Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v13n2p580

Estimation of population mean for logarithmic observation under Poisson distributed study and auxiliary variates

By Khare, Singh, Sharma

Published: 14 October 2020

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribuzione - Non commerciale - Non opere derivate 3.0 Italia License.
For more information see:
http://creativecommons.org/licenses/by-nc-nd/3.0/it/

Electronic Journal of Applied Statistical Analysis
Vol. 13, Issue 02, October 2020, 580-588
DOI: 10.1285/i20705948v13n2p580

Estimation of population mean for logarithmic observation under Poisson distributed study and auxiliary variates

Supriya Khare[a], Rajesh Singh[b], and Prayas Sharm*[c]

[a,c]*Department of Statistics,Institute of Science,Banaras Hindu University,Varanasi-221005*
[c]*School of Business, University of Petroleum and Energy Studies,Dehradun,India*

Published: 14 October 2020

Use of auxiliary information is always suggested at the planning and estimation stage to make the estimators perform more eggiciently.Estimation using auxiliary information is common in sampling literature but using distribution of study and auxiliary information at the estimation stage is uncommon and found useful especially when dealing with rare variable. This study utilizes the auxiliary information and Poisson distributed variates for proposing the log-type estimator and another generalized estimator for finite population mean under simple random sampling without replacement. The Mean Square Error expressions of the proposed estimators are obtained.It is revealed from empirical (point estimation and interval estimation) & theoretical study that use of log type estimators along with suitable auxiliary information for Poisson distributed variates excels the performance of estimators in terms of efficiency.

keywords: Auxiliary variable, Mean Square Error, Poisson distribution, Study variable, Log type estimators.

1 Introduction

In sampling survey, it is widely acknowledged that efficiency of estimators defined for population parameters can be proliferated by using an auxiliary variable highly correlated with study variable. Many authors have suggested estimators for population mean

*Corresponding author: prayassharma02@gmail.com

©Università del Salento
ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

and variance (see Gupta and Shabbir , 2007,Singh et al., 2016,Bhushan, 2012, Kadilar and Cingi, 2004,Al-Jararha, 2018) under simple random sampling without replacement irrespective of the distribution followed by study and auxiliary variable. Poisson distribution introduced by Simone Denis Poison in 1837 has strong theoretical background and very wide spectrum of practical applications. It is found to be highly useful for natural population to represent the probability of number of rare events in day to day life. It can be utilized for estimating population parameters for rare events in case the population is poison distributed, for instance consider Koyuncu and Ozel (2013),Ozel and Inal (2008),Ozel (2011a) andOzel (2011b),Sharma et. al (2017)). The list of applications of Poisson distribution is very long, to name a few we consider here: The number of mutations on a given strand of DNA per unit time, the number of network failures per day, the number of birth, deaths, divorces, marriages, suicides and homicides over a given period of time, the number of errors in a manuscript, the number of file server virus infection at a data centre during a day etc. In this paper, we utilize earthquake data for our study purpose as earthquake is a rare event usually following Poisson distribution. Aftershocks constitute the greatest proportion of shocks in an earthquake and to understand the whole cycle of seismic activity, it is desired to study aftershocks effectively. Estimating the number of aftershocks in seismology has received much attention in recent literature. (see Ozel , 2011b) Ratio type estimators give us information about the number of aftershocks in a specified region (see Ozel , 2013). Here, we use log type estimators for estimating population mean for number of earthquake (which is a rare activity) data that follows Poisson distribution.

The paper is organized as follows. Section 2 contains the Notations used in the paper alongwith the Sampling Methodology used.In section 3, some known estimators in literature with their properties are discussed. Section 4 formulates the suggested estimators.In Section 5, Efficiency comparison is made for the suggested estimators with those discussed in section 3.Detailed Empirical study in the support of suggested Estimators has been conducted and discussed in Section 6.Finally some conclusions are given in Section 7.

