Volume 68, Issue 2, 2024
Journal of Scientific Research
of
The Banaras Hindu University

Estimation of Population Mean Using Auxiliary Information from the Poisson Population

Rajesh Singh [1], Sakshi Rai*[2]

[1] Department of Statistics, Banaras Hindu University, Varanasi -221005, rsinghstat@gmail.com

*[2] Department of Statistics, Pt. Prithi Nath (P.G.) College, Kanpur- 208001, sakshi.rai@bhu.ac.in

Abstract: **In this paper we have proposed class of almost unbiased estimator using single auxiliary variable in case of SRSWOR for the Poison population. It has been shown that the proposed class of estimator is unbiased. The Expressions for mean squared error (MSE) of the proposed class of estimator is derived up to first order of approximation. Empirical study is carried out to show the efficiency of the proposed estimator.**

Index Terms: **Almost unbiased estimator, Mean Squared Error, Poisson distribution, Auxiliary Variable, Efficiency**

I. INTRODUCTION

Sometimes we are concerned about event which occur rarely (such as earthquake) but has much importance to study about it. Natural population in different fields like Earthquake Seismology, Telecommunication, Astronomy, Finance and Insurance management, Optics where the number of occurrence of an event is very rare but has much opportunity to occur, Poisson distribution is best fit distribution for such types of population for further studies. In the field of sampling, auxiliary information can be utilized at planning, selection and estimation stage. Many times auxiliary information is very well available to us other than the character we want to study. Different types of cases can be set up according to the relation between auxiliary and study variable. Ratio, product and regression estimators are some examples that can be used when auxiliary information is accessible to us. When the correlation is positive, ratio estimators are used and for negative correlation, product estimators are used. Several authors including Sisodia and Dwivedi (1981), Upadhyaya and Singh (1999), Kadilar and Cingi (2004), Gupta and Shabbir (2007, 2008), Koyuncu and Kadilar (2009), Singh and Vishwakarma (2010), Shabbir and Gupta (2011) suggested different type of estimators which can be used under different situations but they have not considered the distribution of the population. However in recent times authors including Özel (2013), Koyuncu & Özel (2013), Sharma et al (2016) have suggested estimators for Poisson population. But all these aforesaid estimators are biased in nature.

A. NOTATIONS & TERMIN0OLOGIES

Consider $\Omega = \Omega_{1,}\Omega_2\ \ldots \Omega_N$ be a population of N finite units. Let Y and X be the study and auxiliary variable associated with this population. Here, we have assumed that the population follows Poisson distribution. Let us draw a sample (y_i, x_i) i =1, 2…n from population Ω using simple random sampling without replacement (SRSWOR) to estimate the population mean $\overline{Y}$.

If we assume that variables (Y, X) follows Poisson distribution with parameter (λ_1, λ_2) $\lambda_1, \lambda_2 > 0.$ Using Cochran (1940), when the correlation is positive between study and auxiliary variable we defined a usual ratio estimator in case of Poisson distributed population as:

$$P_{ratio} = \left(\frac{\bar{y}_{po}}{\bar{x}_{po}}\right)\bar{X} \tag{1.1}$$

where, $(\bar{y}_{po}, \bar{x}_{po})$ are sample means of study and auxiliary variable respectively. Here,

$$\bar{X} = \lambda_1, \quad S_x = \sqrt{\lambda_1}, \quad C_x = \frac{S_x}{\bar{X}} = \frac{1}{\sqrt{\lambda_1}}$$

$$\bar{Y} = \lambda_2, \quad S_y = \sqrt{\lambda_2}, \quad C_y = \frac{S_y}{\bar{Y}} = \frac{1}{\sqrt{\lambda_2}}$$

