International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 4, No. 2 (Nov-2017)

Comparative Study on Estimation of Poisson Parameter in Decision Theoretic Approach

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

[1] Department of Statistics, Presidency College, Chennai, India
[2]Department of Mathematics, L N Government College, Ponneri, Tamilnadu, India
[3]Department of Statistics, University of Madras, Chennai, India

Received May 20, 2017, Revised September 21, 2017, Accepted October 11, 2017 Published November 1, 2017

Abstract: It has become imperative in statistical inference to seek beyond the widely practiced squared loss function to accommodate the asymmetric conditions. An attempt in this direction is generally model-specific and the present work has considered inference on count data, as one such model of importance. In particular, the objective is to study the performance of Bayesian estimators for Poisson parameter based on four well established loss functions. The explicit forms are derived and a comprehensive data analyses has been carried out through a simulation study. The study highlights the distinct behaviour of each of the methods to make an appropriate choice based on the small sample behaviour of the data sets.

Keywords: Poisson mean, Loss function, Gamma distribution, Normal approximation

1. Introduction

Estimation of parameters in discrete distributions has witnessed active research debate and discussion; yet it provides ample scope to pursue research for a better understanding and conclusion [7, 10, 14, 15]. This is true in the case of Poisson distribution that has been widely accepted for count data models, especially to rare events. Sample mean has been considered as the most celebrated unbiased minimum variance estimator for Poisson parameter in frequentist approach. However, limitation of this simple statistic is in terms of robustness in that a single value can skew the mean and many standard statistical texts [1, 6, 8] have addressed this problem.

The problem of statistical inference beyond unbiased estimators is also well discussed in the literature [3, 9, 13, 23]. Generally, mean squared error or loss function optimality and Bayes risk are considered as major evaluation criteria in comparison of estimators. This paper is confined to Bayesian approach with risk function optimality to investigate competing estimators for Poisson parameter. Loss function optimality using Bayes risk has drawn research attention in many studies [20,11,19] for obtaining point estimators.

In this direction, an attempt has been made with Poisson model in Bayesian perspective. This study aims to compare the effect of symmetric and asymmetric loss functions. Point estimates are derived from four loss functions. Comparison for their performance is carried out using Bootstrap method based on the observations made in Efron [17, 19, 22].

The following section elaborates on the methods for estimating Poisson mean in the realm of Bayesian inference, and explicit expressions for point estimation. Section 3 and 4 provide details of simulation study and analyses based on two practical data sets respectively and finally the conclusive remarks are presented in the last section.

2. Estimators for Poisson parameter

If $X_1, X_2,\ldots\ldots\ldots,X_n$ are iid random variables distributed as Poisson (θ), then

$Y = \sum X_i \sim$ Poisson ($n\theta$)

with the likelihood of θ/y as

E-mail address: renuka2327@yahool.com, sisufive@gmail.com, mrsvasan59@yahoo.com

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

The major aspect of this study is to obtain estimators $\delta(x_1, x_2, \dots, x_n)$ based on loss function optimality; to choose an action (in terms of δ) that minimizes loss and the risk function (a function of θ). That is, risk function is the average loss $R(\theta, \delta) = E[L(\theta, \delta)]$. The widely used loss function includes squared error loss and absolute error loss that yield respectively mean and median as estimators for θ. Further, many studies [12, 13, 16] have shown the importance of including asymmetric loss functions such as Stein's loss or LINear-EXponential (LINEX) loss functions [8]. Four loss functions have been considered and the loss function optimality in Bayesian inference is a direct approach of minimizing the posterior expected loss.

L1. Squared Error Loss Function (SELF)

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

L2. Scaled Squared Error Loss Function (SSELF)

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

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

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

$= \delta^2 E[\theta^{-k}] - 2\delta E[\theta^{1-k}] + E[\theta^{2-k}]$, a quadratic function in δ.

