STATISTICS IN TRANSITION new series, September 2019
Vol. 20 , No. 3, pp. 1-29, DOI 10.21307/stattrans-2019-022
Submitted – 09.06.2018; Paper ready for publication – 19.03.2019

CHARACTERIZATION OF THE SUM OF BINOMIAL RANDOM VARIABLES UNDER RANKED SET SAMPLING

Vivek Verma[1], Dilip C. Nath[2]

ABSTRACT

In this paper, we examined the characteristics of the sum of independent and non-identical set of binomial ranked set samples, where each set has different order depending success probability. The characterization is done by establishing the general recurrence relations for two different situations based on the number of cycle, which is initially pre-assumed as a constant integer and when it is a random variable. To extend the knowledge about the characteristics of sum in terms of their behaviour and pattern, first four moments *i.e.*, mean, variance, skewness and kurtosis are derive and compared with the sum of binomial simple random samples with same success probability. The proposed procedure has been illustrated through a real-life data on survivorship of children below one year in Empowered Action Groups (EAG) states of India.

Key words: Factorial moment generating function, Skewness; Kurtosis, Poisson distribution.

1. Introduction

The role of Ranked Set Sampling (RSS) as an alternative method of Simple Random Sampling (SRS) have been investigated since the time McIntyre(1952), who first introduced this sampling procedure. Since, then many authors have discussed about the efficacy of RSS either theoretically or analytically. RSS is found to be very effective in contexts where exact measurement of sampling units is expansive in time or toil; but the sample unit can be readily ranked either through subjective or *via* the use of relevant concomitant variables.

[1]**Corresponding Author**. Department of Statistics, Gauhati University, Guwahati, Assam, 781014, India, **Current**: Department of Neurology, All India Institute of Medical Sciences, New Delhi, 110029, India. E-mail: viv_verma456@yahoo.com

[2]Assam University, Silchar, 788011, Assam, India. E-mail: dilipc.nath@gmail.com

Comparing with earlier studies where the variable of interest is continuous, infrequent research discuss about the effectiveness of RSS where the variable is binary. In cases, where the variable of interest is binary, there are two possible outcomes, success (denoted as 1) and failure (denoted as 0) and is supposed to follow Bernoulli with success probability p, say. Here, the probability p can be viewed as a proportion of individuals with certain characteristic in the population. The foregoing studies of RSS, where the response is a binary variable, Terpstra (2004), Chen et al. (2005, 2006, 2007, 2008), Chen (2008), Verma et.al(2017), Das et.al (2017) Lacayo et al. (2002), Terpstra and Miller (2006) and Chen et al. (2009), are mainly concerned about estimation of population proportion and variance, and the comparison with SRS is done using these estimates.

Obtaining the behaviour of a sum of Bernoulli random variables based on simple random sample has found of greater importance in various applications like formalization random walk process (Takacs, 1991), the Stein-Chen method for approximation of Poisson (Barbour and Holst, 1989), obtaining bounds for entropy (Sason, 2013), characterization of flows in internet traffic (Chabchoub et al. 2010), and approximation of rare events (Chen and Rollin, 2013). The problem of estimating the characteristics of sum of independent binary variable in terms of their moments based on simple random samples has already been emphasized by many researchers like Malik (1969), Ahuja (1970), Percus and Percus (1985), Ling (1988), Horvath (1989), Yu and Zelterman (2002), and Kadane (2016). As an alternative procedure of SRS, RSS has found to be more efficient and reliable, but the characterization of a sum of independent and non-identical Bernoulli random variables based on ranked set sample has not been considered in the literature. In this connection the present article has mainly concerned to establish a recurrence relations between the factorial moments of sum of independent and non-identical sets of binary variables, which is never procured in case of RSS. These relations also assists to reduce the number of independent calculations required for evaluation of moments under RSS. And, helps in characterizing the sum of binary variables under RSS by using recurrence relations and compare with SRS.

In this paper, the recurrence relationship of sum of binary variables under SRS and balanced RSS for fixed set size, s, and probabilities of success, p, are obtained under two different situations. In the first case, the relationship is obtained when number of cycles, m is assumed to be known fixed values. In the second case, an attempt is being made to extend the previously mentioned results of recurrence relations among the sum of independent binary variable where the number of cycle is a random variable. To understand the characteristics of sum of binary variables under SRS and RSS, the first four moments are derived using factorial moments that by establishing the recurrence relationships. The technique of asymptotic approximations has been found very helpful in various aspects like in Monte Carlo simulation (Hastings, 1970) and bootstrap techniques (Freedman and Peters, 1984, Brown and Newey, 2002), for obtaining numerical estimates and their asymptotic variance and asymptotic confidence intervals. In characterization of any distributions, asymptotical aspects adds additional information of the distribution and it is an important feature to describe the how large the sample size is required to achieve the asymptotic approximation. A simulation based comparison among SRS and RSS has been discussed to numerically illustrate the requirement of sample size to achieve the asymptotic normality. A practical illustration of the proposed procedure with a real-life data on child survivorship for all eight selected Empowered Action Groups (EAG) Indian states viz., Bihar, Uttaranchal, Chhatisgarh, Jharkhand, Orissa, Rajasthan, Madhya Pradesh and Uttar Pradesh, has also been presented.

2. Sampling Design

Suppose the variable of interest is dichotomous variable, say X, and $n(=ms)$, denotes size of the sample drawn from the population by adopting the procedures of SRS and RSS, respectively, for prefixed set size,s and number of cycles, m. Let $\{X_{[r]i}; r = 1(1)s, i = 1(1)m\}$ symbolizes a ranked set sample of size ms, where $X_{[r]i}$ denotes the i^{th} observation in the r^{th} ranking class. Because of RSS procedure, $X_{[r]i}$'s are independently distributed and corresponding to each r^{th} set, $(X_{[r]1}, X_{[r]2}, \cdots, X_{[r]m})$ are independently and identically (i.i.d.) distributed and $X_{[r]1}$ is the r^{th} order statistic from a simple random sample of s observations on X. Let $\mathbf{X_{SRS}} = (X_1, X_2, \cdots, X_n)$ is an i.i.d. simple random sample from Bernoulli(p) and $W(= \sum_{i=1}^{n} X_i)$, denotes their sum and its density is given by,

$$f_W(w) \quad = \quad \binom{n}{w} p^w \, (1-p)^{n-w} \quad ; w = 0(1)n; 0 \le p \le 1 \tag{2.1}$$

Let $\mathbf{X}_{[\mathbf{r}]} = (X_{[r]1}, X_{[r]2}, \cdots, X_{[r]m})$ is an vector of i.i.d. ranked set samples of r^{th} set from Bernoulli $(p_{[r]})$, for all $r = 1(1)s$ and $Y_r = \sum_{j=1}^{m} X_{[r]j}$, is the number of times the event occurred in r^{th} class, follows Binomial $(m, p_{[r]})$ and is given by,

$$P(Y_r = y_r) = \binom{m}{y_r} p_{[r]}^{y_r} (1 - p_{[r]})^{m-y_r} \quad ; y_r = 0, 1, \cdots, m. \tag{2.2}$$

Here, $p_{[r]} = I_p(s-r+1, r)$, denotes the standard incomplete beta integral and is given by,

$$I_x(a,b) = \frac{1}{\mathscr{B}(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt, \ \ 0 < x < 1.$$

where $\mathscr{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. And, $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_r, \cdots, Y_s)$ is vector of independent Binomial variate with parameters $m > 0$ and $p_{[r]}$ for all $r = 1(1)s$ and $Z = \sum_{r=1}^{s} Y_r$, denotes their sums.

