$	$\begin{aligned} PESEL(\pi, a|\boldsymbol{x}) &= E[L_2(\theta, a)|\boldsymbol{x}] \\ &= E\left[(\theta - a)^2\middle|\boldsymbol{x}\right] \end{aligned}$	$PESEL_2(\pi, \boldsymbol{x}) = PESEL(\pi, a|\boldsymbol{x})\big|_{a=\delta_2^\pi(\boldsymbol{x})}$
$\Theta = (0, 1)$	$\delta_z^\pi(\boldsymbol{x}) = \dfrac{\sqrt{E_1^{z}(\boldsymbol{x})}}{\sqrt{E_1^{z}(\boldsymbol{x})} + \sqrt{E_2^{z}(\boldsymbol{x})}}$ with $E_1^{z}(\boldsymbol{x}) = E\left[\frac{\theta^3}{(1-\theta)^2}\middle|\boldsymbol{x}\right]$ and $E_2^{z}(\boldsymbol{x}) = E[\theta|\boldsymbol{x}]$	$\begin{aligned} PEZL(\pi, a|\boldsymbol{x}) &= E[L_z(\theta, a)|\boldsymbol{x}] \\ &= E\left[\frac{\theta}{\left(\frac{1}{\theta}-1\right)^2 a} + \frac{\theta}{1-a} - \frac{1}{\theta\left(\frac{1}{\theta}-1\right)^2}\middle|\boldsymbol{x}\right] \end{aligned}$	$PEZL_z(\pi, \boldsymbol{x}) = PEZL(\pi, a|\boldsymbol{x})\big|_{a=\delta_z^\pi(\boldsymbol{x})}$

Table 2.
(Table 1 in [36]) The six Bayesian estimators, the PELs, and the smallest PELs.

6. Conclusions and discussions

In this chapter, we have investigated six loss functions: the squared error loss function, the weighted squared error loss function, Stein's loss function, the power-power loss function, the power-log loss function, and Zhang's loss function. Now we give some suggestions on the conditions for using each of the six loss functions. It is worthy to note that among the six loss functions, the first two loss functions are defined on $\Theta = (-\infty, \infty)$ and they penalize overestimation and underestimation equally on $(-\infty, \infty)$, and thus we recommend to use them when the parameter space is $(-\infty, \infty)$. Moreover, the middle two loss functions are defined on $\Theta = (0, \infty)$, and they penalize gross overestimation and gross underestimation equally on $(0, \infty)$, and thus we recommend to use them when the parameter space is $(0, \infty)$. In particular, if one prefers the loss function to have balanced convergence rates or penalties for its argument too large and too small, then we recommend to use the power-power loss function on $(0, \infty)$. Furthermore, the last two loss functions are defined on $\Theta = (0, 1)$, and they penalize gross overestimation and gross underestimation equally on $(0, 1)$, and thus we recommend to use them when the parameter space is $(0, 1)$. In particular, if one prefers the loss function to have balanced convergence rates or penalties for its argument too large and too small, then we recommend to use Zhang's loss function on $(0, 1)$.

For each one of the six loss functions, we can find a corresponding Bayesian estimator, which minimizes the corresponding posterior expected loss. Among the six Bayesian estimators, there exist three strings of inequalities summarized in Theorem 1 (see also Theorem 1 in [36]). However, a string of inequalities among the six smallest PELs does not exist.

We summarize three hierarchical models where the unknown parameter of interest is $\theta \in \Theta = (0, \infty)$, that is, the hierarchical normal and inverse gamma model (9), the hierarchical Poisson and gamma model (10), and the hierarchical normal and normal-inverse-gamma model (11). In addition, we summarize two hierarchical models where the unknown parameter of interest is $\theta \in \Theta = (0, 1)$, that is, the beta-binomial model (30) and the beta-negative binomial model (31).

Now we give some suggestions on the selection of the hyperparameters. One way to select the hyperparameters is through the empirical Bayesian analysis, which relies on a conjugate prior modeling, where the hyperparameters are estimated from the observations and the "estimated prior" is then used as a regular prior in the later inference. The marginal distribution can then be used to recover the prior distribution from the observations. For empirical Bayesian analysis, two common methods are used to obtain the estimators of the hyperparameters, that is, the moment method and the MLE method. Numerical simulations show that the MLEs are better than the moment estimators when estimating the hyperparameters in terms of the goodness of fit of the model to the simulated data. However, the MLEs are very sensitive to the initial estimators, and the moment estimators are usually proved to be good initial estimators.

Conflict of interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author details

Ying-Ying Zhang
Department of Statistics and Actuarial Science, College of Mathematics and Statistics, Chongqing University, Chongqing, China

*Address all correspondence to: robertzhangyying@qq.com; robertzhang@cqu.edu.cn

IntechOpen

References

[1] Robert CP. The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation. 2nd paperback ed. New York: Springer; 2007

[2] James W, Stein C. Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. 1961. pp. 361-380

[3] Brown LD. Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. The Annals of Mathematical Statistics. 1968;**39**:29-48

[4] Brown LD. Comment on the paper by maatta and casella. Statistical Science. 1990;**5**:103-106

[5] Parsian A, Nematollahi N. Estimation of scale parameter under entropy loss function. Journal of Statistical Planning and Inference. 1996;**52**:77-91

[6] Petropoulos C, Kourouklis S. Estimation of a scale parameter in mixture models with unknown location. Journal of Statistical Planning and Inference. 2005;**128**:191-218

