Bayesian Inference for Bernoulli Distribution Using Different Loss Functions

KEYWORDS: Bernoulli distribution, loss functions, posterior risk, Bayes estimates

D. Sheeba Singh
Assistant Professor, Noorul Islam University, Kumaracoil, Thuckalay.

M. Immaculate Mary
Professor, Noorul Islam University, Kumaracoil, Thuckalay

ABSTRACT *Bayesian estimation and inference has a number of advantages in statistical modelling and data analysis. It provides a way of formalising the process of learning from data to update beliefs in accord with recent notions of knowledge synthesis.In this paper the Bernoulli distribution is taken for Bayesian analysis. The properties of Bayes estimates of the parameters are studied under different loss functions through simulated and real life data. Different priors like informative and non-informative are used to estimate the parameters. The loss functions are compared through posterior risk.*

Introduction

Bernoulli distribution is a discreteprobability distribution, which takes value 1 with success probability p and value 0 with failure probability q = 1 - p. If X is a random variable with this distribution, we have:Pr(X = 1) = 1- Pr(X = 0) = 1 - q = p

The probability mass function of this distribution is

$$f(k;p)=\begin{cases} p & if\ k=1 \\ 1-p & if\ k=0 \end{cases}$$

This can be expressed as $f(k;p)=p^k(1-p)^{1-k}\ \ for\ k\in\{0,1\}$. The mean of a Bernoulli random variable X is E(X) = p, and its variance is

Var(X) = p (1 - p)
The cumulative distribution function of Xfollowing Bernoulli distribution is

$$F(x)=\begin{cases} 0 & x<0 \\ 1-p & 0\le x<1 \\ 1 & x\ge 1 \end{cases}$$

Posterior distributions and Likelihood function:

The posterior distribution summarizes available probabilistic information on the parameters in the form of prior distribution and the sample information contained in the likelihood function. The likelihood principle suggests that the information on the parameter should depend only on its posterior distribution. Bayesian scientist's job is to assist the investigator to extract features of interest from the posterior distribution. In this section we will use the Bernoulli distribution as sampling distribution mingles with the informative and noninformative priors for the derivation of posterior distribution

Let the random variable $y_i$follow a Bernoulli distribution with parameter θ.

Then $p(y_i/\theta)=\theta^{y_i}(1-\theta)^{1-y_i}\ \ ,i=1,2,\ldots,n$

The likelihood function is

$$L(y_i,\theta)=\prod_{i=1}^{n}p(y_i/\theta)=\theta^{\sum y_i}(1-\theta)^{n-\sum y_i}$$

Posterior Distribution using Conjugate Prior(CP):

A flexible choice of conjugate prior distribution for a Bernoulli distribution is a beta distribution with parameters αandβ. The probability function of the beta distribution is given by

$$g(\theta)=\frac{\Gamma\alpha+\beta}{\Gamma\alpha\,\Gamma\beta}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

The joint probability density function (pdf) $H(y_1,y_2,\ldots,y_n,\theta)$ is given by

$$H(y_1,y_2,\ldots,y_n,\theta)$$

$$=L(y_1,y_2,\ldots,y_n,\theta)\ g(\theta)$$

where $L(y_1,y_2,\ldots,y_n,\theta)$ is the likelihood function of $P(y_i/\theta)$ and $g(\theta)$ is the prior distribution.

Hence the joint pdf is $H(y_1,y_2,\ldots,y_n,\theta)=$ $\frac{\Gamma\alpha+\beta}{\Gamma\alpha\Gamma\beta}\theta^{\alpha+\Sigma y_i-1}(1-\theta)^{\beta+n-\Sigma y_i-1}$

The marginal pdf of $y_1,y_2,\ldots,y_n$ is given by$p(y_1,y_2,\ldots\ldots y_n)$

$$=\int H(y_1,y_2,\ldots\ldots y_n,\theta)\,d\theta$$

$$=\frac{\Gamma\alpha+\beta}{\Gamma\alpha\Gamma\beta}\frac{\Gamma\alpha+\Sigma y_i\Gamma\beta+n-\Sigma y_i}{\Gamma\alpha+\beta+n}$$