3. Characterization using the Recurrence Relations

Let $G(t) = \sum_{x=0}^{n} t^x P(X = x)$, denotes the probability generating function (*pgf*) of a random variable X having distribution $P(X = x)$, with support $0, 1, \cdots$, $n(= ms) \in \mathbf{Z}^+$.

3.1. Case-I: The number of cycles, m, is a known fixed value

Theorem-1: For fixed m, the recursive relationship among factorial moment of sum W and Z, under SRS and RSS, respectively, is given by

$$\mu'_{[k]}(W) = ms \sum_{j=0}^{k-1} (-1)^j \frac{(k-1)!}{(k-1-j)!} p^{(j+1)} \mu'_{[k-1-j]}(W) \tag{3.1}$$

and

$$\mu'_{[k]}(Z) = m \sum_{j=0}^{k-1} (-1)^j \frac{(k-1)!}{(k-1-j)!} \left(\sum_{i=1}^{s} p_{[i]}^{j+1} \right) \mu'_{[k-1-j]}(Z). \tag{3.2}$$

where $\mu'_{[0]} = 1$.

Proof: Suppose that W, denotes the sum of $n(= ms)$ $i.e., = \sum_{i=1}^{n} X_i$ and Y_r;

$\forall r = 1(1)s$, is the sum of m $i.e., = \sum_{j=1}^{m} X_{[r]j}$, Bernoulli variables with parameters p and $p_{[r]}$ respectively. The factorial moment generating function (*fmgf*) of W and Y_r using equations (2.1)-(2.2) respectively, are given by

$$\begin{aligned} G_W(t+1) &= \prod_{i=1}^{n} G_{X_i}(t+1) = (1+pt)^n & (3.3) \\ G_{Y_r}(t+1) &= E((t+1)^{Y_r}) = \prod_{j=1}^{m} E((t+1)^{X_{[r]j}}) \\ &= (1+p_{[r]}t)^m. & (3.4) \end{aligned}$$

Since, $\{Y_r\}$'s is a set of mutually independent Binomial variate, therefore, the *fmgf* of $Z(= \sum_{r=1}^{s} Y_r)$ using equation-(3.4) is given by

$$G_Z(t+1) = \prod_{r=1}^{s} G_{Y_r}(t+1) = \prod_{r=1}^{s} (1+p_{[r]}t)^m. \tag{3.5}$$

If D denotes the differential operator $i.e., \frac{d}{dt}$, then the recursive relationship between the factorial moments of W, based on simple random samples, can be obtained by successive differentiation of equation-(3.3) and are as follows

$$\begin{aligned} D(G_W(1+t)) &= \frac{np}{1+tp} G_W(1+t), \\ D^2(G_W(1+t)) &= \frac{np}{1+tp} D(G_W(1+t)) - \frac{np^2}{(1+tp)^2} G_W(1+t), \\ D^3(G_W(1+t)) &= \frac{np}{1+tp} D^2(G_W(1+t)) - \frac{2np^2}{(1+tp)^2} D(G_W(1+t)) + \\ &\quad \frac{2np^3}{(1+tp)^3} G_W(1+t), \\ \vdots &= \vdots, \end{aligned}$$

$$D^k(G_W(1+t)) = n \sum_{j=0}^{k-1} (-1)^j \frac{(k-1)!}{(k-1-j)!} \left(\frac{p}{1+tp} \right)^{j+1} D^{k-1-j}(G_Z(1+t)), \tag{3.6}$$

and setting $t=0$ in equation-(3.6) gives

$$\mu'_{[k]}(W) = ms\sum_{j=0}^{k-1}(-1)^j \frac{(k-1)!}{(k-1-j)!} p^{(j+1)} \mu'_{[k-1-j]}(W),$$

where $\mu'_{[0]}(W) = 1$.

The recursive relationship between *fmgf* of Z, which is based on ranked set samples, using equation-(3.5), is given by

$$\begin{aligned}
D(G_Z(1+t)) &= \sum_{i=1}^{s}\left(\frac{mp_{[i]}}{1+tp_{[i]}}\right)\prod_{r=1}^{s}(1+tp_{[r]})^m = \sum_{i=1}^{s}\left(\frac{mp_{[i]}}{1+tp_{[i]}}\right)G_Z(1+t),\\
D^2(G_Z(1+t)) &= D(G_Z(1+t))\sum_{i=1}^{s}\left(\frac{mp_{[i]}}{1+tp_{[i]}}\right) - G_Z(1+t)\sum_{i=1}^{s} m\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)^2\\
D^3(G_Z(1+t)) &= D^2(G_Z(1+t))\sum_{i=1}^{s}\left(\frac{mp_{[i]}}{1+tp_{[i]}}\right) -\\
&\quad D(G_Z(1+t))\sum_{i=1}^{s} 2m\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)^2 *\\
&\quad G_Z(1+t)\sum_{i=1}^{s} 2m\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)^3,\\
\vdots &= \vdots,\\
D^k(G_Z(1+t)) &= m\sum_{j=0}^{k-1}(-1)^j h(k)_j\left(\sum_{i=1}^{s} p_{[i]}^{j+1}\right)D^{k-1-j}(G_Z(1+t)), \qquad (3.7)
\end{aligned}$$

where $h(k)_j = \frac{(k-1)!}{(k-1-j)!}$, setting $t=0$ in equation-(3.7) provides

$$\mu'_{[k]}(Z) = m\sum_{j=0}^{k-1}(-1)^j \frac{(k-1)!}{(k-1-j)!}\left(\sum_{i=1}^{s} p_{[i]}^{j+1}\right)\mu'_{[k-1-j]}(Z),$$

where $\mu'_{[0]}(Z) = 1$.

Corollary 1: The first four factorial moments *i.e.,* for $k = 1,2,3$ and 4, using equation-(3.1) of W and (3.1) of Z, based on simple random samples and ranked set sample, are given in Appendix-(7.1).

3.2. Case-II: The number of cycles, m, is a random variable

Suppose that the number of cycles, N, is a random variable, where $m \in N^+ = \{1,2,,\cdots\}$. The probability mass function of $N=m$ is given by

$$f_N(m) = \frac{e^{-\lambda}\ \lambda^{m-1}}{(m-1)!}\ ; m = 1,2,\cdots. \tag{3.8}$$

i.e., $N-1 \sim \mathscr{P}(\lambda)$ (Poisson with mean λ), $\lambda > 0$.