2 Notations and Methodology

Consider a population U of size N consisting of identifiable and distinct units. Let y and x be study and auxiliary variables associated with each $U_i\,(i = 1, 2, \ldots, N)$of the population. Let us draw a simple random sample of size n using Simple Random Sampling Without Replacement (SRSWOR) from the population. The parent population is assumed to follow a Poisson distribution. It is known in advance that the nature of sampling distribution depends upon a population from which the observations are drawn, so the samples also follow a Poisson distribution. Let the auxiliary variable x has Poisson distribution with parameter $\mu_1 > 0$,then the values of parameters for this distribution are given by $\overline{X} = \mu_1,\ S_X = \sqrt{\mu_{1,}}\ C_X = \frac{S_X}{\overline{X}} = \frac{1}{\sqrt{\mu_{1,}}}$ respectively. Let the study variable y has Poisson distribution with parameter $\mu_2 > 0$,then the values of parameters for this distribution are given in a similar manner as $\overline{Y} = \mu_2,\ S_Y = \sqrt{\mu_{2,}}\ C_Y = \frac{S_Y}{\overline{Y}} = \frac{1}{\sqrt{\mu_{2,}}}$

respectively. Further, to obtain the correlation coefficient between y and x we use trivariate reduction method (see Lai , 1995), a method of constructing bivariate Poisson distribution. Here, we are required to obtain a pair of dependent Poisson distributed variables from three independent Poisson distributed population. For this, let us assume m, w and z are independently distributed Poisson random variables from a Poisson distributed Population. Corresponding bivariate distribution for y and x can be constructed as,$x_j = m_j + z_j$ *and* $y_j = w_j + z_j$, (j=1,2,. . . .n) where, the parameters of m, w and z are γ_1, γ_2 *and* $\gamma_3$respectively. Now, the correlation coefficient between (y, x)can be defined as,

$$\rho_{y_{po}x_{po}} = \frac{Cov(y,x)}{S_X S_Y} = \frac{E(x,y) - E(x)E(y)}{S_X S_Y} = \frac{\gamma_3}{\sqrt{(\gamma_1+\gamma_3)(\gamma_2+\gamma_3)}}$$

Where, $E\left(x, y\right) = \left[(\gamma_1 + \gamma_3)(\gamma_2 + \gamma_3) + \gamma_3\right],\ E\left(x\right) = (\gamma_1 + \gamma_3),\ E\left(y\right) = (\gamma_2 + \gamma_3),$
$S_X = \sqrt{\mu_1} = (\gamma_1 + \gamma_3)$ *and* $S_Y = \sqrt{\mu_2} = (\gamma_2 + \gamma_3)$.Here, $\rho_{y_{po}x_{po}}$ is strictly positive since all of γ_1, γ_2 *and* $\gamma_3 are > 0$. To evaluate the bias and mean square error, we define the formula,
$\overline{y}_{po} = \overline{Y}\left(1 + e_0\right) and\ \overline{x}_{po} = \overline{X}\left(1 + e_1\right)$ such that

$$E\left(e_0\right) = E\left(e_1\right) = 0,\ E\left(e_0^2\right) = \frac{1}{n\mu_2} = \frac{1}{n\left(\gamma_2 + \gamma_3\right)},\ E\left(e_1^2\right) = \frac{1}{n\mu_1} = \frac{1}{n\left(\gamma_1 + \gamma_3\right)} and$$

$$E\left(e_0 e_1\right) = \frac{Cov(y_{po}, x_{po})}{\overline{Y}\overline{X}} = \frac{\rho}{n\sqrt{\mu_1\mu_2}} = \frac{\rho}{n\sqrt{\left(\gamma_1 + \gamma_3\right)\left(\gamma_2 + \gamma_3\right))}}$$

3 Estimators in Literature

Under this sampling design we define usual mean per unit estimator as, $p_1 = \overline{y}_{po}$,

$$Bias\ \ (p_1) = 0\ and\ MSE\left(p_1\right) = \overline{Y}^2 \frac{1}{n\left(\gamma_2 + \gamma_3\right)}$$

where, $\overline{y}_{po} = \sum_{i=1}^{n} y_i/n$ is the sample mean of study variable from a Poisson distribution. Similarly, we define Classical ratio and product estimator respectively as,

$$p_2 = \overline{y}_{po}\frac{\overline{X}}{\overline{x}_{po}}\ ,\ p_3 = \overline{y}_{po}\frac{\overline{x}_{po}}{\overline{X}},$$