The posterior pdf $P(\theta/y_i)$ of θ is given by

Volume : 4 | Issue : 3 | Mar 2014 | ISSN - 2249-555X

$$\pi(\theta / y_1, y_2, ..., y_n) = \frac{H(y_1, y_2, ..., y_n, \theta)}{P(y_1, y_2, ..., y_n)}$$

$$= \frac{\Gamma\alpha+\beta+n}{\Gamma\alpha+\sum y_i \Gamma\beta+n-\sum y_i} \theta^{\alpha+\sum y_i-1}(1-)^{\beta+n-\sum y_i-1}$$

which is again a beta distribution with parameters α = α+$\sum y_i$, β = β+ n-$\sum y_i$ and mean $\frac{\alpha+\sum y_i}{\alpha+\beta+n}$. The posterior mean is used to compute the point estimates for the parameter.

Posterior Distribution using Uniform Prior(UP):

The prior distribution of θ can be taken as general uniform distribution with pdf = $g(\theta) = \frac{1}{\theta}$ $\quad 0 \leq \theta \leq 1$

The Bayesian posterior distribution for the parameter θ is $\pi(\theta / y_1, y_2, ..., y_n)$

$$= \theta^{\sum y_i - 1}(1-\theta)^{n-\sum y_i}$$

which is a beta distribution with parameters α = $\sum y_i$, β = n$-\sum y_i$+1 and mean $\frac{\sum y_i}{n+1}$

Posterior Distributionusing Jeffrey's Prior (JP):

The Jeffery's prior information,

$$g(\theta) = k\left(\frac{1}{\theta(1-\theta)}\right)^{\frac{1}{2}}$$

Posterior probability function is given by

$$\pi(\theta / y_1, y_2, ..., y_n) = \frac{\theta^{(\sum y_i + \frac{1}{2})-1}(1-\theta)^{(n-\sum y_i+\frac{1}{2})-1}}{C}$$

wherec =β($\sum y_i + \frac{1}{2}$, n $-\sum y_i + \frac{1}{2}$) which is again a beta distribution with parameters α =$\sum y_i + \frac{1}{2}$, β = $n - \sum y_i + \frac{1}{2}$ and mean $\frac{\sum y_i - \frac{1}{2}}{n+1}$

Bayesian Estimation Under Different Loss Functions:

Linear ExponentialLoss Function

The linear exponential (LINEX) loss function is an asymmetric loss function. This loss function rises approximately exponentially onone side of zero and approximately linearly on the other side. It is under the assumption that the minimal loss occurs at $\hat{\theta} = \theta$ and is expressed as

$L(\Delta) \propto \exp(a\Delta) - (a\Delta) - 1$; $a \neq 0$ ---------------- (1)

with $\Delta = (\hat{\theta} - \theta)$, where $\hat{\theta}$ is an estimate of θ.The sign and magnitude of the shape parameter 'a' represents the direction and degree of symmetry, respectively. There is overestimation if a > 0 and underestimation if a < 0 but when $a \cong 0$, the LINEX loss function is approximately the squared error loss function. The posterior expectation of the LINEX loss function, according is

$$E_\theta[L(\hat{\theta} - \theta)] \propto \exp(a\hat{\theta}) E_\theta[\exp(-a\theta)] - a(\hat{\theta} - E_\theta(\theta)) - 1$$

------------------ (2)

The Bayes estimator of θ, represented by $\hat{\theta}_L$ under LINEX loss function, is the value of $\hat{\theta}$ which minimizes (2) and is given as

$$\hat{\theta}_L = -\frac{1}{a} \ln E_\theta[\exp(-a\theta)]$$

----------------- (3)

provided that $E_\theta[\exp(-a\theta)]$ exists and is finite. The Bayes estimator $\hat{u}_L$ of a function $u = [\exp(-ab), \exp(-a\eta)]$ is given as

Volume : 4 | Issue : 3 | Mar 2014 | ISSN - 2249-555X

$\hat{u}_L = \mathrm{E}[\exp(-ab), \exp(-a\eta)/y]$

$$= \frac{\iint u[\exp(-ab), \exp(-a\eta)]\, \pi(b,\eta)\, db\, d\eta}{\iint \pi(b,\eta)\, db\, d\eta}$$

