ASR Vol. 30. No. 1 March 2016

SYMMETRIC AND ASYMMETRIC LOSS FUNCTIONS OF POISSON MEAN- CONSENSUS ON JEFFREYS PRIOR UNDER BAYESIAN PARADIGM

Renuka Devi R[1], Subbiah M[2], Srinivasan MR[3]

[1]Department of Statistics, Presidency College, Chennai, India.
renuka2327@yahoo.com

[2]Department of Mathematics, L. N. Govt. College Ponneri, Tamilnadu, India.
sisufive@gmail.com

[3]Department of Statistics, University of Madras, Chennai, India.
mrsvasan8@hotmail.com

Abstract

Most of the Bayesian estimates depend on the posterior mean under squared loss function. However, other loss functions can also be considered to have competing estimators especially when symmetricity cannot be presumed. This work considers Poisson-Gamma model for estimating mean parameter using four loss functions that reflect asymmetry. Explicit form of the estimators that minimize posterior expected loss and their relations together with the conditional Bayes risk are derived. The study has highlighted Jeffreys prior as consensus objective prior using Bayesian conditional risk based on both symmetric and asymmetric loss functions. Appropriate choice of point estimator can be made based on more relevant practical constraints and the associated trade-off between the bias and the variance.

Key Words: Poisson mean, Gamma distribution, Jeffreys prior, Loss functions, MLE.

1. Introduction

Comparison of competing procedures involving statistical inference is one of the interesting research problems to identify the potential choice. In this context, discrete distributions have drawn much attention (Brown et.al. 2001, Agresti and Min, 2005, Campbell 2007, Tuyl et.al. 2008, Sweeting et.al. 2004, Subbiah et.al. 2008, Patil and Kulkarni, 2012), but evaluation and comparison of point estimates have posed interesting research challenges. Evaluation criteria such as mean squared error or loss function optimality are the major approaches; in Bayesian point of view, (1) Bayes risk, (2) risk function as the conditional expectation of an estimator given parameter and (3) posterior expected loss have been discussed (Robert 1994, Lehmann and Casella 1998, Casella and Berger 2002).

Poisson distribution is basically to model rare events and / or count data, and the present work primarily considers rare event cases with many zero occurrences resulting in a low mean value. As a consequence, the available unbiased minimum variance estimator for Poisson mean is not achievable under frequentist approach and thereby looking for an alternative approach of studying the mean. Bayes risk related applications can be found in literature such as Poisson sequential in Hwang and Lee (2013), Lee and Hwang (2011), minimax estimator when the parameter is restricted in Wan et al (2000) and the problem of estimation of a continuous distribution function in Ning and Xie (2007). Most of the existing works deal with one of the asymmetric loss functions like LINEX (Casella and Berger 2002) and an investigation with other loss functions provide a scope for a detailed study.

In this direction, the present study sets the following objectives:

1. Identify different loss function and derive explicit forms for corresponding Bayesian estimators.
2. Understand possible relationships between the estimators that influence the choices of certain constants.

3. Study the risk function as the conditional expectation of an estimator based on the loss functions listed in (1) so as to elaborate the role of Jeffreys prior in Bayesian analysis.

4. Implement and validate using a simulation study through relevant computing algorithms for possible evaluation of different loss functions.

The paper is organized as: Section 2 elaborates the Bayesian estimators obtained through minimizing posterior expected loss under four d loss functions and possible relations between them; details of conditional risk function are provided in Section 3 followed by data analyses in Section 4 and Section 5 provides conclusive remarks.

2. Bayesian Estimators

If $X_1, X_2, \ldots\ldots X_n$ are independent and identically distributed as Poisson (θ) and if $Y = \sum X_i$ then Y is distributed as Poisson ($n\theta$)and likelihood of θ/y is

$$L(\theta / Y) = f(y/\theta) = \begin{cases} e^{-n\theta}\dfrac{(n\theta)^y}{y!} & y = 0,1,2\ldots\ldots \text{ and } \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The conjugate prior density for θ is a Gamma distribution with shape parameter α and scale parameter β.

$$P(\theta) = \begin{cases} \dfrac{1}{\Gamma\alpha\beta^{\alpha}}\theta^{\alpha-1}e^{-\frac{\theta}{\beta}} & \theta > 0;\ \alpha, \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Hence, the Posterior distribution of θ/y is gamma distribution with $y + \alpha$ and $\dfrac{\beta}{1+\beta n}$ as shape and scale parameters respectively. Further, two algebraic results can easily be derived using gamma function.

i. $$\frac{1}{\Gamma a b^a}\int_0^\infty e^{tz} z^{a-1} e^{-z} dz = \left[\frac{1}{1-bt}\right]^a t < \frac{1}{b}$$

ii. $$\frac{1}{\Gamma a b^a}\int_0^\infty e^{r} z^{a-1} e^{-\frac{z}{b}} dz = \frac{b^r}{\Gamma a}\Gamma(r+a)$$

Now, some standard loss functions are listed to obtain respective estimators based on minimizing the posterior expected loss.

a. Squared Error Loss Function (SELF)

$L(\theta, \delta) = (\theta, \delta)^2$ and the respective estimator $\delta_1(x)$ is the mean of the posterior distribution.

b. Scaled Squared Error Loss Function (SSELF)

$$L(\theta, \delta) = \frac{(\theta-\delta)^2}{\theta^k}$$

$$E[L(\theta,\delta)] = E\left[\frac{\theta^2+\delta^2-2\theta\delta}{\theta^k}\right]$$

$$= E[\theta^{2-k} + \delta^2\theta^{-k} - 2\theta^{1-k}\delta]$$

$$= \delta^2 E[\theta^{-k}] - 2\delta E[\theta^{1-k}] + E[\theta^{2-k}],$$

Differentiating with respect to δ and equating to zero, yields

$$2\delta\, E[\theta^{-k}] - 2E[\theta^{1-k}] = 0 \text{ and hence } \delta_2(x) = \frac{E[\theta^{1-k}]}{E[\theta^{-k}]}$$

c. LINEX Error Loss Function (LELF)