On differentiation with respect to δ and equating to zero, we get $\delta_2 = \frac{E[\theta^{1-k}]}{E[\theta^{-k}]}$

L3. LINEX Error Loss Function (LELF)

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

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

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

Differentiating the above with respect to δ and equating to zero, gives

$$\delta_3 = -\frac{1}{c}\ln E[e^{-c\theta}]$$

L4. General Entropy Loss Function (GELF) [5]

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

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

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

Differentiating with respect to δ and equating to zero again, gives

$$\delta_4 = \left[\frac{1}{E(\theta^{-p})}\right]^{\frac{1}{p}}$$

In addition to the above, two more loss functions have extensive applications. It may be noted that with w = p = 1 in GELF, provide Stein's Error Loss Function; $L(\theta,\delta) = \frac{\delta}{\theta} - 1 - \log\frac{\delta}{\theta}$ and with k = 2 in SSELF leads to Quadratic Error Loss Function $L(\theta,\delta) = \left[\frac{\theta-\delta}{\theta}\right]^2$

Subsequently, the study has considered comparing the performance of these point estimators obtained using L1 to L4 with five prior combinations. Three prior specifications are based on the model, $Y = \sum X_i \sim$ Poisson $(n\theta)$ with θ follows Gamma (GAM), or Truncated Normal (TRN), or normal distribution for $\ln\theta$ (FEM); whereas other two prior schemes (REM and ZIP) are working with normal and Bernoulli distributions on the appropriate transformations on θ_i under the likelihood Xi ~ Poisson (θ_i). The comparison of loss functions is carried out through a simulation study followed by real data sets to validate the results and are implemented in R using R2WinBugs package [20].

3. SIMULATION STUDY

The simulation study includes four parametric values of θ (= 0.01, 0.5, 2, and 5) to reflect a plausible range of small to larger positive real value. However, parametric values greater than 5 do not reflect appreciable change in the comparison study and hence are not presented. The inference problem could be affected by the sample size and four representative values are considered for n (= 10, 50, 100, and 500). Suitable constants are assumed to all hyper priors involved in the five prior specifications. Mean squared error is the evaluation criteria based on 500 simulated values of each combination of n and θ and the respective estimates are presented in Table 1.

The simulation largely indicates that MSE due to all four loss functions decreases with increasing values of n and for all values of θ. Also, all loss functions have MSE closer to zero under GAM, TRN uniformly when θ (< 1.0) for all values of n. The loss functions are observed to be having an oscillatory behaviour based on the other three priors, especially FEM and REM, for all n.

SELF favours REM in terms of smaller MSE for all values of θ < 5and n < 200. Under SELF, FEM has uniformly higher values of MSE (greater than 12) when θ = 5 but the comparison is reversed for n = 200 and θ = 5. Similar observation can be made under SSELF that favours FEM for smaller n and θ; whereas for increasing values of n and θ < 5, REM has comparatively smaller MSE.

LELF shows a decreasing MSE under REM when n increases with smaller θ but not favouring FEM in those cases which has MSE larger than 13; when n is small both priors are not so favourable. In the case of GELF, REM shows smaller MSE for intermediate values of n and all values of θ but MSE is 15.02 when n is 10 and θ is 5. Also, FEM seems favourable in n = 10 and 200. For a comparative purpose when n = 10 and θ = 10, maximum MSE can be observed for SELF under FEM (12.319); for SSELF (12.661), LELF (20.364), and GELF (15.02) under REM. However, the behaviour of all loss functions under ZIP shows consistency in MSE that lies between the pairs GAM, TRN and FEM, REM when θ is smaller and for all values of n and increases for higher values of θ.