Theorem-2: The recursive relationship between factorial moments of marginal sums of W and Z respectively, where N is a random variable and follows the Poisson distribution of equation-(3.8), is given by

$$\mu'_{[k]}(W) = s\sum_{j=0}^{k-1} \frac{(k-1)!}{(k-1-j)!} p^{(j+1)} \left((-1)^j + \frac{\lambda(s-1)!}{(s-1-j)!j!}\right) \mu'_{[k-1-j]}(W), \tag{3.9}$$

$$\mu'_{[k]}(Z) = \lambda\left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[k-1]}(Z) + \sum_{j=0}^{k-1}(-1)^j \frac{(k-1)!}{(k-1-j)!}\left(\sum_{i=1}^{s} p_{[i]}^{j+1}\right)\mu'_{[k-1-j]}(Z) \tag{3.10}$$

Proof: Suppose W and Y_r, are the mixtures of binomial distributions with a fixed probability of success p and $p_{[r]}$, respectively, but a variable number of cycles, m, modelled with Poisson distribution discussed in equation-(3.8). The conditional distribution of $W|N=m$ is given by,

$$f_W(w|N=m) \quad = \quad \binom{n}{w}\ p^w\ (1-p)^{n-w} = \binom{ms}{w}\ p^w\ (1-p)^{ms-w}$$

where $w = 0,1,\cdots,ms$ and $0 \le p \le 1$. The *fmgf* of mixture of W using the equations-(2.1) and (3.8) can be derived as,

$$\begin{aligned} G_W(1+t) &= e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}\sum_{w=0}^{ms}\binom{ms}{w}\ ((1+t)p)^w\ (1-p)^{ms-w}, \\ &= e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}(1+tp)^{ms}, \\ &= e^{-\lambda}(1+tp)^s e^{\lambda(1+tp)^s}. \end{aligned} \tag{3.11}$$

If D denotes the differential operator $i.e.,\ \frac{d}{dt}$, then the recursive relationship between the factorial moments of the mixture of W can be obtained by successive differentiation of equation-(3.11) and are as follows,

$$\begin{aligned} D(G_W(1+t)) &= e^{-\lambda}\{ps(1+tp)^{s-1}e^{\lambda(1+tp)^s} + (1+tp)^s e^{\lambda(1+tp)^s}\lambda sp(1+tp)^{s-1}\}, \\ &= G_W(1+t)\left\{\frac{sp}{1+tp} + \lambda sp(1+tp)^{s-1}\right\}, \\ &= s\,G_W(1+t)\left(\frac{p}{1+tp}\right) + s\lambda p\,G_W(1+t)(1+tp)^{s-1}, \end{aligned}$$

$$\begin{aligned} D^2(G_W(1+t)) &= \frac{sp}{1+tp}D(G_W(1+t)) - \frac{sp^2}{(1+tp)^2}G_W(1+t) + \\ & \lambda sp(1+tp)^{s-1}D(G_W(1+t)) + \\ & \lambda s(s-1)p^2(1+tp)^{s-2}G_W(1+t), \end{aligned}$$

$$\begin{aligned} D^3(G_W(1+t)) &= \frac{sp}{1+tp}D^2(G_W(1+t)) - \frac{2sp^2}{(1+tp)^2}D(G_W(1+t)) + \\ & \frac{2sp^3}{(1+tp)^3}G_W(1+t) + \lambda sp(1+tp)^{s-1}D^2(G_W(1+t)) \\ & 2\ \lambda s(s-1)p^2(1+tp)^{s-2}G_W(1+t) \\ & \lambda s(s-1)(s-2)p^3(1+tp)^{s-3}G_W(1+t), \\ \vdots &= \vdots, \end{aligned}$$

$$\begin{aligned} D^k(G_W(1+t)) &= \\ s\sum_{j=0}^{k-1}\frac{(k-1)!}{(k-1-j)!}\left(\frac{p}{1+tp}\right)^{j+1} & \left((-1)^j + \frac{\lambda(s-1)!}{(s-1-j)!j!}(1+tp)^s\right)D^{k-1-j}, \end{aligned} \tag{3.12}$$

setting $t = 0$ in equation-(3.12) gives the recursive relationship,

$$\mu'_{[k]}(W) = s\sum_{j=0}^{k-1}\frac{(k-1)!}{(k-1-j)!}p^{(j+1)}\left((-1)^j + \frac{\lambda(s-1)!}{(s-1-j)!j!}\right)\mu'_{[k-1-j]}(W) \tag{3.13}$$

where $\mu'_{[0]}(Z) = 1$.

The *fmgf* of mixture of Y_r of the r^{th} set of $\mathbf{Y}$ by using the equations-(2.2) and (3.8) can be derived as,

$$\begin{aligned}
G_{Y_r}(t+1) &= e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}\sum_{y_r=0}^{m}(1+t)^{y_r}\binom{m}{y_r}\,p_{[r]}^{y_r}\,(1-p_{[r]})^{m-y_r}\\
&= e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}(1+tp_{[r]})^{m}\\
&= (1+tp_{[r]})e^{-\lambda}\sum_{m=1}^{\infty}\frac{((1+tp_{[r]})\lambda)^{(m-1)}}{(m-1)!}\\
&= (1+tp_{[r]})\,e^{-\lambda(1-(1+tp_{[r]}))} = (1+tp_{[r]})\,e^{\lambda tp_{[r]}}. \qquad (3.14)
\end{aligned}$$

The *fmgf* of Z using the equation-(3.14) is given by,

$$\begin{aligned}
G_Z(1+t) &= \prod_{r=1}^{s}G_{Y_r}(1+t) = e^{t\lambda\sum\limits_{r=1}^{s}p_{[r]}}\prod_{r=1}^{s}(1+tp_{[r]})\\
&= e^{ta}\prod_{r=1}^{s}(1+tp_{[r]}), \qquad (3.15)
\end{aligned}$$

where $z=0,1,\cdots,ms$ and $0\leq p_{[r]}\leq 1$, $a=\lambda sp$ and $p=\sum\limits_{r=1}^{s}p_{[r]}/s$, for all $r=1(1)s$. The recursive relationship between the factorial moments of the mixture of Z can be obtained by successive differentiation of equation-(3.15) and are as follows,

$$\begin{aligned}
D(G_Z(1+t)) &= ae^{ta}\prod_{r=1}^{s}(1+tp_{[r]}) + e^{ta}\sum_{i=1}^{s}\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)\prod_{r=1}^{s}(1+tp_{[r]}),\\
&= aG_Z(1+t) + G_Z(1+t)\sum_{i=1}^{s}\left(\frac{p_{[i]}}{1+tp_{[i]}}\right),\\
D^2(G_Z(1+t)) &= aD(G_Z(1+t)) + D(G_Z(1+t))\sum_{i=1}^{s}\left(\frac{p_{[i]}}{1+tp_{[i]}}\right) -\\
&\quad G_Z(1+t)\sum_{i=1}^{s}\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)^{2},
\end{aligned}$$

$$
\begin{aligned}
D^3(G_Z(1+t)) &= aD^2(G_Z(1+t)) + D^2(G_Z(1+t))\sum_{i=1}^{s}\left(\frac{p_{[i]}}{1+tp_{[i]}}\right) - \\
&\quad D(G_Z(1+t))\sum_{i=1}^{s}2\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)^2 + \\
&\quad G_Z(1+t)\sum_{i=1}^{s}2\left(\frac{p_{[i]}}{1+tp_{[i]}}\right)^3, \\
\vdots &= \vdots, \\
D^k(G_Z(1+t)) &=
\end{aligned}
$$

$$
aD^{k-1}(G_Z(1+t)) + \sum_{j=0}^{k-1}(-1)^j\frac{(k-1)!}{(k-1-j)!}\left(\sum_{i=1}^{s}p_{[i]}^{j+1}\right)D^{k-1-j}(G_Z(1+t)), \quad (3.16)
$$

setting $t=0$ in equation-(3.16) gives recursive relationship as,

$$
\mu'_{[k]}(Z) = a\mu'_{[k-1]}(Z) + \sum_{j=0}^{k-1}(-1)^j\frac{(k-1)!}{(k-1-j)!}\left(\sum_{i=1}^{s}p_{[i]}^{j+1}\right)\mu'_{[k-1-j]}(Z)
$$

where $\mu'_{[0]}(Z) = 1$.

Corollary 2: The first four factorial moments, *i.e.,* $k = 1,2,3$ and 4, of the mixture of W and Z using equation-(3.9) and equation-(3.10) based on simple random sample and ranked set sample, respectively, where the number of cycles, m, is a random variable and follows the Poisson distribution, is given in Appendix-(7.2).