$L(\theta, \delta) = ec^{(\delta-\theta)} - c(\delta - \theta) - 1$ (c > 0)

$$E[L(\theta,\delta)] = E[e^{c\delta - c\theta} - c(\delta-\theta) - 1]$$

$$= e^{c\delta} E[e^{-c\theta}] - c\delta + cE(\theta) - 1$$

Differentiating with respect to δ and equating to zero, yields

$$e^{c\delta} cE[e^{-c\theta}] - c = 0 \text{ and hence } \delta_3(x) = -\frac{1}{c} \text{ In } E[e^{-c\theta}]$$

d. General Entropy Loss Function (GELF)

$$L(\theta,\delta) = w\left[\left(\frac{\delta}{\theta}\right)^p - p\ In\left(\frac{\delta}{\theta}\right) - 1\right] \qquad p > 0,\ w \neq 0$$

$$E[L(\theta,\delta)] = w\left[E\left(\frac{\delta}{\theta}\right)^p - E[p(In\delta - In\theta)] - 1\right]$$

$$= w[\delta^p E(\theta^{-p}) - p\, In\delta + pE(In\theta) - 1]$$

Differentiating with respect to δ and equating to zero,

$$w\left[p\delta^{p-1}E(\theta^{-p}) - p\frac{1}{\delta}\right] = 0 \text{ and hence } \delta_4(x) = \left[\frac{1}{E(\theta^{-p})}\right]^{\frac{1}{p}}$$

Besides the above, there are two more widely used loss functions namely, Stein's Error Loss Function and Quadratic Error Loss Function which are respectively GELF with $w = p = 1$ and SSELF with $k = 2$. Under Poisson-Gamma model, following table presents the explicit form of the Bayesian estimators deduced under different risk functions.

Loss Function	Estimator
SELF	$\delta_1 = E(\theta / X) = \dfrac{\beta}{1+\beta n}(y+\alpha)$
SSELF	$\delta_2 = \left[\dfrac{E(\theta^{1-k})}{E(\theta^{-k})}\right] = \dfrac{\beta}{1+\beta n}(y+\alpha-k) \quad 0<k<\alpha$
LELF	$\delta_3 = -\dfrac{1}{c} In\, E[e^{-c\theta}] = \dfrac{In\left(1+\dfrac{\beta}{1+\beta n}c\right)}{c}(y+\alpha)$
GELF	$\delta_4 = \left[\dfrac{1}{E(\theta^{-p})}\right]^{\frac{1}{p}} = \left[\dfrac{\beta}{1+\beta n}\right]\left[\dfrac{\Gamma(y+\alpha)}{\Gamma(y+\alpha-p)}\right]^{\frac{1}{p}} \quad 0<p<\alpha$

All the expectation evaluations are based on the posterior Gamma distribution (shape: y + α, scale: $\frac{\beta}{1+\beta n}$) and applying the results (i) or (ii) stated earlier. Further, it could be observed from the above display that there is a need to specify constants other than prior scale and shape parameters in these estimators except δ_1. But, the algebraic functions of δ_1, δ_2, δ_3 and δ_4 provide some interesting relationships between them so that the impact of such constants involved in δ_2, δ_3 and δ_4 can be understood especially with respect to prior scale β.

Let $F_1 = \frac{\beta}{1+\beta n}$ so that F1 always lies in (0, 1) based on $\beta \to 0$ or $\beta \to \infty$ and for a fixed n; accordingly for a sufficiently small c in δ_3 $F_1 c \to 0$ whence $\dfrac{In(1+F_1 c)}{F_1 c} \to 1$

Also $\delta_1 = F_1(y+\alpha)$ and $\delta_3 = \dfrac{In\left(1+\dfrac{\beta}{\beta n+1}c\right)}{c}(y+\alpha) = \delta_1 \dfrac{In(1+F_1 c)}{cF_1}$ and equivalently $\delta_3 \cong \delta_1$ whenever c is sufficiently small irrespective of β. On the other hand, if c is large then $F_1 c$ is large (irrespective of β) and hence δ_1 and δ_3 could differ markedly. So, the variation in c of δ_3 allows thinking δ_1 as a limiting case of δ_3 and the choice of c is independent of data and prior parameters under Poisson-Gamma model.

In a similar interest, with respect to k in SSELF

$$\delta_2 = \left(\frac{\beta}{n\beta+1}\right)(y+\alpha-k)$$

$$= \left(\frac{\beta}{n\beta+1}\right)(y+\alpha) - \left(\frac{\beta}{n\beta+1}\right)k$$

$$= \delta_1 - \left(\frac{\beta}{n\beta+1}\right)k$$

This result indicates the role of k and β beyond the original dependency of k on the shape parameter (α) of prior distribution ($k < \alpha$); particularly for sufficiently large β, $\delta_2 \cong \delta_1$ for appropriate choices of n.

Similarly in the case of GELF with a definite relation between p and α $(0 < p < \alpha)$, a limiting case of $p \to \alpha$ yields $\frac{\Gamma(y+\alpha)}{\Gamma(y+\alpha-p)} \simeq \frac{\Gamma(y+\alpha)}{\Gamma y} \leq y^{\alpha}$ (Qi and Luo, 2013). In particular if $\alpha = 0.5$ as in the case of a Jeffreys prior for Poisson-Gamma model, then $\delta_4 < \delta_1$ because, $\delta_4 = F_1\left[\frac{\Gamma(y+0.5)}{\Gamma y}\right]^2 \leq F_1 y < F_1(y+\alpha)$ for $\alpha > 0$; it can also be noted that if β is sufficiently large, then $\delta_4 \leq F_1 y = \frac{y}{n}$, MLE of the Poisson mean parameter θ. However, if p is close to zero δ_4 is very much independent of y which may not be so desirable.

The above discussion provides a methodology for choosing the specific constants that appear in δ_2, δ_3 and δ_4. Now, the interest is on the choice of prior parameters one of the main features of a Bayesian analysis. Listed as the third objective in the previous section, conditional risk function has been considered as a tool to clearly recognize the norms for selecting the two parameters α and β with an attention to non-informative priors.

3. Bayesian Conditional Risk Functions

In this Section, risk Function $R(\theta, \delta)$ as the conditional expectation $E[L(\theta, \delta)| \theta]$ has been derived for each of the above four loss functions.