$$Bias\left(p_2\right) = \frac{\overline{Y}}{n}\left[\frac{1}{(\gamma_1 + \gamma_3)} - \frac{\rho}{\sqrt{\left(\gamma_1 + \gamma_3\right)\left(\gamma_2 + \gamma_3\right))}}\right]$$

$$MSE\left(p_2\right) = \frac{1}{n}\left[\frac{\overline{Y}^2}{\overline{X}^2}\left(\gamma_1 + \gamma_3\right) + \left(\gamma_2 + \gamma_3\right) - 2\rho\frac{\overline{Y}}{\overline{X}}\sqrt{\left(\gamma_1 + \gamma_3\right)\left(\gamma_2 + \gamma_3\right))}\right]$$

$$Bias\left(p_3\right) = \frac{\overline{Y}}{n}\left[\frac{\rho}{\sqrt{\left(\gamma_1 + \gamma_3\right)\left(\gamma_2 + \gamma_3\right))}}\right]$$

$$MSE\left(p_3\right)=\frac{1}{n}\left[\frac{\overline{Y}^2}{\overline{X}^2}\left(\gamma_1+\gamma_3\right)+\left(\gamma_2+\gamma_3\right)+2\rho\frac{\overline{Y}}{\overline{X}}\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)}\right]$$

Where, $\overline{y}_{po}=\sum_{i=1}^{n}y_i/n$ *and* $\overline{x}_{po}=\sum_{i=1}^{n}x_i/n$ are the sample means of study and auxiliary variable respectively.
Koyuncu and Ozel (2013), defined exponential ratio and product type estimators as follows,

$$p_4=\overline{y}_{po}exp\left(\frac{\overline{X}-\overline{x}_{po}}{\overline{X}+\overline{x}_{po}}\right)$$

$$Bias\left(p_4\right)=\frac{\overline{Y}}{8n}\left[\frac{3}{\left(\gamma_1+\gamma_3\right)}-\frac{4\rho}{\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right))}}\right]\ and$$

$$MSE\left(p_4\right)=\frac{\overline{Y}^2}{n}\left[\frac{1}{\left(\gamma_2+\gamma_3\right)}+\frac{1}{4(\gamma_1+\gamma_3)}-\frac{\rho}{\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)}}\right]$$

$$p_5=\overline{y}_{po}exp\left(\frac{\overline{x}_{po}-\overline{X}}{\overline{x}_{po}+\overline{X}}\right)$$

$$Bias\left(p_5\right)=\frac{\overline{Y}}{8n}\left[\frac{4\rho}{\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right))}}-\frac{1}{\left(\gamma_1+\gamma_3\right)}\right]$$

$$MSE\left(p_5\right)=\frac{\overline{Y}^2}{n}\left[\frac{1}{\left(\gamma_2+\gamma_3\right)}+\frac{1}{4(\gamma_1+\gamma_3)}+\frac{\rho}{\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)}}\right]$$

? proposed an estimator for population mean in simple random sampling as,

$$p_6=\overline{y}_{po}exp\left(\frac{\alpha\left(\overline{X}-\overline{x}_{po}\right)}{\overline{X}+\overline{x}_{po}}\right)$$

$$Bias\left(p_6\right)=\alpha\frac{\overline{Y}}{8n}\left[\frac{(\alpha+2)}{\left(\gamma_1+\gamma_3\right)}-\frac{4\rho}{\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right))}}\right]$$

$$MSE\left(p_6\right)=\frac{\overline{Y}^2}{n}\left[1-\rho^2\right]$$

$$under\ \alpha_{opt}=\frac{2\rho\left(\gamma_1+\gamma_3\right)}{\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)}}$$

4 Suggested Estimators

Mishra et. al (2017) suggested a log type estimator under simple random sampling. Moving in this direction we suggest estimators for Poisson distribution under simple random sampling without replacement scheme as follows,

$$t_1=\overline{y}_{po}+\beta log\left(\frac{\overline{x}_{po}}{\overline{X}}\right) \tag{1}$$