Lai (1995) generated a generalized trivariate reduction technique to construct bivariate distribution. Using this, let us define three independent variable u, v, w following Poisson distribution with parameter $\gamma_1, \gamma_2, \gamma_3$ respectively and setting

$$x_i = u_i + w_i$$
$$y_i = v_i + w_i, \quad i = 1, 2 \ldots n$$

then,

$$E(X) = \gamma_1 + \gamma_3, \; E(Y) = \gamma_2 + \gamma_3,$$
$$E(X, Y) = [(\gamma_1 + \gamma_3)(\gamma_2 + \gamma_3) + \gamma_3],$$
$$S_x^2 = \lambda_1 = (\gamma_1 + \gamma_3), \; S_y^2 = \lambda_2 = (\gamma_2 + \gamma_3)$$
$$\rho_{x_{po}y_{po}} = \frac{cov(x, y)}{\sqrt{S_x^2 S_y^2}} = \frac{\gamma_3}{\sqrt{(\gamma_1 + \gamma_3)(\gamma_2 + \gamma_3)}}$$

where, $\rho_{x_{po}y_{po}}$ is the coefficient of correlation between Y and X since γ_1, γ_2 *and* γ_3 are always positive, $\rho_{x_{po}y_{po}}$ is restricted to be strictly positive. The observations (y_i, x_i), i=1, 2… n are drawn from Poisson population with parameter (λ_1, λ_2) we have

$$\bar{x}_{po} = \sum_{i=1}^{n} (u_i + w_i)/n$$
$$\bar{y}_{po} = \sum_{i=1}^{n} (v_i + w_i)/n$$

The covariance of $\bar{x}_{po}$ and $\bar{y}_{po}$ is

$$cov(\bar{x}_{po}, \bar{y}_{po}) = \frac{\gamma_3}{n}$$

To find the bias and MSE we have defined

$$\epsilon_0 = \frac{\bar{y}_{po}}{\bar{Y}} - 1, \epsilon_1 = \frac{\bar{x}_{po}}{\bar{X}} - 1,$$

so that

$$E(\epsilon_0) = E(\epsilon_1) = 0$$
$$E(\epsilon_0^2) = \frac{C_y^2}{n} = \frac{S_y^2}{n\bar{Y}^2} = \frac{1}{n\lambda_2}$$
$$E(\epsilon_1^2) = \frac{C_x^2}{n} = \frac{S_x^2}{n\bar{X}^2} = \frac{1}{n\lambda_1}$$
$$E(\epsilon_0\epsilon_1) = \frac{cov(\bar{x}_{po}, \bar{y}_{po})}{\bar{Y}\bar{X}} = \frac{\rho_{x_{po}y_{po}} C_y C_x}{n} = \frac{\rho_{x_{po}y_{po}}}{n\sqrt{\lambda_1\lambda_2}}$$

Expressing P_{ratio} in terms of $\epsilon's$ we get

$$P_{ratio} = \bar{Y}(1 + \epsilon_0)(1 + \epsilon_1)^{-1}$$
$$= \bar{Y}(1 + \epsilon_0 - \epsilon_1 + \epsilon_1^2 - \epsilon_0\epsilon_1)$$
$$\text{Bias}(P_{ratio}) = E(P_{ratio}) - \bar{Y}$$
$$\text{Bias}(P_{ratio}) = \frac{\lambda_2}{n\lambda_1}\left(1 - \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)$$
$$\text{MSE}(P_{ratio}) = E(P_{ratio} - \bar{Y})^2$$
$$\text{MSE}(P_{ratio}) = \frac{\lambda_2}{n}\left(1 + \frac{\lambda_2}{\lambda_1}\left(1 - 2\rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)\right)$$

II. Estimators in Literature

Koyuncu and Ozel (2013) proposed the following ratio estimator (t_1) and product estimator (t_2) for positive and negative correlation respectively, to estimate unknown population mean as

$$t_1 = \bar{y}_{po} \exp\left((\bar{X} - \bar{x}_{po}) \big/ (\bar{X} + \bar{x}_{po})\right) \qquad (2.1)$$