Under SELF, $L[\theta, \delta] = (\theta - \delta)^2$ so that $R(\theta, \delta) = E[L(\theta, \delta)] = E[(\theta, \delta)^2]$

$R(\theta, \delta) = E[\theta^2 + \delta^2 - 2\theta\delta]$

$$= \theta^2 + E\left[\left(\frac{\beta}{\beta n + 1}\right)^2 (y+\alpha)^2\right] - 2\theta E\left[\left(\frac{\beta}{n\beta + 1}\right)(y+\alpha)\right]$$

$$= \theta^2 + \left(\frac{\beta}{n\beta+1}\right)^2 E\left[y^2 + \alpha^2 + 2\alpha y\right] - 2\theta E\left[\left(\frac{\beta}{n\beta + 1}\right)(y+\alpha)\right]$$

$$= \theta^2 + \left(\frac{\beta}{n\beta+1}\right)^2 \left[n^2\theta^2 + n\theta + \alpha^2 + 2\alpha\theta n\right] - 2\theta\frac{\beta}{n\beta+1}(n\theta+\alpha)$$

$$= \theta^2\left[1 + \left(\frac{n\beta}{n\beta+1}\right)^2 - 2\frac{n\beta}{n\beta+1}\right] + \theta\left(\frac{\beta}{n\beta+1}\right)^2\left[n + 2\alpha\left(n - \frac{n\beta+1}{\beta}\right)\right] + \left(\frac{\alpha\beta}{1+n\beta}\right)^2$$

$$R(\theta,\delta) = \theta^2\left(\frac{1}{n\beta+1}\right)^2 + \theta\left(\frac{\beta}{n\beta+1}\right)^2\left[n - \frac{2\alpha}{\beta}\right] + \left(\frac{\alpha\beta}{1+n\beta}\right)^2,$$ **a quadratic function of θ.**

Under SSELF, $L[\theta,\delta] = \dfrac{(\theta-\delta)^2}{\theta^k}$

So, $R(\theta, \delta) = E\,[L(\theta, \delta(x)]$

$$= E\left[\frac{(\theta-\delta)^2}{\theta^k}\right] = \frac{1}{\theta^k}E\left[\theta^2+\delta^2-2\theta\delta\right]$$

$$=\frac{1}{\theta^k}E\left[\theta^2+\frac{\beta^2}{(n\beta+1)^2}(y+\alpha-k)^2-2\theta\frac{\beta}{n\beta}(y+\alpha-k)\right]$$

$$=\theta^{2-k}-2\theta^{1-k}\frac{\beta}{n\beta+1}E[y+\alpha-k]+\frac{\beta^2\theta^{-k}}{(n\beta+1)^2}E\left[(y+\alpha-k)^2\right]$$

$$=\theta^{2-k}-2\theta^{1-k}\frac{\beta}{n\beta+1}(n\theta+\alpha-k)+\frac{\beta^2\theta^{-k}}{(\beta+1)^2}\left[E(y^2)+(\alpha-k)^2+2(\alpha-k)E(y)\right]$$

$$=\theta^{2-k}-2\theta^{1-k}\frac{\beta}{n\beta+1}(n\theta+\alpha-k)+\frac{\beta^2\theta^{-k}}{(n\beta+1)^2}\left[n^2\theta^2+n\theta+(\alpha-k)^2+2(\alpha-k)n\theta\right]$$

$$=\theta^{2-k}-2\theta^{1-k}\frac{\beta}{n\beta+1}(n\theta+\alpha-k)+\frac{n^2\beta^2}{(n\beta+1)^2}\theta^{2-k}+n\frac{\beta^2}{(\beta+1)^2}\theta^{1-k}+$$

$$n\frac{\beta^2 2(\alpha-k)}{(\beta+1)^2}\theta^{1-k}+(\alpha-k)^2\frac{\beta^2}{(n\beta+1)^2}\theta^{-k}$$

$$\theta^{2-k}\left(1-\frac{n\beta}{n\beta+1}\right)^2-\theta^{1-k}\left[n\frac{\beta^2}{(\beta+1)^2}-2(\alpha-k)\frac{\beta}{n\beta+1}2n(\alpha-k)\left(\frac{\beta}{\beta+1}\right)^2\right]$$

$$+\theta^{-k}\left(\frac{(\alpha-k)\beta}{n\beta+1}\right)^2$$

$$= \theta^{2-k}\left(\frac{1}{n\beta+1}\right)^2 + \theta^{1-k}\left[n\frac{\beta^2}{(n\beta+1)^2} - 2(\alpha-k)\frac{\beta}{n\beta+1}\left(1-\frac{n\beta}{n\beta+1}\right)\right]$$

$$+\theta^{-k}\left[\frac{(\alpha-k)\beta}{n\beta+1}\right]^2$$

$$R(\theta,\delta) = \theta^{2-k}\left(\frac{1}{n\beta+1}\right)^2 + \theta^{1-k}\frac{\beta}{n\beta+1}\left[\frac{n\beta}{n\beta+1} - \frac{2(\alpha-k)}{n\beta+1}\right]$$

$$+\theta^{-k}\left[\frac{(\alpha-k)\beta}{n\beta+1}\right]^2 \qquad 0<k<\alpha$$

Under LINEX, $L(\theta, \delta) = e^{c(\delta-\theta)} - c(\delta-\theta) - 1$ and hence $R(\theta, \delta)$ is

$E[L(\theta, \delta)] = E[e^{c(\delta-\theta)} - c(\delta-\theta) - 1]$

$= e^{c\theta}E[e^{c\delta}] - cE(\delta) + c\theta - 1$

Now, $c\delta = \ln\left(1+\frac{\beta}{n\beta+1}c\right)(y+\alpha)$ so that $e^{c\delta} = \left(1+\frac{\beta}{n\beta+1}c\right)^{(y+\alpha)} = z^{(y+\alpha)}$ where z is free from y

$E[e^{c\delta}] = E[z^{(y+\alpha)}]$

$= z^{\alpha}E[z^{y}] = z^{\alpha}\sum z^{y}p(Y=y)$ where $Y \sim$ Poisson $(n\theta)$ and hence

$$E\left[e^{c\delta}\right] = z^{\alpha}e^{-n\theta(1-z)} = z^{\alpha}e^{-n\theta(1-z)} = z^{\alpha}e^{\frac{n\beta c}{1+n\beta}\theta}$$