[7] Oono Y, Shinozaki N. On a class of improved estimators of variance and estimation under order restriction. Journal of Statistical Planning and Inference. 2006;**136**:2584-2605

[8] Ye RD, Wang SG. Improved estimation of the covariance matrix under stein's loss. Statistics & Probability Letters. 2009;**79**:715-721

[9] Bobotas P, Kourouklis S. On the estimation of a normal precision and a normal variance ratio. Statistical Methodology. 2010;**7**:445-463

[10] Zhang YY. The bayes rule of the variance parameter of the hierarchical normal and inverse gamma model under stein's loss. Communications in Statistics-Theory and Methods. 2017;**46**: 7125-7133

[11] Zhang YY. The bayes rule of the positive restricted parameter under the power-power loss with an application. Communications in Statistics-Theory and Methods. 2019; Under review

[12] Zhang YY, Zhou MQ, Xie YH, Song WH. The bayes rule of the parameter in (0, 1) under the power-log loss function with an application to the beta-binomial model. Journal of Statistical Computation and Simulation. 2017;**87**:2724-2737

[13] Zhang YY, Xie YH, Song WH, Zhou MQ. The bayes rule of the parameter in (0, 1) under zhang's loss function with an application to the beta-binomial model. Communications in Statistics-Theory and Methods. 2019. DOI: 10.1080/03610926.2019.1565840

[14] Stein C. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. Annals of the Institute of Statistical Mathematics. 1964;**16**:155-160

[15] Maatta JM, Casella G. Developments in decision-theoretic variance estimation. Statistical Science. 1990;**5**: 90-120

[16] Casella G, Berger RL. Statistical Inference. 2nd ed. USA: Duxbury; 2002

[17] Lehmann EL, Casella G. Theory of Point Estimation. 2nd ed. New York: Springer; 1998

[18] Raiffa H, Schlaifer R. Applied Statistical Decision Theory. Cambridge: Harvard University Press; 1961

[19] Deely JJ, Lindley DV. Bayes empirical bayes. Journal of the

American Statistical Association. 1981;
76:833-841

[20] Zhang YY, Wang ZY, Duan ZM,
Mi W. The empirical bayes estimators of
the parameter of the poisson
distribution with a conjugate gamma
prior under stein's loss function. Journal
of Statistical Computation and
Simulation. 2019. DOI: 10.1080/
00949655.2019.1652606

[21] Mao SS, Tang YC. Bayesian
Statistics. 2nd ed. Beijing: China
Statistics Press; 2012

[22] Chen MH. Bayesian statistics
lecture. Statistics Graduate Summer
School. China: School of Mathematics
and Statistics; Northeast Normal
University: Changchun; 2014

[23] Xie YH, Song WH, Zhou MQ,
Zhang YY. The bayes posterior
estimator of the variance parameter
of the normal distribution with a
normal-inverse gamma prior under
stein's loss. Chinese Journal of Applied
Probability and Statistics. 2018;**34**:
551-564

[24] Dey D, Srinivasan C. Estimation of
a covariance matrix under stein's loss.
The Annals of Statistics. 1985;**13**:
1581-1591

[25] Sheena Y, Takemura A.
Inadmissibility of non-order-preserving
orthogonally invariant estimators of the
covariance matrix in the case of stein's
loss. Journal of Multivariate Analysis.
1992;**41**:117-131

[26] Konno Y. Estimation of a normal
covariance matrix with incomplete data
under stein's loss. Journal of
Multivariate Analysis. 1995;**52**:308-324

[27] Konno Y. Estimation of normal
covariance matrices parametrized by
irreducible symmetric cones under
stein's loss. Journal of Multivariate
Analysis. 2007;**98**:295-316

[28] Sun XQ, Sun DC, He ZQ. Bayesian
inference on multivariate normal
covariance and precision matrices in a
star-shaped model with missing data.
Communications in Statistics-Theory
and Methods. 2010;**39**:642-666

[29] Ma TF, Jia LJ, Su YS. A new
estimator of covariance matrix. Journal
of Statistical Planning and Inference.
2012;**142**:529-536

[30] Xu K, He DJ. Further results on
estimation of covariance matrix.
Statistics & Probability Letters. 2015;
101:11-20

[31] Tsukuma H. Estimation of a high-
dimensional covariance matrix with the
stein loss. Journal of Multivariate
Analysis. 2016;**148**:1-17

[32] Singh SK, Singh U, Sharma VK.
Expected total test time and Bayesian
estimation for generalized lindley
distribution under progressively type-ii
censored sample where removals follow
the beta-binomial probability law.
Applied Mathematics and Computation.
2013;**222**:402-419

[33] Luo R, Paul S. Estimation for zero-
inflated beta-binomial regression model
with missing response data. Statistics in
Medicine. 2018;**37**:3789-3813

[34] Zhou MQ, Zhang YY, Sun Y, Sun J.
The empirical bayes estimators of the
probability parameter of the beta-
negative binomial model under Zhang's
loss function. Computational Statistics
and Data Analysis. 2019; Under review

[35] Danaher PJ. A markov mixture
model for magazine exposure. Journal of
the American Statistical Association.
1989;**84**:922-926

[36] Zhang YY, Xie YH, Song WH,
Zhou MQ. Three strings of inequalities
among six bayes estimators.
Communications in Statistics-Theory
and Methods. 2018;**47**:1953-1961