Expanding Eq.(1) up to the terms of first order of approximation and taking Expectation, we obtain,

$$Bias\left(t_1\right) = \frac{-\beta}{2n\left(\gamma_1+\gamma_3\right)} \; and \quad MSE_{min}\left(t_1\right) = \frac{\overline{Y}^2}{n}\left[1-\rho^2\right] \tag{2}$$

Under,$\beta_{opt} = -\rho\sqrt{\left(\left(\gamma_1+\gamma_3\right)\right)\left(\left(\gamma_2+\gamma_3\right)\right)}$

$$t_2 = \overline{y}_{po}\left(1+w_1\right) + w_2 log\left(\frac{\overline{x}_{po}}{\overline{X}}\right) \tag{3}$$

Expanding Eq.(3) and retaining terms up to first order of approximation and taking Expectation, we get,

$$Bias\left(t_2\right) = \overline{Y}w_1 - \frac{w_2}{2n\left(\gamma_1+\gamma_3\right)}$$

$$MSE_{min}\left(t_2\right) = C + \frac{BC^2+AD^2-2CDE}{E^2-AB} \tag{4}$$

Under $w_{1opt} = \frac{AD-EC}{E^2-AB}$ $\quad$ $and\ w_{2opt} = \frac{BC-DE}{E^2-AB}$
Where, $A = \overline{Y}^2\left(1+\frac{1}{n(\gamma_2+\gamma_3)}\right)$, $\quad B = \frac{1}{n(\gamma_1+\gamma_3)}$, $\quad C = \overline{Y}^2\left(\frac{1}{n(\gamma_2+\gamma_3)}\right)$,

$$D = \overline{Y}\frac{\rho}{n\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)\right)}}, \quad E = \overline{Y}\left(\frac{\rho}{n\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)\right)}} - \frac{2}{n\left(\gamma_1+\gamma_3\right)}\right)$$

$$t_3 = \overline{y}_{po}\left(1+w_3\right) + \frac{\overline{X}}{\overline{x}_{po}}w_4 \tag{5}$$

Expanding Eq.(5) up to the terms of first order of approximation and taking Expectation, we obtain,

$$Bias\left(t_3\right) = \overline{Y}w_3 + w_4 + \frac{w_4}{n\left(\gamma_1+\gamma_3\right)}$$

$$MSE_{min}\left(t_3\right) = C_1 + \frac{B_1C_1^2+A_1D_1^2-2C_1D_1E_1}{E_1^2-A_1B_1} \tag{6}$$

Under $w_{3opt} = \frac{A_1D_1-C_1E_1}{E_1^2-A_1B_1}$ $\quad$ $and\ w_{4opt} = \frac{B_1C_1-D_1E_1}{E_1^2-A_1B_1}$

$$where,\ \ A_1 = \overline{Y}^2\left(1+\frac{1}{n\left(\gamma_2+\gamma_3\right)}\right), \quad B_1 = \left(1+\frac{3}{n\left(\gamma_1+\gamma_3\right)}\right)$$

$$C_1 = \overline{Y}^2\left(\frac{1}{n\left(\gamma_2+\gamma_3\right)}\right), \quad D_1 = -\overline{Y}\frac{\rho}{n\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)\right)}}$$

$$E_1 = \overline{Y}\left(1-\frac{\rho}{n\sqrt{\left(\gamma_1+\gamma_3\right)\left(\gamma_2+\gamma_3\right)\right)}} + \frac{1}{n\left(\gamma_1+\gamma_3\right)}\right)$$

5 Efficiency Comparison

Here, we compare the efficiency of proposed estimator t_3 with estimators proposed by Koyuncu and Ozel (2013) and Sharma et al. (2015).

