Applied Mathematical Sciences, Vol. 3, 2009, no. 35, 1739 - 1744

On the Estimation of Population Proportion

Sukhjinder Singh Sidhu

Department of Mathematics, Statistics and Physics
Punjab Agricultural University, Ludhiana, Punjab 141004, India
ssidhu@hotmail.com

Rajesh Tailor

School of Studies in Statistics, Vikram University
Ujjain 456010, India
rtailor@yahoo.com

Sarjinder Singh

Department of Mathematics
University of Texas at Brownsville and Texas Southmost College
Brownsville, Texas 78520, USA
sarjinder@yahoo.com

Abstract

In this paper, a new estimator of the population proportion has been proposed by using moment generating function and compared with the usual estimator from the minimum mean squared error point of view. The empirical trials show that the proposed estimator can be made more efficient than the usual estimator. The FORTRAN codes used for such study are also provided. One of the practical utilities of the proposed estimator is that it could be used in all the randomized response models while estimating proportion of a sensitive attribute by following Warner (1965).

Keywords: Estimation of proportion, moment generating function, relative efficiency

1 INTRODUCTION

The problem of estimation of population proportion is well known and can be found in all the basic book of statistics like Singh (2003, 2006), or Warner (1965). We say if $X \sim B(n,\ p)$, that is, X follows a binomial distribution with parameters n and p with probability mass function:

$$P(X = x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{(n-x)} \tag{1.1}$$

where $q = 1 - p$ and $x = 0, 1, 2, ..., n$ denote the number of successes in n trials.

Then, an unbiased estimator of the parameter p is given by

$$\widehat{p} = \frac{x}{n} \tag{1.2}$$

with variance

$$\mathrm{V}(\widehat{p}) = \frac{p(1-p)}{n} \tag{1.3}$$

In the next section, we suggest a new estimator of the population proportion p and compare its mean squared error with respect to the variance of the usual estimator in (1.3).

2 PROPOSED ESTIMATOR

We know that if $X \sim B(n,\ p)$, then the moment generating function is given by

$$\mathrm{E}\left(e^{tx}\right) = \left(q + pe^t\right)^n \tag{2.1}$$

By the method of moments, from (2.1), a new estimator of the parameter p is given by

$$\widehat{p}_{new} = \frac{e^{\frac{tx}{n}} - 1}{e^t - 1}, \qquad t \neq 0 \tag{2.2}$$

Note that if $x = 0$, then $\widehat{p}_{\text{new}} = 0$ and if $x = n$, then $\widehat{p}_{\text{new}} = 1$.

Now we have the following theorem:

Theorem 2.1. If $t \to 0$, then $\widehat{p}_{\text{new}} = 0$ is unbiased estimator of population proportion p.

Proof. The expected value of (2.2) is given by

$$\mathrm{E}\left(\widehat{p}_{new}\right) = \frac{\left(q + pe^{\frac{t}{n}}\right)^n - 1}{e^t - 1} \tag{2.3}$$

Taking the limit as $t - 0$, and applying the L-Hospital Rule, we have

$$\lim_{t \to 0} E\left(\widehat{p}_{\text{new}}\right) = p$$

which proves the theorem.

Thus from (2.3) the bias in the new estimator $\widehat{p}_{\text{new}}$ can be written as

$$\begin{aligned} \mathrm{B}\left(\widehat{p}_{\text{new}}\right) &= \mathrm{E}\left(\widehat{p}_{\text{new}}\right) - p \\ &= \frac{\left(q + pe^{\frac{t}{n}}\right)^n - \left(q + pe^t\right)}{e^t - 1} \end{aligned} \tag{2.4}$$

Again by L-Hospital Rule, as $t \to 0$ then

$$B\left(\widehat{p}_{\text{new}}\right) = 0 \tag{2.5}$$

Theorem 2.2. The variance of the new estimator $\widehat{p}_{\text{new}}$ is given by

$$\mathrm{V}\left(\widehat{p}_{\text{new}}\right) = \frac{\left(q + pe^{\frac{2t}{n}}\right)^n - \left(q + pe^{\frac{t}{n}}\right)^{2n}}{(e^t - 1)^2} \tag{2.6}$$

Proof. It follows from the fact that

$$V\left(\widehat{p}_{\text{new}}\right)=E\left(\widehat{p}_{new}^{2}\right)-\left\{E\left(\widehat{p}_{new}\right)\right\}^{2}$$

From (2.4) and (2.6), the mean squared error (MSE) of the new estimator $\widehat{p}_{new}$ is

$$\begin{aligned}\text{MSE}\left(\widehat{p}_{\text{new}}\right) &= V\left(\widehat{p}_{\text{new}}\right)+\left\{B\left(\widehat{p}_{\text{new}}\right)\right\}^{2}\\ &= \frac{1}{\left(e^{t}-1\right)^{2}}\left[\left(q+pe^{\frac{2t}{n}}\right)^{n}-\left(q+pe^{\frac{t}{n}}\right)^{2n}+\left\{\left(q+pe^{\frac{t}{n}}\right)^{n}-\left(q+pe^{t}\right)\right\}^{2}\right]\end{aligned} \tag{2.7}$$

Again by applying the L-Hospital Rule, one can easily see that

$$\lim_{t\to 0}\text{MSE}\left(\widehat{p}_{\text{new}}\right)=\frac{p\left(1-p\right)}{n}=V\left(\widehat{p}\right) \tag{2.8}$$