4. Comparison of Moments

In this section, a comparison is being made among the factorial moments of W and Z. Let $\mathscr{D}_{[k]} = \mu'_k(W) - \mu'_k(Z)$, denotes the difference among factorial moments of W and Z, of order k. For the situation, where m, is a known constant, the difference, $\mathscr{D}_{[k]}$, by using equations-(3.1) and (3.2), is given by

$$
\mathscr{D}_{1[k]} = m\sum_{j=0}^{k-1}(-1)^j\frac{(k-1)!}{(k-1-j)!}\left(sp^{(j+1)}\mu'_{[k-1-j]}(W) - \mu'_{[k-1-j]}(Z)\sum_{i=1}^{s}p_{[i]}^{j+1}\right). \quad (4.1)
$$

Under the assumption that m, is a random variable, $\mathscr{D}_{[k]}$ can be obtained by using the equations-(3.9) and (3.10) and is given by,

$$\mathscr{D}_{2[k]} =$$

$$\sum_{j=0}^{k-1}\left(sp^{(j+1)}\ \gamma_j\ \mu'_{[k-1-j]}(W) - (-1)^j\left(\sum_{i=1}^{s} p_{[i]}^{j+1}\right)\mu'_{[k-1-j]}(Z)\right) \times \frac{(k-1)!}{(k-1-j)!} - \lambda\left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[k-1]}(Z), \quad (4.2)$$

where $\gamma_j = \left((-1)^j + \frac{\lambda(s-1)!}{(s-1-j)!j!}\right)$.

Note-1: Since, $p_{[i]} = I_p(s-i+1,i)$, $\forall\ i = 1,2,\cdots,s$, therefore, a sum of $p_{[i]}$'s is given by,

$$\begin{aligned}\sum_{i=1}^{s} p_{[i]} &= \sum_{i=1}^{s} I_p(s-i+1,i) = \sum_{i=1}^{s}\int_0^p \frac{t^{s-i}(1-t)^{i-1}}{\mathscr{B}(s-i+1,i)}dt \\ &= \int_0^p \sum_{i=1}^{s}\binom{s-1}{i-1} t^{s-i}(1-t)^{i-1}dt = s\int_0^p dt \ = \ sp.\end{aligned}$$

Note-2: Let $\gamma(v) = \frac{1}{p^v}\sum_{i=1}^{s} p_{[i]}^v$, is a constant depends on the order v and for $v = 1$, $\gamma(1) = sp/p = s$, *i.e.*, the minimum value of $\gamma(v)$ is s, that implies $\gamma(v) \geq s\ ; \forall v$. Suppose that,

$$\sum_{i=1}^{s} p_{[i]}^v = Cp^v,$$

where $C > 0$, is a proportionality constant such that,

$$\sum_{i=1}^{s} p_{[i]}^v - Cp^v \ = \ 0 \text{ if } C = \gamma(v) \quad (4.3)$$

$$\sum_{i=1}^{s} p_{[i]}^v - Cp^v \ > \ 0 \text{ if } C \in (0,\gamma(v)] \quad (4.4)$$

$$\sum_{i=1}^{s} p_{[i]}^v - Cp^v \ < \ 0 \text{ if } C > \gamma(v). \quad (4.5)$$

Note-3: The difference equations of $\mathscr{D}_{1[k]}$ and $\mathscr{D}_{2[k]}$, of equations-(4.1)-(4.2), respectively, for $k = 1,2$ are such that

$$\mathscr{D}_{1[1]} = m\left(sp - \sum_{i=1}^{s} p_{[i]}\right) = 0 \tag{4.6}$$

$$\mathscr{D}_{1[2]} = m\left(\sum_{i=1}^{s} p_{[i]}^2 - sp^2\right) > 0; \text{ from-equations (4.6) and (4.4)} \tag{4.7}$$

$$\mathscr{D}_{2[1]} = sp(1+\lambda) - \sum_{i=1}^{s} p_{[i]} - \lambda \sum_{i=1}^{s} p_{[i]} = 0 \tag{4.8}$$

$$\begin{aligned}\mathscr{D}_{2[2]} &= sp(1+\lambda)\mu'_{[1]}(W) + sp^2(-1+\lambda(s-1)) - \mu'_{[1]}(Z)\left[\lambda sp + \sum_{i=1}^{s} p_{[i]}\right] + \sum_{i=1}^{s} p_{[i]}^2 \\ &= \left[\sum_{i=1}^{s} p_{[i]}^2 - sp^2\right] + sp^2\lambda(s-1) > 0; \text{from-equations (4.8) and (4.4)}\end{aligned} \tag{4.9}$$

Since, $\sum_{i=1}^{s} p_{[i]} = sp$ and $s \in (0, \gamma(v)]; \forall v$, therefore, from equation-(4.4), we find that the difference between, $\left(\sum_{i=1}^{s} p_{[i]}^{v} - sp^{v}\right) > 0$.

5. Simulations

To assess the performance and changes in the moments of sums, when set size, m, is known and unknown, a simulation study is done for different combination of $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, $s = 2,4$ and 6, and $m = \lambda = 10, 50, 100, 200$ and 500, under SRS and RSS, are presented in Table 1-2 of Appendix. To compare the accuracy of the estimator under RSS with respect to SRS, the relative efficiency (RE)

$$RE = \frac{\mu_2(SRS)}{\mu_2(RSS)}$$

is also obtained. In addition of that pattern of the skewness and kurtosis of sums based on both SRS and RSS regarding their asymptotic behaviour are also depicted in Figure 1-4 of Appendix.

Discussion: From Table 1, it has observed that the RE of the estimator under RSS with respect to SRS, always greater than 1 for all combination of s, m and p. For fixed number of cycles, m and proportion, p, the RE of the estimator under RSS with respect to SRS, is also increasing in most of the cases with increase in set size, s, such as for $m = 10$ and $p = 0.1$, with increase the value of $s = 2,4,6$, the RE is 1.21, 1.32 and 1.39, respectively. When the set size, s, is fixed and number of cycles are significantly high values, the RE of the estimator under RSS with respect to SRS has followed an increasing trend with increase in proportion p, for e.g., when $m = 500$ and $s = 6$, the RE has obtained as 1.38, 1.81, 2.02, 2.26 and 2.31 for $p = 0.1, 0.2, \cdots 0.5$, respectively. Table 2 is based on the assumption that the number of cycles, is a random quantity and follows zero truncated poisson(λ). Under this truncated poisson assumption, similar pattern of the RE of the estimator under RSS with respect to SRS as previous has obtained. Results has shown that with increase in s and p higher will be the values of RE. It has also found that for fixed s and p, changing in poisson parameter λ does not affect the efficiency of estimators under RSS as comparing to SRS, and remains almost same. It is found from the simulated results that even though the mean under both SRS and RSS are same (Verma et.al (2017)) but the variances of sums based on SRS are often higher than that of RSS, for all m and p, which shows that the number of success obtained using RSS is more reliable and efficient as compare to SRS.

The interaction of sample size, $n = ms$, and skewness and kurtosis, respectively, of sums under SRS and RSS has depicted in Figure 1-2, where the number of cycles m is a fixed quantity. Figures 1 and 2 represents the pattern of five skewness and kurtosis curves, respectively, obtained for fixed set size, s, at different choices of $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. When m is a random quantity and follows zero truncated poisson(λ), the pattern of five skewness and kurtosis curves for different choices of p have also presented in Figure 3-4. Through these figures, one can compare and calculate the required sample size, n, to meet that asymptotic normality (Small, 1980, Bai and NG,2005, Sunklodas, 2014, and Butler and Stephens, 2017), *i.e.,* $skewness = 0$ and $kurtosis = 3$, under RSS as compare to SRS, for given p. For fixed set size, s, and proportion p, it has found that with a minimum number of cycles, m or parameter λ, one can achieve asymptotic normality, under RSS as compare to the required number of sample based on SRS.