$$MSE\,(p_4) - MSE_{min}\,(t_3) \geq 0$$

$$\frac{\overline{Y}^2}{n}\left[\frac{1}{(\gamma_2+\gamma_3)}+\frac{1}{4(\gamma_1+\gamma_3)}-\frac{\rho}{\sqrt{(\gamma_1+\gamma_3)\,(\gamma_2+\gamma_3)}}\right] \geq C_1 + \frac{B_1C_1^2+A_1D_1^2-2C_1D_1E_1}{E_1^2-A_1B_1} \tag{7}$$

$$MSE\,(p_5) - MSE_{min}\,(t_3) \geq 0$$

$$\frac{\overline{Y}^2}{n}\left[\frac{1}{(\gamma_2+\gamma_3)}+\frac{1}{4(\gamma_1+\gamma_3)}+\frac{\rho}{\sqrt{(\gamma_1+\gamma_3)\,(\gamma_2+\gamma_3)}}\right] \geq C_1 + \frac{B_1C_1^2+A_1D_1^2-2C_1D_1E_1}{E_1^2-A_1B_1} \tag{8}$$

$$MSE\,(p_6) - MSE_{min}\,(t_3) \geq 0$$

$$\frac{\overline{Y}^2}{n}\left[1-\rho^2\right] \geq C_1 + \frac{B_1C_1^2+A_1D_1^2-2C_1D_1E_1}{E_1^2-A_1B_1} \tag{9}$$

$$MSE\,(t_2) - MSE_{min}\,(t_3) \geq 0$$

$$C + \frac{BC^2+AD^2-2CDE}{E^2-AB} \geq C_1 + \frac{B_1C_1^2+A_1D_1^2-2C_1D_1E_1}{E_1^2-A_1B_1} \tag{10}$$

Conditions mentioned in Eq. (7) to Eq. (10) holds true and it can be observed that proposed estimator is more efficient than estimators proposed by Koyuncu and Ozel(2013) and Sharma et al. (2015).Also, it is most efficient among proposed estimators.

6 Empirical Study

Source: Sharma et. al (2017)

Here, we utilize the earthquake data of Turkey for the numerical comparison of the proposed estimators and other existing estimators in simple random sampling without replacement. The data is obtained from the data base of Kandilli Observatory, Turkey. Earthquake is an unavoidable natural disaster for Turkey since a significant portion of turkey is subject to frequent destructive mainshocks, their foreshocks and aftershocks sequence.Consider 109 mainshocks that occurred in 1900 and 2011 having the surface wave magnitudes Ms≥0.5, foreshocks within 5days week with Ms≥0.3 and aftershocks within 1 month with Ms≥ 0.4.Here, the number of aftershocks is a study variable and the number of foreshocks is an auxiliary variable. The population consists of the destructive earthquakes. In the population data set the number of aftershocks is a study variable and the number of foreshocks is an auxiliary variable.In order to obtain the distribution of these variables, we fit the poisson distribution to earthquake dataset. To obtain the ρ_{xpoypo} for the Poisson distributed data, Turkey is divided into three main neotectonic domains based on the neotectonic zones of Turkey. The foreshocks in Turkey are separated according to these neotectonic zones. In this way, the parameters 1, 2 and 3 are obtained. According to the goodness of the fit test, it is obvious that the Poisson distribution fits the number of shocks for Region 1 with parameter 1 =4.1813 _2 = 0.048, p value = 0.043_ and 2 = 8.104 _2 = 0.014, p value = 0.032_for Region 2, and 3 = 2.112 _2 = 0.013, p value = 0.025_ for Region 3 (consider n=20).Then the correlation between the study variable and auxiliary variable is positive and equal to 0.712.The MSE and PRE values for proposed and existing estimators along with the length of confidence interval for MSE are summarized in the table 1.

Table 1: MSE and PRE of Proposed Estimators

S.No.	Estimators	MSE	PRE	C.I for MSE
1	p_1	0.511	100.000	1.912
2	p_2	0.413	123.608	1.771
3	p_3	2.266	22.534	3.462
4	p_4	0.255	200.531	1.500
5	p_5	1.181	43.234	2.641
6	p_6	0.252	202.817	1.494
7	t_1	0.252	202.817	1.494
8	t_2	0.249	204.640	1.490
9	t_3	0.090	566.266	1.099