TABLE 1. MEAN SQUARED ERROR OF THE ESTIMATORS OBTAINED FROM FOUR LOSS FUNCTIONS AND FIVE PRIOR DISTRIBUTIONS FOR DIFFERENT SAMPLE SIZE (N) AND THE PARAMETER (Θ)

Loss function	**Prior**	***n = 10***				***n = 50***				***n = 100***				***n = 200***			
		θ = 0.01	***θ = 0.5***	***θ = 2***	***θ = 10***	***θ = 0.01***	***θ = 0.5***	***θ = 2***	***θ = 10***	***θ = 0.01***	***θ = 0.5***	***θ = 2***	***θ = 10***	***θ = 0.01***	***θ = 0.5***	***θ = 2***	***θ = 10***
SELF	GAM	0.003	0.049	0.225	1.117	0.000	0.011	0.034	0.189	0.000	0.005	0.021	0.123	0.000	0.003	0.009	0.049
	TRN	0.010	0.056	0.222	1.150	0.001	0.010	0.042	0.175	0.000	0.005	0.020	0.098	0.000	0.027	0.011	0.052
	FEM	0.001	0.124	0.648	12.319	0.000	0.045	0.527	8.075	0.000	0.033	0.413	11.328	0.000	0.033	0.296	9.699
	REM	2.188	1.635	0.630	7.723	2.191	2.163	2.204	0.420	2.155	2.840	4.149	1.782	2.288	3.991	8.179	11.879
	ZIP	0.202	0.088	0.314	4.663	0.105	0.054	0.246	4.186	0.112	0.055	0.239	4.027	0.108	0.055	0.238	4.059
SSELF	GAM	0.001	0.049	0.223	1.110	0.000	0.011	0.035	0.190	0.000	0.005	0.021	0.122	0.000	0.003	0.009	0.049
	TRN	0.004	0.047	0.213	1.145	0.000	0.010	0.042	0.175	0.000	0.005	0.020	0.098	0.000	0.027	0.011	0.052
	FEM	0.000	0.060	0.364	3.831	0.000	0.018	0.156	2.118	0.000	0.011	0.123	2.810	0.000	0.009	0.087	2.549
	REM	0.113	0.504	0.302	12.661	0.155	0.645	1.061	0.616	0.231	0.871	2.049	1.258	0.299	1.255	3.925	8.173
	ZIP	0.065	0.059	0.480	8.416	0.027	0.042	0.363	7.792	0.022	0.043	0.353	7.453	0.025	0.043	0.343	7.720
LELF	GAM	0.003	0.046	0.210	1.119	0.000	0.011	0.034	0.196	0.000	0.005	0.021	0.120	0.000	0.003	0.009	0.049
	TRN	0.009	0.051	0.204	1.153	0.001	0.010	0.042	0.181	0.000	0.005	0.020	0.098	0.000	0.027	0.011	0.052
	FEM	0.001	0.062	0.207	14.133	0.000	0.023	0.095	13.055	0.000	0.013	0.053	12.157	0.000	0.011	0.048	11.648
	REM	0.923	0.660	0.216	20.634	0.906	0.782	0.766	1.842	0.913	1.095	1.325	0.453	0.840	1.371	2.185	2.596
	ZIP	0.165	0.068	0.435	9.267	0.085	0.050	0.333	8.230	0.090	0.051	0.319	7.942	0.086	0.051	0.315	7.980

GELF	GAM	0.001	0.047	0.223	1.115	0.000	0.011	0.035	0.190	0.000	0.005	0.021	0.122	0.000	0.003	0.009	0.049
	TRN	0.002	0.044	0.210	1.144	0.000	0.009	0.042	0.176	0.000	0.005	0.020	0.098	0.000	0.027	0.011	0.052
	FEM	0.000	0.052	0.321	3.273	0.000	0.014	0.134	2.654	0.000	0.009	0.114	2.000	0.000	0.007	0.091	1.617
	REM	0.001	0.147	0.185	15.020	0.001	0.203	0.599	0.708	0.002	0.254	1.069	1.007	0.002	0.327	1.813	6.418
	ZIP	0.021	0.046	0.580	11.075	0.005	0.036	0.428	9.673	0.003	0.038	0.409	9.533	0.002	0.037	0.402	9.533

4. DATA ANALYSES

Count models are extensively used in many applications, and the present work has identified three distinctive data sets; (I) Singapore Auto data extracted from insurance Data, an R package that provides automobile accident frequency; (II) number of deaths of women on daily basis derived from Bohning [4]; (III) collision losses from private passenger United Kingdom automobile insurance policies derived from McCullagh and Nelder [2].