Also, $E[\delta] = \dfrac{\left(1+\frac{\beta}{\beta n+1}c\right)}{c}(n\theta+\alpha)$

$$\therefore R(\theta,\delta)=e^{-c\theta}z^{\alpha}e^{\frac{n\beta c}{1+n\beta}\theta}-c\frac{In\left(1+\frac{\beta c}{n\beta+1}\right)}{c}(n\theta+\alpha)+c\theta-1$$

$$=\left(1+\frac{\beta c}{n\beta+1}\right)^{\alpha}e^{-c\left(1-\frac{n\beta}{\beta n+1}\right)\theta}-In\left(1+\frac{\beta c}{\beta n+1}\right)(n\theta+\alpha)+c\theta-1$$

$$R(\theta,\delta)=\left[c-In\left(1+\frac{\beta c}{\beta n+1}\right)n\right]\theta+\left(1+\frac{\beta c}{n\beta+1}\right)^{\alpha}e^{-c\left(\frac{1}{\beta n+1}\right)\theta}$$

$$-\alpha\, In\left(1+\frac{\beta c}{\beta n+1}\right)-1c>0$$

If c is close to 0, then LINEX will be closer to SELF hence to study the effect of c it is retained as $c > 1$.

Under GELF, $L(\theta,\delta)=w\left[\left(\frac{\delta}{\theta}\right)^{p}-p\,In\left(\frac{\delta}{\theta}\right)-1\right]$, without loss of generality, assume $w =$ 1 so that $L(\theta,\delta)=\left(\frac{\delta}{\theta}\right)^{p}-p\,In\left(\frac{\delta}{\theta}\right)-1=\left(\frac{\delta}{\theta}\right)^{p}-p\,In\;\delta+p\,In\theta-1$

So $R(\theta,\delta)=E[L(\theta,\delta)]=E\left[\left(\frac{\delta}{\theta}\right)^{p}-p\,In\delta+p\,In\theta-1\right]$

$$R(\theta,\delta)=\frac{1}{\theta^{p}}E\left[\delta^{p}\right]-p\,E[In\,\delta]+p\,In\,\theta-1$$

Now, $\delta^{p}=\left(\frac{\beta}{\beta n+1}\right)+\frac{1}{p}\,In\frac{\Gamma(y+\alpha)}{\Gamma(y+\alpha-p)}0<p<\alpha$

$\ln\ \delta = In\left(\frac{\beta}{\beta n+1}\right)+\frac{1}{p}In\frac{\Gamma(y+\alpha)}{\Gamma(y+\alpha-p)}0<p<\alpha$. Computation of expectations for δ^p and lnδ is quite involved and an approximation procedure may be considered and this expression can further be discussed as limiting case for p.

The behaviour of conditional expectation of four loss functions can be studied further as a function of θ and specific choice of other quantities involved. All the forms have revealed the impact of prior scale parameter (β) and hence it becomes the foremost choice to discuss followed by the choice of other constants (k, c, and p). That is if β is sufficiently large such that $\frac{1}{\beta n+1}$ and $\frac{\beta}{\beta n+1}$ tend to zero and $\frac{1}{n}$ respectively then $R(\beta, \delta)$ under SELF is linear in θ, whereas the choice of α and $k > 0$ in $(\alpha - \varepsilon, \alpha)$ will decide the shape of $R(\theta, \delta)$ under SSELF as increasing or decreasing function of θ accordingly $\alpha < 1$ or $\alpha > 1$. If β is small, $R(\theta, \delta)$ is a quadratic function of θ for SELF and SSELF has similar property except for the limits for α greater or less than two.

In the case of LINEX, $R(\theta, \delta)$ is linear in θ when β is large, with the coefficient θ is $n[c/n - ln\ (1 + c/n)] = n\{c/n - [c/n - c^2/2n^2 + c^3/3n^3 -\ldots]\} = c^2/n[1/2 - c/3n +\ldots]$ which is of order $1/n$ for fixed c and similar pattern is visible for smaller β also. In the absence of a closed form for $R(\theta, \delta)$ under GELF, it can be observed that if p is sufficiently closer to α, then $\frac{\Gamma(y+\alpha)}{\Gamma(y+\alpha-p)}\to\frac{\Gamma(y+\alpha)}{\Gamma y}\to y^{\alpha}$. Also for large β, $\delta^p\to\frac{y^{\alpha}}{n^p}$ and $In\,\delta\to In\left(\frac{1}{n}\right)+\frac{1}{p}In\,y^{\alpha}=\frac{\alpha}{p}In\left(\frac{y}{n}\right)$. Further, first order Taylor's approximation for $E(\delta^p)$ and $E(In\delta)$ will be $\frac{\theta^{\alpha}}{n^p}$ and $\frac{\alpha}{p}In\left(\frac{\theta}{n}\right)$ and hence $R(\theta, \delta)$ will have a form of

$$\frac{\theta^{\alpha-p}}{n^p}-\alpha\,In\left(\frac{\theta}{n}\right)+pIn\theta-1=\frac{\theta^{\alpha-p}}{n^p}-(\alpha-p)In\theta+\alpha In(n)-1.$$

This will make $R(\theta, \delta)$ free from θ for p closer to α. However, small value of β will increase $R(\theta, \delta)$ irrespective of all other quantities which may not be desirable.

In summary, prior distribution of Poisson-Gamma model could be a smaller shape (α) and larger scale (β). This observation using conditional risk function draws attention to the notion of Jeffreys prior which is $\alpha = 0.5$ and $\beta \rightarrow \infty$ in the case of Poisson-Gamma model. A visible further interest would be to understand the impact of β under different parametric combinations (n and θ) together with different choices of constants in δ_2, δ_3 and δ_4; a detailed simulation study has been carried out towards this purpose.

4. Simulation Study

The analyses include a detailed simulation plan based on different choices of parametric value namely Poisson mean (θ) and the sample size (n). six values 0.10, 0.25, 0.50, 0.75, 1.00, and 2.00 for θ are considered; choice of n includes 5, 10, 20, 50, and 100; Prior Gamma parameters: Shape (α) = 0.5 and Scale (β) from three options (0, 5), (5, 10) and (10, ∞); other chosen constants are $c = 3$ in LINEX (to reflect the possible difference from SELF), p in GELF and k in SSELF are less than but close to $\alpha = 0.5$.