Bias and MSE of the estimator t_1 is given by

$$\text{Bias}(t_1) = \frac{\lambda_2}{2n\lambda_1}\left(\frac{3}{4} - \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right) \qquad (2.2)$$

$$\text{MSE}(t_1) = \frac{\lambda_2}{n}\left[1 + \frac{\lambda_2}{\lambda_1}\left(\frac{1}{4} - \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)\right] \qquad (2.3)$$

The estimator t_2 is given as

$$t_2 = \bar{y}_{po} \exp\left(\frac{\bar{x}_{po} - \bar{X}}{\bar{x}_{po} + \bar{X}}\right) \qquad (2.4)$$

Bias and MSE of the estimator t_2 is given by

$$\text{Bias}(t_2) = \frac{\lambda_2}{2n\lambda_1}\left(\rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}} - \frac{1}{4}\right) \qquad (2.5)$$

$$\text{MSE}(t_2) = \frac{\lambda_2}{n}\left[1 + \frac{\lambda_2}{\lambda_1}\left(\frac{1}{4} + \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)\right] \qquad (2.6)$$

Özel (2011), suggested a generalized class of ratio estimators for the population mean as

$$t_3 = \frac{\bar{y}_{po}}{a\bar{x}_{po} + b}(a\bar{X} + b); \; a\bar{x}_{po} + b \neq 0 \qquad (2.7)$$

where a and b are either constants or function of known parameters of the population such as a = 1 or C_x, $\beta_1(x)$, $\beta_2(x)$, $\rho_{x_{po}y_{po}}$ etc.

Bias and MSE of the estimator t_3 is given by

$$\mathrm{Bias}(t_3) = \frac{\alpha\lambda_2}{n\lambda_1}\left(\alpha - \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right) \quad (2.8)$$

$$\mathrm{MSE}(t_3) = \frac{\lambda_2}{n}\left[1 + \frac{\alpha\lambda_2}{\lambda_1}\left(\alpha - 2\rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)\right] \quad (2.9)$$

where

$$\alpha = \frac{a\bar{X}}{a\bar{X}+b}$$

Table I. Values of a, b and consequently α used in existing estimator t_3

a	**b**	**α**
1	0	1
1	C_x	$\frac{\bar{X}}{\bar{X}+C_x}$
1	$\beta_1(x)$	$\frac{\bar{X}}{\bar{X}+\beta_1(x)}$
1	$\beta_2(x)$	$\frac{\bar{X}}{\bar{X}+\beta_2(x)}$
1	$\rho_{x_{po}y_{po}}$	$\frac{\bar{X}}{\bar{X}+\rho_{x_{po}y_{po}}}$
1	D_x	$\frac{\bar{X}}{\bar{X}+D_x}$
C_x	$\beta_1(x)$	$\frac{C_x\bar{X}}{C_x\bar{X}+\beta_1(x)}$
C_x	$\beta_2(x)$	$\frac{C_x\bar{X}}{C_x\bar{X}+\beta_2(x)}$
C_x	$\rho_{x_{po}y_{po}}$	$\frac{C_x\bar{X}}{C_x\bar{X}+\rho_{x_{po}y_{po}}}$
C_x	D_x	$\frac{C_x\bar{X}}{C_x\bar{X}+D_x}$
$\beta_1(x)$	C_x	$\frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+C_x}$
$\beta_1(x)$	$\beta_2(x)$	$\frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+\beta_2(x)}$
$\beta_1(x)$	D_x	$\frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+D_x}$
$\beta_2(x)$	C_x	$\frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+C_x}$
$\beta_2(x)$	$\beta_1(x)$	$\frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+\beta_1(x)}$
$\beta_2(x)$	$\rho_{x_{po}y_{po}}$	$\frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+\rho_{x_{po}y_{po}}}$
$\beta_2(x)$	D_x	$\frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+D_x}$
$\rho_{x_{po}y_{po}}$	C_x	$\frac{\rho_{x_{po}y_{po}}\bar{X}}{\rho_{x_{po}y_{po}}\bar{X}+C_x}$
$\rho_{x_{po}y_{po}}$	$\beta_1(x)$	$\frac{\rho_{x_{po}y_{po}}\bar{X}}{\rho_{x_{po}y_{po}}\bar{X}+\beta_1(x)}$