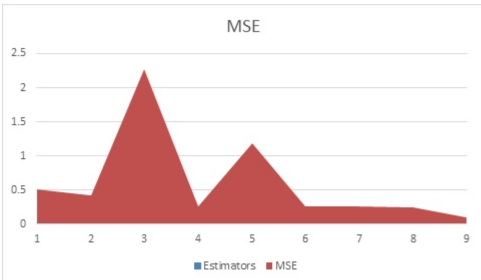

Figure 1: MSE of Estimators

7 Concluding Remarks

From the statistics given in table 1, it can be depicted that the proposed estimators $\boldsymbol{t_1}, \boldsymbol{t_2}$ ***and*** $\boldsymbol{t_3}$ are more efficient than $\boldsymbol{p_1, p_2, p_3, p_4, p_5}$ ***and*** $\boldsymbol{p_6}$ whereas, $\boldsymbol{p_6}$ is equally effi-

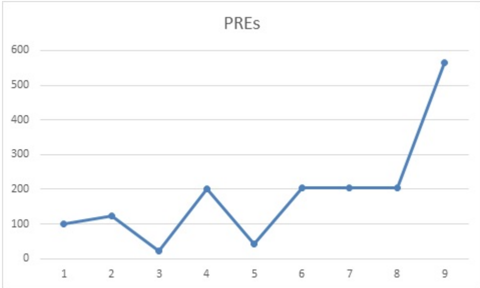

Figure 2: PRE of Estimators

cient to $\boldsymbol{t_1}$. It can also be observed from table 1 that proposed estimators possess lesser MSE as compared to other estimators making it suitable for application in the estimation of population mean under logarithmic approach for a Poisson distributed variates. Length of the Confidence Interval for MSE values given in Table 1 depict the desirable trend that is, the proposed estimators have lesser length of confidence interval as compared to other estimators, $\boldsymbol{t_3}$ being minimum among all. Overall the theoretical results based on efficiency comparison and numerical results based on Table 1, figure 1 and figure 2 we can conclude that the proposed estimator performs better as compared to other estimators considered here therefore, suggested for practical application as shown here in the case of earthquake data and are applicable for other similar situations as well.

References

Al-Jararha, J. (2018). Unbiased Estimator of Finite Population Variance Based on a Ratio Type Estimator. *Electronic Journal of Applied Statistical Analysis*, 11(1), 196-207.

Bhushan, S. (2012). Some Efficient Sampling Strategies based on ratio type estimator *Electronic Journal of Applied Statistical Analysis*, 5(1), 196-207.

Gupta, S. and Shabbir, J.(2007). On the use of transformed auxiliary variable in estimating Population mean. *J. Stat. Plan. Inference*, 137, 1606-1611.

Kadilar, C. and Cingi, H. (2004). Ratio Estimators in Simple Random Sampling. *Applied Mathematics and Computation*, 3(2), 745-755.

Koyuncu, N., Ozel, G. (2013). Exponential estimators using characteristics of poisson distribution: an application to earthquake data. *11th International Conference of Numerical Analysis andApplied Mathematics, API Conference Proceedings*, 1558, 1470–1473.

Lai, C. D. (1995). Construction of bivariate distribution by a generalized trivariate reduction technique. *Statistics and Probability Letters*, 25, 265–270.

Mishra, P., Adichwal, N. K. & Singh, R.,(2017). A new- log Product type estimator using auxiliary information. *Journal of Scientific Research, BHU, Varanasi*, 61,179–183.

Ozel G, Inal C (2008). The Probability function of the compound Poisson Process and an application aftershock sequences in Turkey. *Environmetrics*, 19(1), 79–85.

Ozel, G. (2011a). A bivariate compound Poisson model for the occurrence of foreshock and aftershock sequences in Turkey. *Environmetrics*, 22(7), 847–856.

Ozel, G. (2011b). On certain properties of a class of bivariate compound poisson distributions and an application to earthquake data. *Revista Colombiana de Estadistica*, 34(3), 545–566.

Ozel, G. (2013). Ratio Estimators using characteristics of Poisson Distribution with Application to Earthquake Data. *7^{th} International Days of Statistics and Economics, Prague, September*, 19–21.

Sharma, P., Verma, H. K., Adichwal, N. K., & Singh, R. (2017). Generalized Estimators Using Characteristics of Poisson Distribution. *International Journal of Applied and Computational Mathematics*, 3, v745–755.

Singh, R., Verma, H. K. and Sharma, P. (2016). Efficient Estimators of finite Population Mean Using Predictive Estimation in Simple Random Sampling. *Journal of Statistics*, 23(1), 1-11.