For each combination, 5000 Monte Carlo samples of size n are generated from Poisson (θ) to obtain the Bayesian estimates. To summarize mean of $(\delta - \theta)^2$ is calculated for the four Bayesian estimators and the results are presented in Table 1. This will help to understand the sensitivity of an estimator for the changes in θ, β and Table 2 provides similar summary to study the impact of three more method-specific constants (c, k and p). For comparative purpose results from maximum likelihood estimator is also presented in Table 1.

It can be observed from that for $n = 5$ LELF is the least for all θ and β with an exception when $\theta = 2$ and $1 < \beta < 10$ SELF is minimum; GELF is less than SSELF for smaller values of θ and across all ranges of β and tend to be closer for larger values of

θ and β MLE is the maximum for all values of θ and β except when $\beta > 10$ where SELF provides the maximum and all measures get closer with an increase of θ and $\beta > 10$. Similar observations can be made when $n = 10, 20, 50$ over three ranges of β and all values of θ. However, starting from $n = 20$, GELF and SSELF tend to have same values whenever $10 < \beta$ and the comparisons become less important with an increasing values of n, θ and β. This difference fades much more visibly when n becomes more than 50 and for any value of θ and β.

Further, influence of changes in prior scale parameter (β) is less for increasing $n(n > 20)$ across the choice of θ; in the case of $n = 5$ and 10 LELF shows less sensitivity over the values of θ strangely with an exception of $\theta = 2$. All other methods show an increasing pattern with respect to increase in β in these two cases. Irrespective of n, all estimators show a slow but increasing trend in the mean squared difference with increasing values of θ.

From these observations, β in (1, 5) has been considered for further comparisons between LELF, SSELF and GELF for the impact of c and p. It can be observed from Table 2 that LELF is quite sensitive for the changes in c with respect to θ; estimate shows less mean squared difference when θ and c are inversely related irrespective of changes in n. However, for larger n ($n > 20$) impact of p and k (closer to α) is minimal for SSELF and GELF; when n is small these estimates tend to show increasing difference when p is getting closer to α. The results obtained by simulation are validated through an illustrative data discussed in the next section.

Table 1: Simulated results of mean value of the squared difference between the estimates for Poisson(θ) and θ over five different values of n, six values of θ and three combinations of prior scale parameter (θ)

Sample Size	θ	1 < β < 5				5 < β < 10				10 < β				MLE
		SELF	SSELF	LELF	GELF	SELF	SSELF	LELF	GELF	SELF	SSELF	LELF	GELF	
5	0.10	0.0182	0.0138	0.0102	0.0115	0.0286	0.0195	0.0150	0.0156	0.0306	0.0207	0.0159	0.0157	0.020
	0.25	0.0370	0.0359	0.0232	0.0349	0.0554	0.0472	0.0296	0.0428	0.0593	0.0506	0.0315	0.0454	0.051
	0.50	0.0720	0.0765	0.0557	0.0811	0.1016	0.0937	0.0593	0.0916	0.1099	0.1009	0.0631	0.0982	0.102
	0.75	0.1066	0.1172	0.0993	0.1278	0.1478	0.1434	0.0993	0.1445	0.1523	0.1440	0.0965	0.1440	0.150
	1.00	0.1427	0.1585	0.1531	0.1736	0.1953	0.1893	0.1398	0.1909	0.2075	0.1979	0.1404	0.1981	0.200
	2.00	0.3265	0.3649	0.5036	0.3964	0.3770	0.3759	0.3791	0.3809	0.4089	0.4020	0.3811	0.4054	0.408
10	0.10	0.0120	0.0097	0.0084	0.0084	0.0125	0.0102	0.0087	0.0088	0.0124	0.0101	0.0086	0.0087	0.010
	0.25	0.0258	0.0241	0.0186	0.0237	0.0267	0.0247	0.0190	0.0240	0.0276	0.0254	0.0195	0.0247	0.025
	0.50	0.0472	0.0463	0.0367	0.0472	0.0507	0.0490	0.0383	0.0494	0.0531	0.0504	0.0389	0.0506	0.050
	0.75	0.0691	0.0687	0.0575	0.0703	0.0728	0.0719	0.0596	0.0731	0.0768	0.0740	0.0589	0.0743	0.074
	1.00	0.0949	0.0957	0.0854	0.0980	0.0964	0.0951	0.0815	0.0961	0.1005	0.0979	0.0813	0.0983	0.101
	2.00	0.1882	0.1913	0.2111	0.1952	0.1943	0.1947	0.2039	0.1970	0.2046	0.2020	0.1968	0.2024	0.202
20	0.10	0.0051	0.0047	0.0043	0.0046	0.0055	0.0050	0.0045	0.0047	0.0058	0.0051	0.0047	0.0049	0.005
	0.25	0.0125	0.0122	0.0106	0.0123	0.0129	0.0124	0.0108	0.0125	0.0134	0.0128	0.0111	0.0128	0.013
	0.50	0.0240	0.0244	0.0218	0.0249	0.0252	0.0248	0.0217	0.0250	0.0261	0.0255	0.0222	0.0255	0.025
	0.75	0.0357	0.0363	0.0337	0.0370	0.0371	0.0368	0.0333	0.0372	0.0371	0.0366	0.0325	0.0367	0.038
	1.00	0.0461	0.0473	0.0462	0.0482	0.0505	0.0499	0.0456	0.0501	0.0509	0.0501	0.0452	0.0501	0.050
	2.00	0.0916	0.0942	0.1078	0.0957	0.0976	0.0974	0.0999	0.0979	0.1010	0.1006	0.1014	0.1009	0.100
50	0.10	0.0021	0.0020	0.0019	0.0020	0.0020	0.0019	0.0019	0.0019	0.0021	0.0020	0.0019	0.0020	0.002
	0.25	0.0049	0.0048	0.0045	0.0048	0.0051	0.0051	0.0048	0.0051	0.0051	0.0050	0.0047	0.0050	0.005
	0.50	0.0105	0.0104	0.0098	0.0104	0.0098	0.0097	0.0092	0.0098	0.0095	0.0094	0.0090	0.0095	0.010
	0.75	0.0148	0.0148	0.0142	0.0148	0.0151	0.0150	0.0143	0.0150	0.0149	0.0148	0.0141	0.0148	0.015
	1.00	0.0203	0.0203	0.0196	0.0204	0.0202	0.0201	0.0194	0.0201	0.0203	0.0202	0.0196	0.0203	0.020
	2.00	0.0393	0.0395	0.0409	0.0397	0.0405	0.0404	0.0407	0.0405	0.0412	0.0411	0.0413	0.0412	0.041
100	0.10	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0009	0.0010	0.0010	0.0010	0.0010	0.0010	0.001
	0.25	0.0025	0.0024	0.0024	0.0024	0.0025	0.0025	0.0024	0.0025	0.0026	0.0026	0.0025	0.0026	0.002
	0.50	0.0051	0.0051	0.0050	0.0051	0.0050	0.0050	0.0049	0.0050	0.0051	0.0050	0.0049	0.0050	0.005
	0.75	0.0077	0.0077	0.0075	0.0077	0.0076	0.0076	0.0074	0.0076	0.0074	0.0074	0.0072	0.0074	0.007
	1.00	0.0099	0.0099	0.0098	0.0099	0.0101	0.0101	0.0100	0.0101	0.0101	0.0101	0.0099	0.0101	0.010
	2.00	0.0198	0.0198	0.0200	0.0198	0.0190	0.0190	0.0193	0.0191	0.0205	0.0204	0.0204	0.0204	0.020