$\rho_{x_{po}y_{po}}$	$\beta_2(x)$	$\frac{\rho_{x_{po}y_{po}}\bar{X}}{\rho_{x_{po}y_{po}}\bar{X}+\beta_2(x)}$
$\rho_{x_{po}y_{po}}$	D_x	$\frac{\rho_{x_{po}y_{po}}\bar{X}}{\rho_{x_{po}y_{po}}\bar{X}+D_x}$
D_x	C_x	$\frac{D_x\bar{X}}{D_x\bar{X}+C_x}$
D_x	$\beta_1(x)$	$\frac{D_x\bar{X}}{D_x\bar{X}+\beta_1(x)}$
D_x	$\beta_2(x)$	$\frac{D_x\bar{X}}{D_x\bar{X}+\beta_2(x)}$
D_x	$\rho_{x_{po}y_{po}}$	$\frac{D_x\bar{X}}{D_x\bar{X}+\rho_{x_{po}y_{po}}}$

III. PROPOSED ESTIMATOR

Let

$$T_1 = \bar{y}_{po}, T_2 = \bar{y}_{po}\left\{\frac{\bar{X}^*_{po}}{\bar{x}^*_{po}}\right\}, T_3 = \bar{y}_{po}\exp\left\{\frac{\bar{X}^*_{po}-\bar{x}^*_{po}}{\bar{X}^*_{po}+\bar{x}^*_{po}}\right\} \quad (3.1)$$

Where,

$$\bar{x}^*_{po} = a\bar{x}_{po} + b \quad (3.2)$$

$$\bar{X}^*_{po} = a\bar{X}_{po} + b \quad (3.3)$$

Let G is a set of all possible estimators to estimate the population mean $\bar{Y}$, hence $T_1, T_2, T_3 \in G$ and $T_g = \sum_{i=1}^{3} g_i T_i \in G$ such that

$$\sum_{i=1}^{3} g_i = 1, \text{where } g_i (i = 1,2,3) \in R \text{ are constants.} \quad (3.4)$$

By definition, the set T_g is linear and it may be written as

$$T_g = g_1T_1 + g_2T_2 + g_3T_3 \quad (3.5)$$

$$T_g = g_1\bar{y}_{po} + g_2\bar{y}_{po}\left(\frac{\bar{X}^*_{po}}{\bar{x}^*_{po}}\right) + g_3\bar{y}_{po}\exp\left(\frac{\bar{X}^*_{po}-\bar{x}^*_{po}}{\bar{X}^*_{po}+\bar{x}^*_{po}}\right) \quad (3.6)$$

In order to get the bias and MSE expression of the estimator T_g we write the equation in term of $\epsilon's$

$$T_g = g_1\bar{Y}(1+\epsilon_0) + g_2\bar{Y}(1+\epsilon_0)\left\{\frac{\bar{X}^*_{po}}{\bar{X}^*_{po}+a\bar{X}_{po}\epsilon_1}\right\} + g_3\bar{Y}(+\epsilon_0)\exp\left\{\frac{\bar{X}^*_{po}-(\bar{X}^*_{po}+a\bar{X}_{po}\epsilon_1)}{\bar{X}^*_{po}+(\bar{X}^*_{po}+a\bar{X}_{po}\epsilon_1)}\right\}$$

$$= \bar{Y}(1+\epsilon_0)\left[g_1 + g_2\left\{\frac{\bar{X}^*_{po}}{\bar{X}^*_{po}+a\bar{X}_{po}\epsilon_1}\right\} + g_3 exp\left\{\frac{\bar{X}^*_{po}-(\bar{X}^*_{po}+a\bar{X}_{po}\epsilon_1)}{\bar{X}^*_{po}+(\bar{X}^*_{po}+a\bar{X}_{po}\epsilon_1)}\right\}\right]$$

$$= \bar{Y}(1+\epsilon_0)\left[g_1 + g_2\{1+A\epsilon_1\}^{-1} + g_3\exp\left\{\frac{-A}{2}\epsilon_1\left(1+\frac{A}{2}\epsilon_1\right)^{-1}\right\}\right] \quad (3.7)$$