6. Illustration with Real-life Data

To illustrate a practical significance of the discussed methodology, a real-life data on children aged 0-1 years to mothers aged 15 to 39 years, who are residing in eight Indian states ((a) Bihar (b) Uttaranchal (c) Chhatisgarh (d) Jharkhand (d) Orissa (e) Rajasthan (f) Madhya Pradesh and (g) Uttar Pradesh,) has considered. These selected eight states are socio-economically backward and reports highest infant mortality rates ($<$ 50 per 1000), and are also known as Empowered Action Groups (EAG) states of India. The data has been obtained from the National Family Health Survey-3 (2005-06), preceding five years of the survey. Here, our objective is to characterize the number of babies that remains alive in EAG states of India, under both SRS and RSS.

The event of survivorship of a child is positively correlated with mother's age (Finlay et.al (2011) and Selemani et.al (2014)). One can say that chance of survivorship of a child is low in mother of lower ages than that of those of higher ages. So mother's age (in months) is used as an auxiliary variable for ranking purpose in ranked set sampling. The procedure adopted for sampling through RSS has discussed below (Das et al. 2017):

1. A simple random sample of s^2 units, say $X_i; i = 1,2,\cdots,s$, is drawn from the target population and are randomly partitioned into s sets each having s units, , say X_{rj} for all $r = 1,2,\cdots,s;\ j = 1,2,\cdots,s$.

2. In each of s sets the units are ranked according to the mother's age, denoted as $X_{[r]j}$. In situation of ties the observations are ordered systematically in the sequence, as discussed by Terpstra and Nelson (2005)

3. From the first set, the unit corresponding to the mother with lowest age is selected ($X_{[1]1}$). From the second set, the unit corresponding to mother with second lowest age is selected ($X_{[2]2}$) and so on. Finally, from the s^{th} set, the unit corresponding to the mother with highest age ($X_{[s]s}$) is selected. The remaining $s(s-1)$ sampled units are discarded from the data set.

4. The Steps 1 - 3, called a cycle, are repeated m times to obtain a ranked set sample of size $n = ms$.

Using the sample we have computed various moments discussed in previous sections under both SRS and RSS. The results have reported in Tables 3 and 4. When the number of cycles m has assumed as a fixed quantity, the obtained result have shown that characterization of the sums based on ranked set sampling for all states are much reliable and its efficiency lies to 10-34%. The kurtosis based on both simple random sample and ranked set sample have found closer to 3, but significant deviation from 0 have observed in the skewness (negatively skewed). Under the assumption that m is a random quantity, the efficiency increases 3 to 4 times that of earlier case. It has also observed that the variance under RSS is converging towards the mean, which shows the asymptotic convergence to possion distribution. The kurtosis based on both simple random sample and ranked set sample have found far away from 3 and a significant deviation from 0 have observed in the skewness (positively skewed). Statistical Analysis System (SAS) package, University edition has used for sampling units and all other computation is carried out by using R package (version-3.0.3).

7. Conclusion

The goal of present article is to characterize a sum of independent and non-identical set of binomial ranked set samples and compare it with a sum of independent and identical binomial simple random samples for two different situations based on the number of cycles, which is first pre-assumed as a constant integer and when it is a random variable. Our comparison depends only on establishing the variability and their behaviour using some moments. Results show that the sum based on ranked set samples, which is same as that of simple random sample, are more precise and achieve asymptotic normality using comparatively with smaller sample than that of simple random sample. In the context of real-life data study related to child's survivorship in selected eight EAG Indian states, it is found that RSS provides much reliable and accurate estimates than that of SRS for all selected states taken into account.

Acknowledgement

The first author would like to express his deepest gratitude and sincere thanks to the Department of Science & Technology, India (Grant No - IF130365) for funding, while working at Gauhati University. The authors would like to thank the anonymous referees and editorial board for their constructive comments and suggestions to improve the quality of this manuscript.

REFERENCES

AHUJA, J. C., (1970). On the distribution of sum of independent positive binomial variables. Canad. Math. Bull, 13(1), pp. 151–152.

BAI, J., NG, S., (2005)., Tests for skewness, kurtosis, and normality for time series data. Journal of Business & Economic Statistics, 23(1), pp. 49–60.

BARBOUR, A. D., HOLST, L., (1989). Some applications of the Stein-Chen method for proving Poisson convergence. Advances in Applied Probability, 21(1), pp. 74–90.

BROWN, B. W., NEWEY, W. K., (2002). Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business & Economic Statistics, 20(4), pp. 507–517.

BUTLER, K., STEPHENS, M. A., (2017). The distribution of a sum of independent binomial random variables. Methodology and Computing in Applied Probability, 19(2), pp. 557–571.

CHABCHOUB, Y., FRICKER, C., GUILLEMIN, F., ROBERT, P., (2010). On the statistical characterization of flows in Internet traffic with application to sampling. Computer Communications, 33(1), pp. 103–112.

CHEN, H., (2008). Alternative Ranked set Sample Estimators for the Variance of a Sample Proportion. Applied Statistics Research Progress. Nova Publishers, pp. 35–38.

CHEN, L. H., ROLLIN, A., (2013). Approximating dependent rare events. Bernoulli, 19(4), pp. 1243-1267.

CHEN, H. STASNY E. A., WOLFE, D. A., (2005). Ranked Set Sampling for Efficient Estimation of a Population Proportion. Statistics in Medicine, 24, pp. 3319–3329.

CHEN, H., STASNY, E. A., WOLFE, D. A., (2008). Ranked set sampling for ordered categorical variables. Canadian Journal of Statistics, 36(2), pp. 179–191.

CHEN, H., STASNY, E. A., WOLFE, D. A., (2006). Unbalanced Ranked Set Sampling for estimating a Population Proportion. Biometrics, 62,, pp. 150–158.

CHEN, H., STASNY, E. A., WOLFE, D. A., (2007). Improved Procedures for Estimation of Disease Prevalence Using Ranked Set Sampling. Biometrical Journal, 49(4),, pp. 530–538.

CHEN, H., STASNY, E. A., WOLFE, D. A., MACEACHERN, S. N., (2009). Unbalanced Ranked Set Sampling for Estimating a Population Proportion Under Imperfect Rankings. Communications in Statistics: Theory and Methods, 38(12), pp. 2116–2125.

DAS, R. VERMA, V., NATH, C. N., (2017). Bayesian Estimation of Measles Vaccination Coverage under Ranked Sets Sampling. Statistics in Transition new series, 18(4), pp. 589–608.

FINLAY, J. E., OZALTIN, E., CANNING, D., (2011). The association of maternal age with infant mortality, child anthropometric failure, diarrhoea and anaemia for first births: evidence from 55 low-and middle-income countries. BMJ open, 1(2), e000226.

FREEDMAN, D. A., PETERS, S. C., (1984). Bootstrapping a regression equation: Some empirical results. Journal of the American Statistical Association, 79(385), pp. 97–106.

HASTINGS, W. K., (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1), pp. 97–109.

HORVATH, L., (1989). The limit distributions of likelihood ratio and cumulative sum tests for a change in a binomial probability. Journal of Multivariate Analysis, 31(1), pp. 148–159.