Table 2: Mean value of the squared difference three estimators with different constants (*c* in LELF, *k* in SSELF, *p* in GELF) and specific choices for *n* and θ; in all the cases β is in (1, 5)

Sample Size	θ	c = 3, k = p = 0.40			c = 5, k = p = 0.45			c = 10, k = p = 0.475		
		SSELF	LELF	GELF	SSELF	LELF	GELF	SSELF	LELF	GELF
5	0.10	0.0138	0.0102	0.0115	0.0179	0.0099	0.0141	0.0176	0.0056	0.0137
	0.25	0.0359	0.0232	0.0349	0.0461	0.0230	0.0420	0.0449	0.0181	0.0409
	0.50	0.0765	0.0557	0.0811	0.0921	0.0548	0.0912	0.0906	0.0609	0.0902
	0.75	0.1172	0.0993	0.1278	0.1381	0.1023	0.1413	0.1396	0.1303	0.1418
	1.00	0.1585	0.1531	0.1736	0.1848	0.1599	0.1894	0.1782	0.2306	0.1843
	2.00	0.3649	0.5036	0.3964	0.3763	0.5425	0.3873	0.3627	0.9038	0.3754
10	0.10	0.0097	0.0084	0.0084	0.0095	0.0066	0.0082	0.0090	0.0044	0.0078
	0.25	0.0241	0.0186	0.0237	0.0239	0.0160	0.0234	0.0233	0.0135	0.0230
	0.50	0.0463	0.0367	0.0472	0.0471	0.0355	0.0481	0.0473	0.0391	0.0482
	0.75	0.0687	0.0575	0.0703	0.0726	0.0619	0.0743	0.0706	0.0759	0.0719
	1.00	0.0957	0.0854	0.0980	0.0959	0.0916	0.0979	0.0934	0.1315	0.0958
	2.00	0.1913	0.2111	0.1952	0.1942	0.2677	0.1977	0.1888	0.4618	0.1924
20	0.10	0.0047	0.0043	0.0046	0.0047	0.0039	0.0046	0.0050	0.0033	0.0047
	0.25	0.0122	0.0106	0.0123	0.0121	0.0097	0.0122	0.0120	0.0087	0.0121
	0.50	0.0244	0.0218	0.0249	0.0238	0.0208	0.0242	0.0239	0.0225	0.0243
	0.75	0.0363	0.0337	0.0370	0.0366	0.0345	0.0372	0.0363	0.0399	0.0367
	1.00	0.0473	0.0462	0.0482	0.0488	0.0493	0.0495	0.0487	0.0651	0.0493
	2.00	0.0942	0.1078	0.0957	0.0999	0.1283	0.1009	0.1005	0.2064	0.1013
50	0.10	0.0020	0.0019	0.0020	0.0020	0.0018	0.0020	0.0020	0.0016	0.0020
	0.25	0.0048	0.0045	0.0048	0.0050	0.0045	0.0050	0.0049	0.0042	0.0049
	0.50	0.0104	0.0098	0.0104	0.0098	0.0092	0.0099	0.0096	0.0094	0.0096
	0.75	0.0148	0.0142	0.0148	0.0145	0.0141	0.0146	0.0154	0.0167	0.0155
	1.00	0.0203	0.0196	0.0204	0.0198	0.0198	0.0199	0.0192	0.0236	0.0193
	2.00	0.0395	0.0409	0.0397	0.0409	0.0464	0.0411	0.0399	0.0667	0.0401
100	0.10	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0009	0.0010
	0.25	0.0024	0.0024	0.0024	0.0025	0.0023	0.0025	0.0025	0.0023	0.0025
	0.50	0.0051	0.0050	0.0051	0.0050	0.0049	0.0051	0.0050	0.0050	0.0050
	0.75	0.0077	0.0075	0.0077	0.0073	0.0072	0.0073	0.0074	0.0082	0.0075
	1.00	0.0099	0.0098	0.0099	0.0096	0.0096	0.0096	0.0097	0.0113	0.0098
	2.00	0.0198	0.0200	0.0198	0.0199	0.0213	0.0199	0.0199	0.0283	0.0200

5. Data Analyses

The study has considered six data sets (I - IV) Prussian Horse Kick data which indicate the number of deaths by horse-kicks in the Prussian Army from 1875 to 1894 for 14 army corps for the purpose of illustration and demonstration of the procedures. Ten of the corps have a similar structure of 20 squadrons each and performed similar duties. The corps that has different structures (Guard Corps, Corps I,VI, and XI) has been considered; (V) Crab data (Agresti 2002) from a study of nesting horseshoe that investigates factors affecting whether the female crab has any other males, called satellites, residing nearby and the variable of interest for this study is the number of satellites; (VI) A data set on the number of awards earned by students at one high school (http://www.ats.ucla.edu/stat/r/dae/poissonreg.htm.).