Where,

$$A = \frac{a\bar{X}_{po}}{\bar{X}^*_{po}} \quad (3.8)$$

Retaining the terms of $\epsilon's$ up to the first order of approximation, we have

$$T_g = \bar{Y}\left\{1 + \epsilon_0 - Ag\epsilon_1 + A^2\epsilon_1^2\left(g_2 + \frac{3}{8}g_3\right) - Ag\epsilon_0\epsilon_1\right\} \quad (3.9)$$

Where

$$g = \left(g_2 + \frac{g_3}{2}\right) \quad (3.10)$$

Taking expectation on both sides and then subtracting population mean, we get

$$Bias(T_g) = \frac{\bar{Y}}{n}\left[A^2C_x^2\left(g_2 + \frac{3g_3}{8}\right) - Ag\rho_{x_{po}y_{po}}C_yC_x\right] \quad (3.11)$$

$$Bias(T_g) = \frac{\lambda_2}{n}\left[\frac{A^2}{\lambda_1}\left(g_2 + \frac{3}{8}g_3\right) - \frac{Ag\rho_{x_{po}y_{po}}}{\lambda_1\lambda_2}\right] \quad (3.12)$$

$$T_g - \bar{Y} = \bar{Y}(\epsilon_0 - Ag\,\epsilon_1) \quad (3.13)$$

Squaring both sides and taking expectations, we get

$$\mathrm{MSE}\left(\mathrm{T_g}\right) = \frac{\lambda_2}{n}\left[1 + \frac{Ag\lambda_2}{\lambda_1}\left(Ag - 2\rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)\right] \quad (3.14)$$

$MSE\left(T_g\right)$ is minimum when

$$g = \frac{\rho_{x_{po}y_{po}}}{A}\sqrt{\frac{\lambda_1}{\lambda_2}} \quad (3.15)$$

Putting the value of g in MSE $(\mathrm{T_g})$, we get

$$\text{Min MSE}\left(T_g\right) = \frac{\lambda_2}{n}\left[1 - \rho^2_{x_{po}y_{po}}\right] \quad (3.16)$$

which is same as the usual regression estimator.

Since we have only two equations (3.4) and (3.10) but there are three unknowns $g_1, g_2 \& \ g_3$. In order to find the unique values of these unknowns, we have imposed a linear restriction as

$$g_1 B(T_1) + g_2 B(T_2) + g_3 B(T_3) = 0 \quad (3.17)$$

Where,

$$B\left(T_{1p}\right) = 0$$

$$B\left(t_{2p}\right) = \frac{\lambda_2}{n\lambda_1}\left(1 - \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)$$

$$B\left(t_{3p}\right) = \frac{A\lambda_2}{n\lambda_1}\left(A - \rho_{x_{po}y_{po}}\sqrt{\frac{\lambda_1}{\lambda_2}}\right)$$

Solving the equations (3.4), (3.10) and (3.17) to get the values of g_1, g_2 and g_3 as

$$g_1 = 1 - \frac{A}{\left\{\frac{1}{2} - \frac{B(t_{3p})}{B(t_{2p})}\right\}}\left\{1 - \frac{B(t_{3p})}{B(t_{2p})}\right\}$$

$$g_2 = \frac{A}{\left\{\frac{B(t_{3p})}{B(t_{2p})} - \frac{1}{2}\right\}}\left\{\frac{B(t_{3p})}{B(t_{2p})}\right\}$$

$$g_3 = \frac{A}{\left\{\frac{1}{2} - \frac{B(t_{3p})}{B(t_{2p})}\right\}}$$

Use these g_i(i =1, 2, 3) remove the bias up to terms of order (n^{-1})