------------------ (4)

From equation (4), it can be observed that it contains a ratio of integrals which cannot be solved analytically, and for that we employ Linley's approximation procedure to estimate the parameters. Lindley considered an approximation for the ratio of integrals for evaluating the posterior expectation of an arbitrary function $\hat{u}(\theta)$ as

$$E[u(\theta)/x)] = \frac{\int u(\theta) v(\theta) \exp[L(\theta)] d\theta}{\int v(\theta) \exp[L(\theta)] d\theta}$$

Lindley's expansion can be approximated asymptotically by

$$\hat{\theta} = u + \frac{1}{2}[(u_{11}\,\delta_{11}) + (u_{22}\,\delta_{22})] + u_1\,\rho_1\,\delta_{11}$$

$$+ u_2\,\rho_2\,\delta_{22} + \frac{1}{2}[(L_{30}\,u_1\,\delta_{11}^2)(L_{03}\,u_2\,\delta_{22}^2)]$$

where L is the log- likelihood function, and

$$u(b) = \exp(-ab),\ u_1 = \frac{\partial u}{\partial b} = -a\exp(-ab),$$

$$u_{11} = \frac{\partial^2 u}{\partial b^2} = -a^2 \exp(-ab)$$

$$u(\eta) = \exp(-a\eta),$$

$$u_2 = \frac{\partial u}{\partial \eta} = -a\exp(-a\eta),$$

$$u_{22} = \frac{\partial^2 u}{\partial \eta^2} = -a^2 \exp(-a\eta)$$

$$\rho = \text{log of prior},\ \rho_1 = \frac{\partial \rho}{\partial b}, \rho_2 = \frac{\partial \rho}{\partial \eta}\ and$$

$$\delta_{11} = (-L_{20})^{-1}, \delta_{22} = (-L_{02})^{-1}$$

Entropy Loss Function

Another useful asymmetric loss function is a entropy loss function (ENLF) which is given as $L(\hat{\theta} - \theta) \propto$

$$\left(\frac{\hat{\theta}}{\theta}\right)^k - k\ln\left(\frac{\hat{\theta}}{\theta}\right) - 1$$

The Bayes estimator $\hat{\theta}_G$ of θ under the entropy loss function is

$\hat{\theta}_G = [E_\theta(\theta^{-k})]^{-\frac{1}{k}}$ provided $E_\theta(\theta^{-k})$ exists and is finite. The Bayes estimator for this loss function is

$$\hat{u}_G = E\{u[(b)^{-k}, (\eta)^{-k}]/t\}$$

$$= \frac{\iint u[(b)^{-k}, (\eta)^{-k}]\, \pi(b,\eta)\, db\, d\eta}{\iint \pi(b,\eta)\, db\, d\eta}$$

Similar Lindley approach is used for the general entropy loss function as in the LINEX loss but here the Lindley approximation procedure as stated in (4), where u_1,u_{11} and u_2,u_{22} are the first and second derivatives for α and β, respectively, and are given as

$$u = (b)^{-k}, u_1 = \frac{\partial u}{\partial b} = -k(b)^{-k-1},$$

$$u_{11} = \frac{\partial^2 u}{\partial b^2} = -(-k^2 - k)(b)^{-k-2},$$

$u_2 = u_{22} = 0$,

$$u = (\eta)^{-k}, u_2 = \frac{\partial u}{\partial \eta} = -k(\eta)^{-k-1},$$

$$u_{22} = \frac{\partial^2 u}{\partial \eta^2} = -(-k^2 - k)(\eta)^{-k-2},$$

$u_1 = u_{11} = 0$

Squared Error Loss Function

The squared error loss function(SELF) is symmetric in nature and is given by $l(\hat{\theta} - \theta) = (\hat{\theta} - \theta)^2$. The Bayes estimator of a function $u = u(b,\eta)$ of the unknown parameters under squared error loss function is the posterior mean. where

$$\hat{u} = E[u(b,\eta)/t] = \frac{\iint u(b,\eta)\, \pi^*(b,\eta)\, db\, d\eta}{\iint \pi^*(b,\eta)\, db\, d\eta}$$