The size of the above three data sets are quite varying as to be 7483, 366, and 129 and interestingly the low counts of 0's and 1's is also highly dispersed as 99.5%, 1.4% and 8.53% respectively. Further, the maximum frequency of II is 14 as compared to 434 in III; data set III is quite inquisitive in terms of frequency variation such as 20.16% are less than 11 (excluding 0s and 1s), 48.84% are between 11 and 100 and 22.48% values are more than 100. Figure 1 depicts these data distributions.

Table 2 presents the empirical results of the procedures comprising twenty schemes of five priors and four loss functions. This includes a bootstrap (500 samples of original size) estimation of parameter and corresponding standard error of the estimators. When n is large (data set I), both GAM and TRN have less standard errors (SE) though bootstrap estimates differ but not in a larger extent. Their original estimates are stable across all four loss functions. FEM and REM also not much varied but a slight higher point estimates are visible in ZIP model but with lesser SE.

In the case of data set II which has a moderate n, TRN is quite consistent across four loss functions in terms of its SE but estimates are similar with GAM. Similar pattern of SE is visible among FEM and REM models whereas estimates due to the latter model are quite comparable with other priors, which may not be the case for FEM. A compromising estimates and SE are due to ZIP that lies between the least (TRN) and GAM that has higher estimated values.

A more dispersed but less sized data set (III) has revealed that three prior schemes (GAM, TRN, and ZIP) are consistent across the loss functions; ZIP has least SE but estimates are uniformly higher than GAM and TRN. FEM and REM are too far from such consistency though they have less SE compared to other models. This observation indicates a slightly less compatible with simulation findings with data set III, which may be due to its dispersion; on the other hand, other two data sets are quite comparable with the observations made through simulation results.

Nevertheless, a consolidated finding of results listed below may be helpful to compare the approaches;

- LELF has the least bootstrap error followed by SELF under Gamma prior.
- GELF and SSELF can be a choice other than the usual SELF in the case of Truncated Normal prior.
- LELF and SSELF can be an alternate loss functions to SELF while Gamma or ZIP priors are used.
- In case of GELF,TRN has smaller error followed by ZIP and Gamma
- FEM and REM have different bootstrap error when loss functions differ.

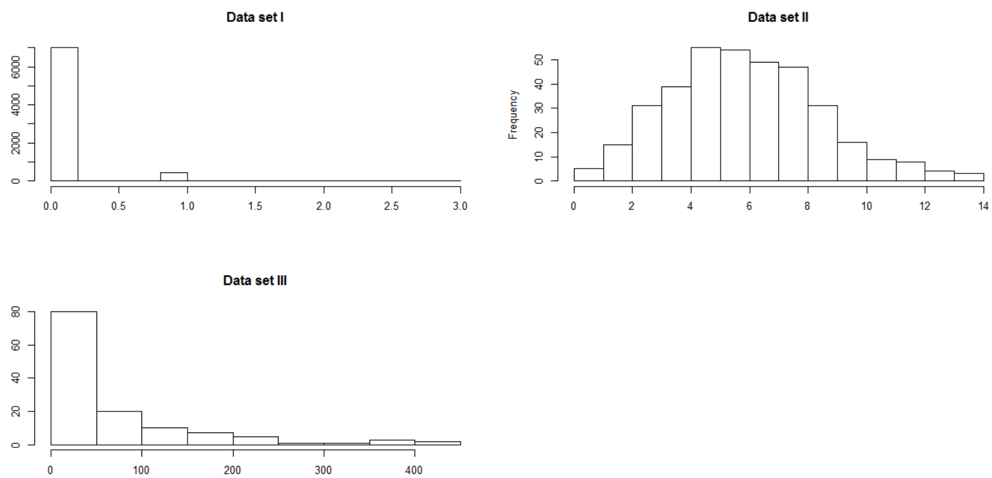

Figure 1. Histogram of three data sets considered for the data analysis of Bayesian model with five priors and four loss functions