IV. EMPIRICAL STUDY

In order to analyze the efficiency of proposed estimators, we examine the earthquake data of Turkey for the numerical comparisons of the proposed and existing estimators in the simple random sampling. The data is taken 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 main shocks, their foreshocks and aftershocks sequence.

We consider the main shocks that occurred between 1900 and 2011 having surface wave magnitudes MS ≥ 0.5, their foreshocks within 5 days week with MS ≥ 0.3 and aftershocks within 1month with MS ≥ 0.4. In this area, 109 main shocks with surface magnitude MS ≥ 0.5 have occurred between 1900 and 2011. The population comprises the destructive earthquakes. In the population data set number of aftershocks is a study variable and number of foreshocks is an auxiliary variable.

Note that we take sample size n = 20. The MSE values of the proposed estimators are computed considering the distribution of study and auxiliary variables. To obtain the distribution of these variables we fit the Poisson distribution to the earthquake data set. To obtain the $\rho_{x_{po}y_{po}}$ 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 γ_1, γ_2 and γ_3 are obtained.

According to the goodness of fit test, it is obvious that the Poisson distribution fits the number of shocks for Region1 with parameter $\gamma_1 = 4.1813$ $(\chi 2 = 0.048, p\ value = 0.043), \gamma_2 = 8.104$ $(\chi 2 = 0.014, p\ value = 0.032)$ for Region2 $and\ \gamma_3 = 2.112$ $(\chi 2 = 0.013, p\ value\ 0.025)$ $for\ Region$ 3. Then the correlation between the study variable and auxiliary variable is positive $\rho_{x_{po}y_{po}} = 0.712$ and it can be noted that the number of foreshocks is related to the number of aftershocks therefore, ratio estimators are preferable in this case.

Percent relative efficiency is calculated by using expression

$$\mathrm{PRE} = \frac{\mathrm{var}\left(\bar{y}_{po}\right)}{\mathrm{MSE}(.)} \times 100$$

Table II. Bias, MSE and PRE of Existing and Proposed Estimator

Estimators	Bias	MSE	PRE
y_{po}	0	0.5108	100
P_{ratio}	0.0358	0.4132	123.6087
t_1	0.0078	0.2547	200.5318
t_2	0.0125	0.9742	52.4341
t_3	0.0329	0.3957	129.0796
$\boldsymbol{T_g}$	**0.0001**	**0.2516**	**202.8167**

CONCLUSION

When the population is poisson distributed and proposed estimator is almost unbiased then from the empirical study it is shown that the efficiency of proposed estimator T_g will be equal to the regression estimator and T_g will be more efficient than other listed existing estimators.

REFERENCES

1) Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes. *Applied Mathematics and Computation*, *218*(22), 10900-10905.

2) Koyuncu, N., & Özel, G. (2013, October). Exponential estimators using characteristics of Poisson distribution: An application to earthquake data. In *AIP Conference Proceedings* (Vol. 1558, No. 1, pp. 1470-1473). American Institute of Physics.

3) Lai, C. D. (1995). Construction of bivariate distributions by a generalised trivariate reduction technique. *Statistics & probability letters*, *25*(3), 265-270.

4) Özel, G. (2011). 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.

5) 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*(2), 745-755.

6) Singh, R., Chauhan, P., Smarandache, F., & Sawan, N. (2007). *Auxiliary information and a priori values in construction of improved estimators*. Infinite Study.

7) Singh, R., & Singh, H. P. (1998). Almost unbiased ratio and product-type estimators in systematic sampling. *Qüestiió: quaderns d'estadística i investigació operativa*, *22*(3).