Volume : 4 | Issue : 3 | Mar 2014 | ISSN - 2249-555X

Applying the same Lindley approach here with u_1, u_{11} and u_2, u_{22} being the first and second derivatives for α and β, respectively, we have

$$u = b \quad , u_1 = \frac{\partial u}{\partial b} = 1, \; u_{11} = u_2 = u_{22} = 0,$$

$$u = \eta \quad , u_2 = 1 \quad , u_{11} = u_1 = u_{22} = 0$$

Numerical Data Analysis:

In Kanyakumari districtthere are four taluksvizAgastheeswaram, Kalkulam, Thovalai, and Vilavancode. The dengue patients reported from the health centres of the above four taluks from the year 2009 to 2012 are taken for the study. In each year, out of a sample of 'n' patients affected by fever, $\sum y_i$ were affected by dengue.

The original data follows a Bernoulli distribution with parameter θ. The flexible choice of a prior distribution for a Bernoulli probability is θ ~ Beta (α,β) that is, θ has a beta distribution with specified parameters α and β.

The posterior summary of the real life data is shown in Table 1. The Bayesian estimates and the posterior risk under different loss functions is given in Table 2.

Simulation Study

Here a simulation criterion is used and for each n = 50, 75 and 100, the Bayes estimates and the Bayes posterior risks are calculated under different loss functions along with different priors. The comparison of Bayes posterior risk under different loss function using different priors has been made through which we can conclude from Table 3 to Table 5that within each loss functions the conjugate prior provides less Bayes posterior risk so it is more suitable for the class of life-time distributions and amongst loss functions, LINEX loss function, is more preferable as compared to all other loss functions which are provided here because under this loss function Bayes posterior risk is small for each and every value of parameter .

CONCLUSION

The Bayesian analysis of the Bernoulli distribution is done using different priors. After analysis the conjugate prior of Beta distribution is compatible for the unknown parameter of the distributions and preferable over all other competitive priors because of having less posterior variance along with less Skewness combined with less Kurtosis.As far as choice of loss function is concerned, one can easily observe based on evidence of different properties as discussed above that the LINEX loss function has smaller posterior risk. As we increase sample size posterior risk comes down. In future, this work can be extended using class of life-time truncated distributions and considering location parameter. The study is useful for the researchers, practitioners and also for scientists in the field of physics and chemistry or other fields where life-time distribution are extensively used.

Table 1 Posterior Summary using the real data with different priors for the year 2009, 2010, 2011, 2012

Year	2009			2010		
	CP	UP	JP	CP	UP	JP
Mean	0.19777	0.1979	0.19815	0.19684	0.19134	0.1914
Mode	0.19747	0.1976	0.19787	0.19655	0.19104	0.19111
Variance	0.00007	0.00008	0.00008	0.00007	0.00007	0.00007
Skewnes	0.06731	0.06775	0.06766	0.06604	0.06843	0.06841
Kurtosis	0.00381	0.00389	0.00387	0.00373	0.00417	0.00416
	2011			2012		
Mean	0.17894	0.1777	0.17792	0.16353	0.1645	0.16373
Mode	0.17865	0.17741	0.17763	0.16327	0.16423	0.16347
Variance	0.00062	0.00064	0.00065	0.54673	0.54739	0.54726
Skewnes	0.07048	0.07104	0.07095	0.07272	0.07222	0.07264
Kurtosis	0.00479	0.0049	0.00489	0.00553	0.00543	0.00552

Table 2 Bayes estimates and posterior risk using different loss functions for 2009,2010, 2011, 2012 for the real life data