TABLE 2. POINT ESTIMATES OF ORIGINAL (EST) AND BOOTSTRAP (B.EST) SAMPLES TOGETHER WITH BOOTSTRAP STANDARD ERRORS (B.SE) FOR THE THREE DATA SETS. RESULTS ARE PRESENTED FOR FOUR DIFFERENT LOSS FUNCTIONS

Data	**GAM**			**TRN**			**REM**			**FEM**			**ZIP**		
	Est	***B.Est***	***B.SE***	***Est***	***B.Est***	***B.SE***	***Est***	***B.Est***	***B.SE***	***Est***	***B.Est***	***B.SE***	***Est***	***B.Est***	***B.SE***
SELF															
I	0.07	0.12	0.05	0.07	0.16	0.09	0.09	0.09	0.09	0.05	0.04	0.05	0.14	0.32	0.05
II	6.37	6.42	0.49	4.99	4.73	0.06	8.39	7.94	0.73	6.17	6.10	0.54	6.38	6.16	0.37
III	72.79	75.24	25.50	72.78	71.62	6.65	94.75	116.51	48.75	27.82	39.19	14.97	75.76	71.87	29.01
SSELF															
I	0.06	0.01	0.01	0.07	0.03	0.06	0.0004	0.001	0.001	0.05	0.01	0.01	0.12	0.01	0.01
II	6.36	6.03	0.49	4.99	4.66	0.09	0.08	0.54	0.23	6.16	4.92	0.54	6.37	5.67	0.37
III	72.77	75.11	25.50	72.75	71.59	6.65	10.49	49.99	18.26	25.73	29.48	13.16	75.75	71.71	29.01
LELF															
I	0.07	0.11	0.05	0.07	0.15	0.09	0.08	0.05	0.06	0.05	0.04	0.05	0.14	0.22	0.04
II	6.33	5.48	0.42	4.99	4.47	0.15	2.26	1.59	0.11	6.11	4.04	0.42	6.34	4.98	0.29
III	71.84	70.53	24.29	71.87	70.67	6.66	5.13	8.51	3.59	19.79	9.51	3.82	74.75	66.31	27.19
GELF															
I	0.07	0.01	0.01	0.07	0.03	0.06	0.01	0.002	0.002	0.05	0.01	0.01	0.13	0.01	0.02
II	6.36	6.25	0.49	4.99	4.71	0.07	2.49	2.61	0.45	6.16	5.60	0.53	6.37	5.95	0.37
III	72.78	75.01	25.50	72.77	71.60	6.65	34.04	21.95	11.15	26.33	22.62	11.59	75.74	71.59	28.05

5. CONCLUSION

The study on the estimation of Poisson mean with emphasis on lower values of θ including zeroes is the crux of the study as rare events constitute integral part of many real time situations. The study has exploited frequentist criterion for evaluating four competing Bayesian estimates under five priors. The study has brought out the features of different Bayesian estimators and the need for a careful choice of unknowns that are part of the loss functions.

Choice of a loss function can be beyond the predominant squared error loss and associated point estimate for Poisson parameter. For small n and θ, GELF may be appropriate with possible priors could be Gamma, Truncated normal or Zero inflated Poisson whereas SSELF is an option with the same prior distribution for small samples and for any θ. GELF and LELF may be suitable with GAM and TRN priors and for a moderate sized sample (say 50 to 100) and θ is small. However, as θ increases, SELF with TRN could be chosen for moderate n. GELF with ZIP is considered to be a cautious choice for all θ; whereas FEM and REM require more careful and in-depth study even when n is large.

The comparison has pointed out that with a proper choice of loss function is more important in obtaining better estimates of parameter especially in the case of modeling rare events using Poisson distribution. The study through frequentist evaluations has emphasized to stepping away from the usual squared error loss function and considering other loss functions in a decision making process under Bayesian modeling. However, a possible limitation could be in choosing more pragmatic priors that may largely depend on the problem at hand and the researchers' perspective of the same. This work also provides few prior specifications other than conjugate model to work once a decision on appropriate loss function has been made with regard to the given problem environment.