Table 3 provides a summary of these six data sets; dispersion refers the ratio of variance to mean that indicates over or under dispersion relative to Poisson distribution. All data sets are over dispersed except data sets I and IV which are respectively no and under dispersed. Also the number of zeros is considerably high in the data set VI and II; moderately in I, III, IV and low in V. Data sets I to IV, V and VI can easily be grouped as large and small samples.

Table 3: Descriptive statistics for the illustrative six data sets under consideration

Description	Data sets					
	I	II	III	IV	V	VI
Sample size	20	20	20	20	173	200
Maximum	3	3	3	2	15	6
% of zeros	45	55	45	50	31	62
Dispersion	1	1.395	1.149	0.854	3.396	1.763

Table 4 provide Bayesian point estimates together with bootstrap estimates of bias and standard error (SE) for each of the six data sets. In the case of Bayesian estimates for the six data sets, the impact of prior choice especially on scale parameter and the choice of other unknowns (p in GELF and k in SSELF) can directly be observed from Table 4. The most visible comparison between SELF and LELF can be understood from the choice of c, in that both estimates are considerably closer for smaller values of c but not so otherwise, such as when $c = 5$ the choice made in this analysis.

For large sample with high proportion of zeros (data set VI), LELF has minimum SE when the prior parameter $\beta < 1$ and it has the least SE when compared to that of other three procedures. Point estimates from all the methods are closer to MLE irrespective of variation in β which might be due to sufficiently large sample. On the values of bias all are lower up to the order of 10^{-4} except for GELF when $\beta \leq 1$. Under the similar scenario but with less zeros (data set V) LELF has smaller SE and not affected by the changes of β but MLE and SELF estimates are comparably closer with smaller SE by SELF. SSELF behaves very similar to SELF in all the aspects whereas GELF has larger SE than others though its point estimates are closer.

On the other hand, for small samples with moderate zeros with dispersion equal to 1 (data set I) point estimates are similar when $\beta > 1$ except LELF that also yields smaller SE and it tends to increase with an increase of β. Further, with the increase of β, bias due to GELF increases but other methods show smaller bias values. Similar behaviour is visible in other small sample cases too with more zeros (data sets II and III) but the SE of all the four Bayesian procedures are comparatively lower than MLE.

Table 4: Bayesian point estimates using four loss functions and bootstrap standard error (SE) and bias for each of the estimators. Gamma prior parameters α, β and the constant c that appears in the LINEX estimator

$\alpha = 0.5, \beta < 1, c = 5$				$\alpha = 0.5, \beta = 1, c = 5$				$\alpha = 0.5, \beta > 1, c = 5$			
SELF	SSELF	LELF	GELF	SELF	SSELF	LELF	GELF	SELF	SSELF	LELF	GELF
0.7816	0.7783	0.7015	0.7526	0.7857	0.7848	0.7048	0.7593	0.8195	0.8145	0.7320	0.7911
0.1820	0.1820	0.1633	0.1819	0.1855	0.1855	0.1664	0.1854	0.1915	0.1915	0.1710	0.1914
-0.0005	-0.0005	-0.0005	-0.0516	-0.0007	-0.0007	-0.0006	-0.0552	0.0011	0.0011	0.0010	-0.0581
0.6447	0.6354	0.5889	0.6220	0.7857	0.7847	0.7048	0.7612	0.8122	0.8103	0.7261	0.7856
0.1816	0.1816	0.1658	0.1815	0.2212	0.2212	0.1984	0.2211	0.2273	0.2273	0.2032	0.2273
0.0047	0.0047	0.0043	-0.0284	0.0005	0.0005	0.0005	-0.0559	-0.0020	-0.0020	-0.0018	-0.0615
0.7869	0.7799	0.7098	0.7622	0.8333	0.8304	0.7475	0.8042	0.8688	0.8629	0.7760	0.8388
0.1932	0.1932	0.1743	0.1932	0.2024	0.2024	0.1816	0.2024	0.2165	0.2165	0.1934	0.2165
-0.0020	-0.0020	-0.0018	-0.0544	0.0017	0.0017	0.0016	-0.0550	0.0023	0.0023	0.0020	-0.0605
0.5978	0.5974	0.5400	0.5703	0.6429	0.6411	0.5767	0.6161	0.6644	0.6585	0.5940	0.6391
0.1436	0.1436	0.1297	0.1436	0.1560	0.1560	0.1399	0.1559	0.1608	0.1608	0.1438	0.1608
0.0007	0.0007	0.0007	-0.0296	-0.0005	-0.0005	-0.0004	-0.0399	-0.0005	-0.0005	-0.0004	-0.0456
2.9005	2.8993	2.8596	2.8971	2.9052	2.9049	2.8642	2.9018	2.9158	2.9155	2.8745	2.9122
0.2361	0.2361	0.2328	0.2361	0.2360	0.2360	0.2327	0.2360	0.2383	0.2383	0.2349	0.2383
-0.0002	-0.0002	-0.0002	-0.0377	0.0035	0.0035	0.0034	-0.0341	-0.0009	-0.0009	-0.0009	-0.0386
0.6237	0.6225	0.6162	0.6208	0.6294	0.6292	0.6217	0.6266	0.6320	0.6318	0.6243	0.6293
0.0724	0.0724	0.0716	0.0724	0.0747	0.0747	0.0738	0.0747	0.0741	0.0741	0.0732	0.0741
0.0001	0.0001	0.0001	-0.0045	-0.0003	-0.0003	-0.0003	-0.0052	-0.0020	-0.0020	-0.0019	-0.0070

6. Conclusions

Evaluation and comparison of Bayesian models are quite pervasive in statistical literature. This includes the sensitivity of posterior distribution to the prior, optimizing loss functions to obtain non-standard inferences and having data-analytic performance (Lin et al, 2006 and Sanjay Kumar Singh et al, 2014) and information criterions (Gelman et al, 2014). Some studies have explored the ways to favour objective priors in specific likelihood-prior-posterior situations (Tuyl et al, 2008 and 2009).