KADANE, J. B., (2016). Sums of possibly associated Bernoulli variables: The Conway-Maxwell-binomial distribution. Bayesian Analysis, 11(2), pp. 403–420.

LING, K. D., (1988). On binomial distributions of order k. Statistics & Probability Letters, 6(4), pp. 247–250.

MALIK, H. J., (1969). Distribution of the sum of truncated binomial variates. Canadian Mathematical Bulletin, 12(3), pp. 334–336.

PERCUS, O. E., PERCUS, J. K., (1985). Probability bounds on the sum of independent nonidentically distributed binomial random variables. SIAM Journal on Applied Mathematics, 45(4), pp. 621–640.

SASON, I., (2013). Entropy bounds for discrete random variables via coupling. In 2013 IEEE International Symposium on Information Theory (pp. 414–418). IEEE.

SELEMANI, M., MWANYANGALA, M. A., MREMA, S., SHAMTE, A., KAJUNGU, D., MKOPI, A., NATHAN, R., (2014). The effect of mother?s age and other related factors on neonatal survival associated with first and second birth in rural, Tanzania: evidence from Ifakara health and demographic surveillance system in rural Tanzania. BMC pregnancy and childbirth, 14(1), p. 240.

SMALL, N. J. H., (1980). Marginal skewness and kurtosis in testing multivariate normality. Applied Statistics, pp. 85–87.

SUNKLODAS, J. K., (2014). On the normal approximation of a binomial random sum. Lithuanian Mathematical Journal, 54(3), pp. 356–365.

TAKACS, L., (1991). A Bernoulli excursion and its various applications. Advances in Applied Probability, 23(3), pp. 557–585.

TERPSTRA, J. T., (2004). On estimating a population proportion via ranked set sampling. Biometrical Journal, 46,, pp. 264–272.

TERPSTRA, J. T., MILLER, Z. A., (2006). Exact Inference for a Population

Proportion Based on a Ranked Set Sample, Communications in Statistics: Simulation and Computation, 35(1), pp. 19–27.

TERPSTRA, J. T., NELSON, E. J., (2005). Optimal Rank Set Sampling Estimates for a Population Proportion, Journal of Statistical Planning and Inference, 127, pp. 309–321.

VERMA, V., NATH, D. C., DAS, R., (2017). Bayesian bounds for population proportion under ranked set sampling. Communications in Statistics-Simulation and Computation, pp. 1–16. DOI: 10.1080/03610918.2017.1387659.

YU, C., ZELTERMAN, D., (2002). Sums of dependent Bernoulli random variables and disease clustering. Statistics & probability letters, 57(4), pp. 363–373.

APPENDIX

INTER-RELATIONSHIP BETWEEN FACTORIAL, RAW AND CENTRAL MOMENTS

$$
\begin{aligned}
\mu'_{[1]} &= \mu'_1 = \mu_1 \\
\mu'_{[2]} &= \mu'_2 - \mu'_1 \; ; \mu'_2 = \mu'_{[2]} + \mu'_1 \; ; \; \mu_2 = \mu'_2 - \mu_1'^2 \\
\mu'_{[3]} &= \mu'_3 - 3\mu'_2 + 2\mu'_1 \; ; \; \mu'_3 = \mu'_{[3]} + 3\mu'_2 - 2\mu'_1 \; ; \; \mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^3 \\
\mu'_{[4]} &= \mu'_4 - 6\mu'_3 + 11\mu'_2 - 6\mu'_1 \; ; \; \mu'_4 = \mu'_{[4]} + 6\mu'_3 - 11\mu'_2 + 6\mu'_1; \\
& \quad \mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6\mu_1'^2\mu'_2 - 3\mu_1'^4
\end{aligned}
$$

7.1. Case-I: The number of cycles, m, is a known fixed value

Using equation-(3.1) the factorial moments based on simple random sample are given by

$$
\begin{aligned}
\mu'_{[1]}(W) &= \frac{n!}{(n-1)!}p \\
\mu'_{[2]}(W) &= \frac{n!}{(n-2)!}p^2 \\
\mu'_{[3]}(W) &= \frac{n!}{(n-3)!}p^3 \\
\mu'_{[4]}(W) &= \frac{n!}{(n-4)!}p^4
\end{aligned}
$$

Using equation-(3.2) the factorial moments based on ranked set sample are given by

$$
\begin{aligned}
\mu'_{[1]}(Z) &= m\left(\sum_{i=1}^{s} p_{[i]}\right) \\
\mu'_{[2]}(Z) &= m\left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[1]} - m\left(\sum_{i=1}^{s} p_{[i]}^2\right) \\
\mu'_{[3]}(Z) &= m\left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[2]} - 2m\left(\sum_{i=1}^{s} p_{[i]}^2\right)\mu'_{[1]} + 2m\left(\sum_{i=1}^{s} p_{[i]}^3\right) \\
\mu'_{[4]}(Z) &= m\left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[3]} - 3m\left(\sum_{i=1}^{s} p_{[i]}^2\right)\mu'_{[2]} + 6m\left(\sum_{i=1}^{s} p_{[i]}^3\right)\mu'_{[1]} - 6m\left(\sum_{i=1}^{s} p_{[i]}^4\right)
\end{aligned}
$$

$$\mu'_{[k]}(W) = s\sum_{j=0}^{k-1} \frac{(k-1)!}{(k-1-j)!} p^{(j+1)} \left((-1)^j + \frac{\lambda(s-1)!}{(s-1-j)!j!} \right) \mu'_{[k-1-j]}(W). \quad (7.1)$$

7.2. Case-II: The number of cycles, m, is a random variable

Using equation-(3.9) the first four factorial moments based on simple random samples, where $N \sim$ Poisson (λ) ;$N = 1,2,\cdots$, are given by

$$\begin{aligned}
\mu'_{[1]}(W) &= sp(1+\lambda) \\
\mu'_{[2]}(W) &= s\{p(1+\lambda)\mu'_{[1]} + p^2(-1+\lambda(s-1))\} \\
\mu'_{[3]}(W) &= s\{p(1+\lambda)\mu'_{[2]} + 2p^2(-1+\lambda(s-1))\mu'_{[1]} + 2p^3(1+0.5\lambda(s-1)(s-2))\} \\
\mu'_{[4]}(W) &= s\{p(1+\lambda)\mu'_{[3]} + 3p^2(-1+\lambda(s-1))\mu'_{[2]} \\
&\quad +6p^3(1+\frac{\lambda(s-1)(s-2)}{2})\mu'_{[1]} + 6p^3(-1+\frac{\lambda(s-1)(s-2)(s-3)}{3!})\ \}
\end{aligned}$$

$$\mu'_{[k]}(Z) = \lambda\left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[k-1]}(Z) + \sum_{j=0}^{k-1}(-1)^j \frac{(k-1)!}{(k-1-j)!}\left(\sum_{i=1}^{s} p_{[i]}^{j+1}\right)\mu'_{[k-1-j]}(Z). \quad (7.2)$$

Using equation-(3.10) the first four factorial moments based on ranked set samples, where $N \sim$ Poisson (λ) ;$N = 1,2,\cdots$, are given by

$$\begin{aligned}
\mu'_{[1]}(Z) &= a + \left(\sum_{i=1}^{s} p_{[i]}\right) \\
\mu'_{[2]}(Z) &= a\mu'_{[1]} + \left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[1]} - \left(\sum_{i=1}^{s} p_{[i]}^2\right) \\
\mu'_{[3]}(Z) &= a\mu'_{[2]} + \left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[2]} - 2\left(\sum_{i=1}^{s} p_{[i]}^2\right)\mu'_{[1]} + 2\left(\sum_{i=1}^{s} p_{[i]}^3\right) \\
\mu'_{[4]}(Z) &= a\mu'_{[3]} + \left(\sum_{i=1}^{s} p_{[i]}\right)\mu'_{[3]} - 3\left(\sum_{i=1}^{s} p_{[i]}^2\right)\mu'_{[2]} + 6\left(\sum_{i=1}^{s} p_{[i]}^3\right)\mu'_{[1]} - 6\left(\sum_{i=1}^{s} p_{[i]}^4\right)
\end{aligned}$$

where $a = \lambda sp$ and $p = \frac{1}{s}\sum_{r=1}^{s} p_{[r]}$.