Theorem2.4. If $t\to 0$ then $\widehat{p}_{\text{new}}$ and $\widehat{p}$ are same. If $t\neq 0$, then there may exist a non-zero t such that

$$\text{MSE}\left(\widehat{p}_{\text{new}}\right)<V\left(\widehat{p}\right) \tag{2.9}$$

In the next section, we wrote FORTRAN codes in IMSL subroutines using double precision to search the values of t such that the condition (2.9) is satisfied, and we computed the percent relative efficiency (RE) of the proposed estimator $\widehat{p}_{\text{new}}$ with respect to the usual estimator $\widehat{p}$ as

$$\text{RE}=\frac{V\left(\widehat{p}\right)}{\text{MSE}\left(\widehat{p}_{new}\right)}\times 100\% \tag{2.10}$$

3 FORTRAN CODES

```
!               PROGRAM
                  IMPLICIT NONE
                  REAL P, Q, N, T1, T2, T3, T4, T, MSE, VAR,RE
```

```
          CHARACTER*20 OUT_FILE
          WRITE(*,'(A)') 'NAME OF THE OUTPUT FILE'
          READ(*,'(A20)') OUT_FILE
          OPEN(42, FILE=OUT_FILE, STATUS='UNKNOWN')
          WRITE(42,118)
118        FORMAT(3X,'N',3X,'P',3X,'T',3X,'RE')
          DO 111 N = 5, 20, 5
           DO 111 P = 0.1, 0.91, 0.1
          Q = 1-P
          DO 111 T = -1.50, +1.50, 0.01
           IF ( T.NE.0 ) THEN
           T1 = ( Q + P*EXP(2*T/N) )**N
          T2 = ( Q + P*EXP(T/N) )**(2*N)
          T3 = ( Q + P*EXP(T/N) )**N
          T4 = ( Q + P*EXP(T) )
          MSE = (T1-T2+(T3-T4)**2) /(EXP(T)-1)**2
          VAR = P*Q/N
          RE = VAR*100/MSE
           IF((MSE.GT.0).AND.(MSE.LE.VAR)) THEN
          WRITE(42,117)N,P,T,RE
117        FORMAT(2X,F5.1,3X,F5.3,3X,F7.3,3X,F14.2)
         ENDIF
         ENDIF
111        CONTINUE
          END
```

4 RESULTS AND DISCUSSION

Using the above FORTRAN codes, it has been observed the if $0 < p < 0.5$, then the value of t remains positive for the relative efficiency to be more than 100%, and if $0.5 < p < 1$, then the value of t remains negative for the relative efficiency to be more than 100%. For $p = 0.5$, the new estimator $\widehat{p}_{new}$ remains as efficient as $\widehat{p}$ for t very close to zero. Table 4.1 gives the ranges of the values of t and relative efficiency. Note that in the present study we considered only $n = 5$ and 20, and t between -1.5 to $+1.5$ with step of 0.1, and true proportion p between 0.1 and 0.9 with a step 0.1, and found that the percent relative efficiency (RE) changes from 100% to 218% with median efficiency of 115%, first quartile is 105% and the third quartile is 136%.

Table 6.1. Precent relative efficiency (RE) of $\widehat{p}_{new}$ with respect to $\widehat{p}$.

Sample size (n)	True parameter (p)	Range of t	Range of RE (%)
5	0.1	$0.1 < t < 1.4$	$106 <$ RE < 211
	0.2	$0.1 < t < 1.4$	$104 <$ RE < 148
	0.3	$0.1 < t < 1.4$	$103 <$ RE < 113
	0.4	$0.1 < t < 0.9$	$101 <$ RE < 103
	0.6	$-0.9 < t < -0.1$	$100 <$ RE < 103
	0.7	$-1.5 < t < -0.1$	$105 <$ RE < 114
	0.8	$-1.5 < t < -0.1$	$104 <$ RE < 148
	0.9	$-1.5 < t < -0.1$	$106 <$ RE < 218
20	0.1	$0.1 < t < 1.4$	$107 <$ RE < 137
	0.2	$0.1 < t < 0.8$	$100 <$ RE < 111
	0.3	$0.1 < t < 0.4$	$100 <$ RE < 104
	0.4	$0.0 < t < 0.1$	$100 <$ RE < 101
	0.6	$-0.1 < t < 0.0$	$100 <$ RE < 101
	0.7	$-0.4 < t < -0.1$	$100 <$ RE < 104
	0.8	$-0.8 < t < -0.1$	$100 <$ RE < 111
	0.9	$-1.5 < t < -0.1$	$107 <$ RE < 137

Acknowledgements: The authors are thankful to the Editor and the referee for their comments on the original version of the manuscript.

References

[1] Singh, S. (2003). *Advanced Sampling theory with applications: How Michael Selected Amy.* Kluwer Academic Publishers, The Netherlands.

[2] Singh, S.(2006). *Thinking Statistically: Elephants Go to School.* Kendall/Hunt Publishing Company, Iowa, USA.

[3] Warner, S.L. (1965). Randomized response : A survey technique for eliminating evasive answer bias. *Jour. Amer. Statist. Assoc.*, 60, 63–69.

Received: November, 2008