Year	2009			2010		
Prior	SELF	ELF	LINEX	SELF	ELF	LINEX
CP	0.19876(0.00172)	0.79825(0.359224)	0.19596(0.002494)	0.19684(0.00168)	0.642712(0.34622)	0.09792(0.02494)
UP	0.19794(0.00176)	0.505306(0.18369)	0.19514(0.001251)	0.19131(0.00154)	0.522629(0.17299)	0.09515(0.01251)
JP	0.19815(0.00178)	0.250874(0.12389)	0.19535(0.000835)	0.19156(0.00132)	0.254191(0.14669)	0.09528(0.00835)
	2011			2012		
CP	0.17894(0.00159)	0.487532(0.31522)	0.08897(0.001994)	0.16449(0.00142)	0.426743(0.31322)	0.08174(0.001824)
UP	0.17771(0.00149)	0.562525(0.15299)	0.08835(0.000951)	0.16352(0.00138)	0.611509(0.11169)	0.08125(0.000781)
JP	0.17792(0.00128)	0.26148(0.13569)	0.08846(0.000435)	0.16373(0.00114)	0.270248(0.10599)	0.08136(0.000375)

Bayes estimates and respective posterior risk under simulation study

Table 3 Bayes estimates and posterior risk using priors under Squared Error Loss Function

θ	0.5			1		
n	CP	UP	JP	CP	UP	JP
50	0.17519 (0.000038)	0.16352 (0.000036)	0.16373 (0.000041)	0.29966 (0.00089)	0.44874 (0.05525)	0.61935 (0.01958)
75	0.16483 (0.000044)	0.16244 (0.000038)	0.16229 (0.000051)	0.39114 (0.01634)	0.61663 (0.0613)	0.56892 (0.0178)
100	0.16531 (0.000045)	0.16256 (0.000037)	0.16233 (0.000050)	0.32757 (0.00804)	0.66381 (0.00156)	0.284 (0.0066)

Table 4 Bayes estimates and posterior risk using priors under LINEX loss function

θ	0.5			1		
n	CP	UP	JP	CP	UP	JP
50	0.00018 (0.000055)	6.14024 (0.000054)	2.70828 (0.000054)	0.000011 (0.000065)	2.22846 (0.000055)	2.05953 (0.000035)
75	0.00011 (0.000045)	6.15612 (0.000044)	2.71211 (0.000044)	0.000476 (0.000055)	1.62171 (0.000043)	2.01927 (0.000037)
100	0.00008 (0.000035)	6.15157 (0.000034)	2.71184 (0.000034)	0.000565 (0.000035)	1.50645 (0.000032)	2.21761 (0.000027)

Volume : 4 | Issue : 3 | Mar 2014 | ISSN - 2249-555X

Table 5 Bayes estimates and posterior risk using priors under Entropy loss function

θ	0.5			1		
n	CP	UP	JP	CP	UP	JP
50	0.17519 (0.45011)	0.163522 (0.45922)	0.16373 (0.46076)	0.28169 (1.00056)	0.17935 (1.00068)	0.18058 (1.00132)
75	0.16483 (0.45734)	0.176011 (0.45910)	0.16427 (0.45189)	0.27133 (1.01432)	0.16899 (1.02391)	0.17022 (1.03472)
100	0.16531 (0.45830)	0.177043 (0.449378)	0.16745 (0.41489)	0.27181 (1.03425)	0.16947 (1.05642)	0.17073 (1.07638)

REFERENCE 1. Al-Hussaini, E. K. (1999), Predicting observables from a general class of distributions", Journal of Planning and Statistical Inference, vol. 79, 79-91. | 2. Dey, S. and Sudhansu, S.M. (2010). Bayesian estimation of the parameter of Maxwell distribution under different loss functions. J. Stat. Theory Pract., 4, 279-287. | 3. D. K. Dey, M. Ghosh, and C. Srinivasan, (1987), Simultaneous estimation of Parameters under entropy loss, J. Statist. Plann. Inference,347--363. | 4. Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). Bayesian Data Analysis, 2nd ed. Chapman and Hall/CRC, New York. | 5. Lindley, D.V., (1980). Approximate Bayesian Methods. Trabojos de Estadistica, vol. 31: 223-245. | 6. Mu Zhu, Arthur Y. Lu(2004),The Counter-intuitive Non-informative Prior for the Bernoulli Family, Journal of Statistics Education Volume 12, Number 2 . | 7. Zellner A. (1986), Bayesian estimation and prediction using asymmetric loss func- tions. Journal of the American Statistical Association; vol. 81:446–51. |