Table 1: Mean and relative precision of sum of Binomial variate under SRS and RSS, for the given set size s, m and p.

p	m	s=2			s=4			s=6		
		Mean		RE	Mean		RE	Mean		RE
		SRS	RSS		SRS	RSS		SRS	RSS	
0.1	10	3	3	1.21	3	3	1.32	5	5	1.39
	50	14	14	1.19	18	18	1.23	29	29	1.53
	100	16	16	1.10	42	42	1.23	46	46	1.60
	200	38	38	1.12	83	83	1.36	103	103	1.42
	500	101	101	1.11	196	196	1.29	273	273	1.38
0.2	10	5	5	1.50	7	7	1.41	12	12	1.66
	50	20	20	1.25	42	42	1.98	58	58	1.51
	100	44	44	1.20	89	89	1.66	139	139	2.17
	200	91	91	1.22	157	157	1.48	234	234	1.82
	500	197	197	1.20	375	375	1.48	593	593	1.81
0.3	10	5	5	1.50	12	12	1.83	18	18	1.50
	50	31	31	1.33	60	60	1.99	92	92	2.33
	100	67	67	1.46	120	120	1.63	175	175	1.98
	200	117	117	1.19	247	247	1.54	356	356	2.08
	500	298	298	1.27	597	597	1.70	899	899	2.02
0.4	10	3	3	1.21	23	23	2.51	29	29	2.11
	50	39	39	1.29	75	75	1.60	122	122	2.07
	100	79	79	1.10	168	168	1.79	240	240	2.18
	200	164	164	1.33	297	297	1.77	481	481	2.25
	500	422	422	1.30	792	792	1.91	1205	1205	2.26
0.5	10	10	10	1.56	17	17	2.51	28	28	4.39
	50	49	49	1.33	107	107	2.38	154	154	1.99
	100	108	108	1.24	212	212	1.78	297	297	2.53
	200	204	204	1.43	400	400	1.83	596	596	2.05
	500	508	508	1.38	1000	1000	1.89	1509	1509	2.31

Table 2: Marginal mean and relative precision of sum of Binomial variate under SRS and RSS, for the given set size s, λ and p.

p	λ	s=2			s=4			s=6		
		Mean		RE	Mean		RE	Mean		RE
		SRS	RSS		SRS	RSS		SRS	RSS	
0.1	10	2.2	2.2	1.10	4.4	4.4	1.30	6.6	6.6	1.50
	50	10.2	10.2	1.10	20.4	20.4	1.30	30.6	30.6	1.50
	100	20.2	20.2	1.10	40.4	40.4	1.30	60.6	60.6	1.50
	200	40.2	40.2	1.10	80.4	80.4	1.30	120.6	120.6	1.50
	500	100.2	100.2	1.10	200.4	200.4	1.30	300.6	300.6	1.50
0.2	10	4.4	4.4	1.20	8.8	8.8	1.60	13.2	13.2	1.99
	50	20.4	20.4	1.20	40.8	40.8	1.60	61.2	61.2	2.00
	100	40.4	40.4	1.20	80.8	80.8	1.60	121.2	121.2	2.00
	200	80.4	80.4	1.20	160.8	160.8	1.60	241.2	241.2	2.00
	500	200.4	200.4	1.20	400.8	400.8	1.60	601.2	601.2	2.00
0.3	10	6.6	6.6	1.30	13.2	13.2	1.89	19.8	19.8	2.49
	50	30.6	30.6	1.30	61.2	61.2	1.90	91.8	91.8	2.50
	100	60.6	60.6	1.30	121.2	121.2	1.90	181.8	181.8	2.50
	200	120.6	120.6	1.30	241.2	241.2	1.90	361.8	361.8	2.50
	500	300.6	300.6	1.30	601.2	601.2	1.90	901.8	901.8	2.50
0.4	10	8.8	8.8	1.40	17.6	17.6	2.19	26.4	26.4	2.98
	50	40.8	40.8	1.40	81.6	81.6	2.20	122.4	122.4	3.00
	100	80.8	80.8	1.40	161.6	161.6	2.20	242.4	242.4	3.00
	200	160.8	160.8	1.40	321.6	321.6	2.20	482.4	482.4	3.00
	500	400.8	400.8	1.40	801.6	801.6	2.20	1202.4	1202.4	3.00
0.5	10	11.0	11.0	1.49	22.0	22.0	2.48	33.0	33.0	3.47
	50	51.0	51.0	1.50	102.0	102.0	2.50	153.0	153.0	3.49
	100	101.0	101.0	1.50	202.0	202.0	2.50	303.0	303.0	3.50
	200	201.0	201.0	1.50	402.0	402.0	2.50	603.0	603.0	3.50
	500	501.0	501.0	1.50	1002.0	1002.0	2.50	1503.0	1503.0	3.50

Table 3: State-wise mean, variance, skewness, kurtosis and relative precision of sum of Binomial variate under SRS and RSS, for the given set size s, m and p.

State	p	s	m	Mean	SRS			RSS			RE
					Vari	Skew	Kurt	Vari	Skew	Kurt	
Bihar	0.94	4	140	523	34.56	-0.148	3.018	29.11	-0.108	3.001	1.19
Uttaranchal	0.95	4	75	285	14.25	-0.238	3.050	12.96	-0.204	3.025	1.10
Chhatisgarh	0.92	4	90	328	29.16	-0.152	3.018	22.56	-0.091	2.993	1.29
Jharkhand	0.93	4	100	373	25.18	-0.172	3.025	21.15	-0.126	3.002	1.19
Orissa	0.93	5	60	278	20.39	-0.189	3.029	14.63	-0.090	2.981	1.39
Rajasthan	0.93	5	70	324	24.07	-0.174	3.024	19.34	-0.116	2.997	1.24
Madhya Pr.	0.93	5	100	465	32.55	-0.151	3.019	26.23	-0.101	2.998	1.24
Uttar Pr.	0.91	5	80	367	30.28	-0.152	3.018	23.44	-0.093	2.995	1.29

Table 4: State-wise marginal mean, variance, skewness, kurtosis and relative precision of sum of Binomial variate under SRS and RSS, for the given set size s, λ and p.