ACKNOWLEDGMENT

The authors would like to thank the editor and an anonymous referee who kindly reviewed the earlier version of this manuscript and provided valuable suggestions and comments.

REFERENCES

[1] Johnson, N.L. & Kotz, S. (1969). Distribution in Statistics. Discrete Distributions, Willey/Houghton-Mifflin.

[2] McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models, no. 37 in Monograph on Statistics and Applied Probability.

[3] Ashley, R. (1990). Shrinkage Estimation with General Loss Functions: An Application of Stochastic Dominance Theory. International Economic Review, 31, 301-313.

[4] Bohning, D. (1994). A note on a test for Poisson over dispersion.Biometrika, 81(2), 418-419.

[5] Calabria, R., & Pulcini, G. (1996). Point estimation under asymmetric loss function for left truncated exp. Samples. Communications in Statistics (Theory and Methods), 25 (3), 585-600.

[6] Lehmann, E.L. & Casella, G. (1998). *Theory of point estimation*. Springer (India) Private Limited, New Delhi.

[7] Brown, L.D., Cai, T.T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. Statistical Science, 16, 101 – 133.

[8] Casella, G. & Berger, R.L. (2002). *Statistical Inference*, Duxbury: Belmont: CA.

[9] Hardy, M. (2003). An Illuminating Counter example. American Mathematical Monthl, 110 (3), 234-238.

[10] Sweeting, J.M., Sutton, J.A., 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.

[11] Jianhui Ning, J. & Xie, M. (2007). Minimax Invariant Estimator of Continuous Distribution Function Under LINEX Loss. JrlSystSci & Complexity. 20, 119-126.

[12] Graham Elliot, Ivana Komunjer, Allan Timmermann. (2008). Biases in Microeconomic Forecasts: Irrationality or Asymmetric loss? Journal of the European Economic Association, 6 (1), 122-157.

[13] Marcus Alexander & Nicholas A. Christakis. (2008). Bias and asymmetric loss in expert forecast; A study of physician prognostic behaviour with respect to patient survival. Journal of Health Economics, 27, 1095-1108.

[14] Subbiah, M., Kumar, B., Srinivasan, M.R. (2008). Bayesian approach to multicentre sparse data. Communications in Statistics - Simulation and Computation, 37, 687 – 696.

[15] Tuyl, F., Gerlach, R., Mengersen, K.A. (2008). Comparison of Bayes-Laplace, Jeffreys and Other Priors: The case of zero events. The American Statistician, 62, 40 – 44.

[16] Aretz, K. Bartram, SM. Pope, PF. (2011) Asymmetric loss function and the rationality of expected stock returns, International Journal of Forecasting, Volume 27, Issue 2: 413-437.

[17] Efron, B. (2011). The bootstrap and Markov chain Monte Carlo. J Biopharm, Stat, 2, 1052–1062.

[18] Lee, C.H. & Hwang, L.H. (2011). Asymptotic optimal estimation of Mean under LINEX loss function. Communications in Statistics – Theory and Methods, 40, 4308-4321.

[19] Efron, B. (2012). Bayesian inference and the parametric bootstrap. Ann. Applied Statistics, 6, 1971-1997.

[20] Hwang, L.H. & Lee, C.H. (2013). Bayes sequential estimation for a Poisson process under a LINEX loss function. Statistics, 47, 672-687.

[21] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URLhttp://www.R-project.org/.

[22] Efron, B. (2014). Frequentist accuracy of Bayesian estimates. Journal of the Royal Statistical Society: Series B (Statistical Methodology), doi:10.1111/rssb.12080

[23] Shakil, M. Kibria, B.M.G. and Singh, J.N. (2016) . Characterization of a New Class of Generalized Pearson Distribution by Truncated Moment. Int. J. Comp. Theo. Stat. 3, No. 2 pp 91-100.

International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 4, No. 2 (Nov-2017)

E-mail address: renuka2327@yahool.com, sisufive@gmail.com, mrsvasan59@yahoo.com