This work focuses on loss function optimality for obtaining estimators apart from the frequently used candidate, the posterior mean (Bolstad, 2013) so as to accommodate symmetric and asymmetric loss functions. The other component of the paper is to quantify the prior parameters through the behaviour of conditional risk of estimators which is akin to decide or favour an objective prior for a model through predictive distributions. The Poisson-Gamma model has been considered in the deduction that it provides an active scope to study over and under estimation issues in a Bayesian perspective.

Though this univariate model is a regular inclusion in standard texts of statistics of any paradigm, the present work has made an attempt to derive explicit forms for obtaining Bayesian estimates based on posterior expected loss and conditional risk associated with four loss functions. The study has favoured Jeffreys prior for this model based on risk function and its behaviour for various inputs and prior parameters. For an empirical evaluation on the performance of estimates, a simulation study has been carried out using different values of sample size (n), Poisson parameter (θ), scale parameter (β) of prior Gamma distribution and two model specific constants.

- Closed form estimates are made available under the four loss functions for Poisson-Gamma model.
- Conditional risk function can also be obtained in closed form except for GELF in which case, its behaviour (in an approximate sense) can be studied using properties of Gamma functions and Delta method.

- For Poisson-Gamma model, Jeffreys prior can be made consensus prior and the impact of shape parameter needs attention.
- SELF is tend to overestimate for $\theta < 0.5$, for larger n and all choices of $\beta > 0$; this becomes more stable when n increases.
- LINEX tend to under estimate the parameter when $n < 50$ irrespective the choice of β which might be the influence of smaller c in LINEX.
- Other two Bayesian estimates due to SSELF and GELF behave very similar even in the case of $n < 50$; the trade-off between under and over-estimation is based on the choice of smaller and larger values of β.
- As expected, impact of the prior (β) is very minimal when n is 50 or more and all the estimates are quite similar in comparison to sample mean, the MLE of this model. A few cases of over estimation is also visible for lower n.
- For any n, the constant c in LINEX and parameter θ is inversely related and hence larger θ requires smaller c but not less than 3. Very small values of c may not distinguish LINEX from SELF.
- But the constant p (GELF) and k (SSELF) need not be made too close to the prior shape parameter α for any θ and when $n < 50$; this can be relaxed as n increases.

Constructing plausible priors through subjectivism or hierarchical method finds the research attention in applied problems; still it is imperative to study the impact of an objective prior through any appropriate criterion that could be useful for applications (Ellison, 2004). Many practical applications have benefitted from relaxing symmetric assumptions (Elliot et al, 2008), and hence a comprehensive evaluation may indicate the choice of a reasonable prior to begin with a simple model which is a basic principle in a applied statistical investigation (Gelman et al, 2014) before attempting larger models.

References

1. Agresti, A. and Min, Y. (2005): Frequentist performance of Bayesian confidence intervals for comparing proportions in 2 × 2 contingency tables. Biometrics, 61: 515-523.
2. Bolstad, W.M. (2013): Introduction to Bayesian Statistics, John Wiley & Sons
3. Brown, L.D., Cai, T.T. and Das Gupta, A. (2001): Interval estimation for a binomial proportion. Statistical Science, 16: 101-133.
4. Campbell, I. (2007): Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Statistics in Medicine, 26: 3661-3675.
5. Casella. G. and Berger. R.L.(2002): Statistical Inference, Duxbury, Belmont, CA.
6. Elliott, G., Komunjer, I. and Timmermann, A. (2008): Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss? Journal of the European Economic Association, 6, 1: 122-157.
7. Ellison, A.M. (2004): Bayesian inference in ecology. Ecology Letters, 7: 509-520.
8. Gelman, A., Hwang,. and Vehtari, A. (2014): Understanding predictive information criteria for Bayesian models, Statistics and Computing, 24, 6: 997-1016.
9. Hwang, L.H. and Lee, C.H. (2013): Bayes sequential estimation for a Poisson process under a LINEX loss function, Statistics, 47;:672-687.
10. Jianhui Ning, J. and Xie, M. (2007): Minimax Invariant Estimator of Continuous Distribution Function Under LINEX Loss, JrlSystSci & Complexity 20,1:119-126.
11. Lee, C.H. and Hwang, L.H. (2011): Asymptotic optimal estimation of Mean under LINEX loss function, Communications in Statistics - Theory and Methods, 40: 4308-4321.
12. Lehmann, E.L. and Casella. G. (1998): Theory of point estimation. Springer (India) Private Limited, New Delhi.
13. Lin, R., Louis, A.T., Paddock ,S.M. and Ridgeway, G. (2006): Loss Function Based Ranking in Two-Stage, Hierarchical Models, Bayesian Analysis, 1, Number 4:915-945.

14. Patil, V. V., and Kulkarni, H.V. (2012): Comparison of confidence intervals for the Poisson mean: some new aspects, REVSTAT - Statistical Journal, 10: 211-227.
15. Qi and Luo (2013): Bounds for the ratio of two gamma functions: from Wendel's asymptotic relation to Elezović-Giordano-Pećarić's theorem. Journal of Inequalities and Applications, 2013: 542.
16. Robert, C.P. (1994): The Bayesian choice. A decision-theoretic motivation. Springer-Verlag, New York.
17. Sanjay Kumar Singh, Umesh Singh and Manoj Kumar (2014): Estimation for the Parameter of Poisson-Exponential Distribution under Bayesian Paradigm, Journal of Data Science, 12: 157-173.
18. Subbiah, M., Kumar, B. and Srinivasan, M.R. (2008): Bayesian approach to multicentre sparse data. Communications in Statistics-Simulation and Computation:, 3: 687 - 696.
19. Sweeting, J.M., Sutton, J.A. and Lambert, C.P. (2004): What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Statistics in Medicine, 23: 1351 -1375.
20. Tuyl, F., Gerlach, R. and Mengersen, K. (2008): A comparison of Bayes-Laplace, Jeffreys and other Priors: The case of zero events. The American Statistician, 62: 40 - 44.
21. Tuyl, F., Gerlach, R. and Mengersen, K. (2009): Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameter, Bayesian Analysis, 4(1): 151-158
22. Wan Tze-Kin Alan, Guohua Zou and Lee, A.H. (2000): Minimax and Γ - minimax estimation for the Poisson distribution under LINEX loss when the parameter space is restricted. Statistics and Probability letters, 50.1: 23-32.