State	p	s	λ	Mean	SRS			RSS			RE
					Vari	Skew	Kurt	Vari	Skew	Kurt	
Bihar	0.94	4	140	527.74	1995.18	0.086	3.008	524.20	0.043	3.021	3.81
Uttaranchal	0.95	4	75	288.80	1097.44	0.117	3.014	285.17	0.058	3.041	3.85
Chhatisgarh	0.92	4	90	335.69	1250.82	0.108	3.012	332.23	0.054	3.033	3.76
Jharkhand	0.93	4	100	374.71	1403.58	0.102	3.011	371.22	0.051	3.030	3.78
Orissa	0.93	5	60	284.67	1325.64	0.131	3.017	280.25	0.059	3.049	4.73
Rajasthan	0.93	5	70	328.63	1524.07	0.122	3.015	324.27	0.055	3.042	4.70
Madhya Pr.	0.93	5	100	469.65	2195.13	0.102	3.010	465.26	0.046	3.029	4.72
Uttar Pr.	0.91	5	80	369.56	1697.65	0.114	3.013	365.30	0.052	3.036	4.65

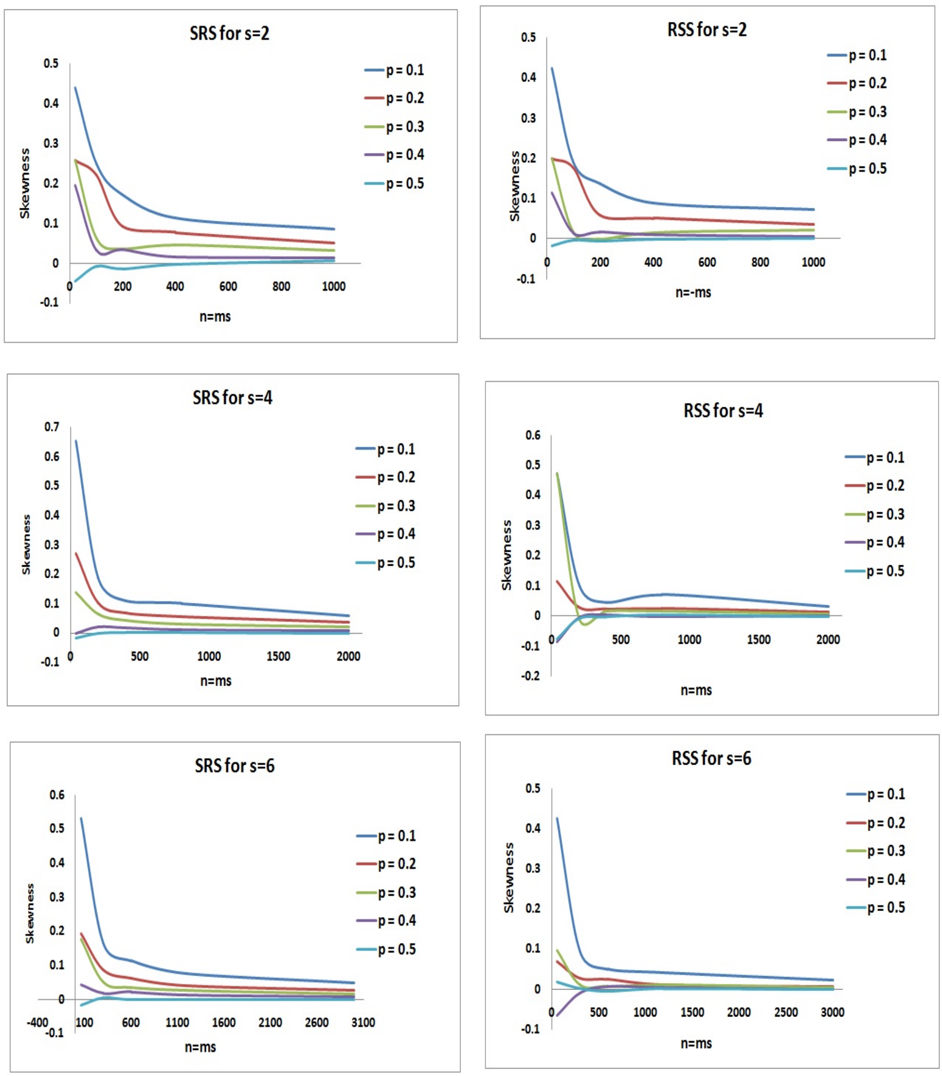

Figure 1: Skewness pattern of sum of Binomial variable under SRS and RSS for set size $s = 2,4,6$ and sample size $n = ms$, where $m = \{10,50,100,200,500\}$, and for fixed p.

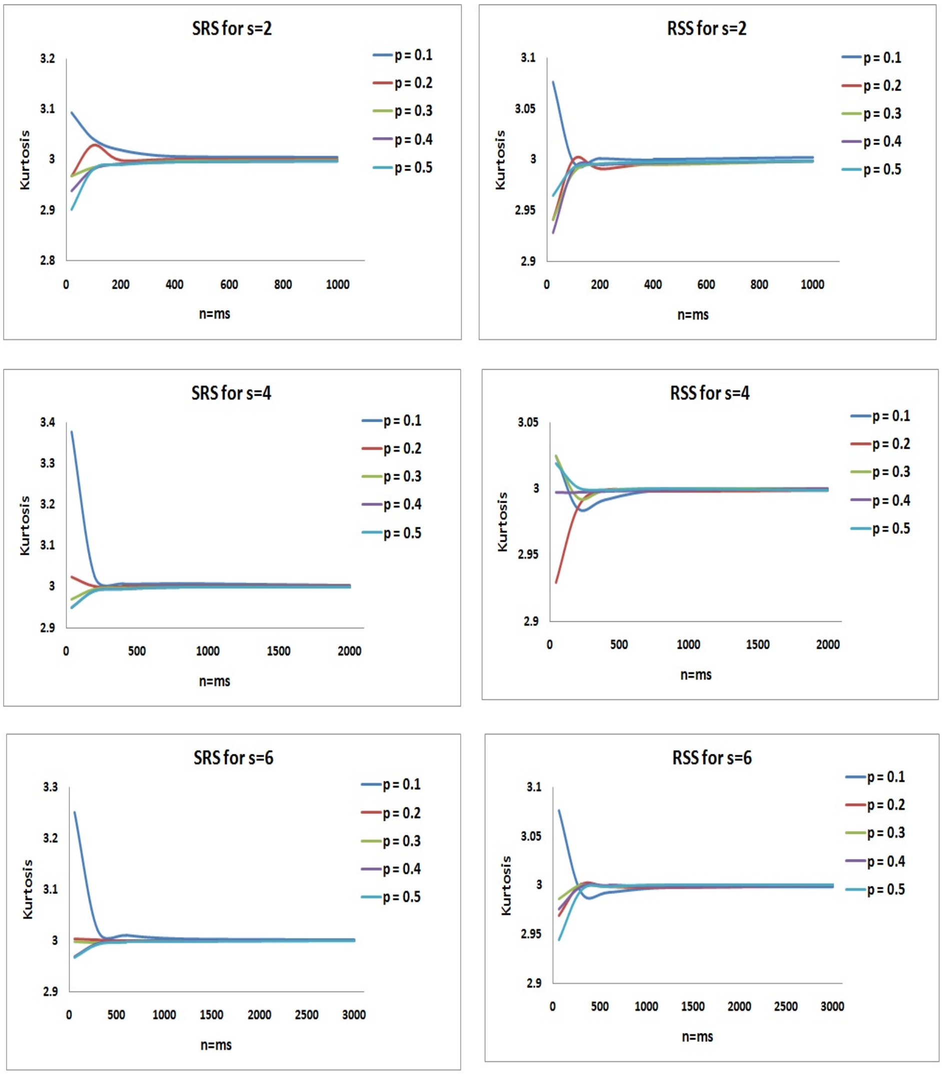

Figure 2: Kurtosis pattern of sum of Binomial variable under SRS and RSS for set size $s = 2,4,6$ and sample size $n = ms$, where $m = \{10,50,100,200,500\}$, and for fixed p.

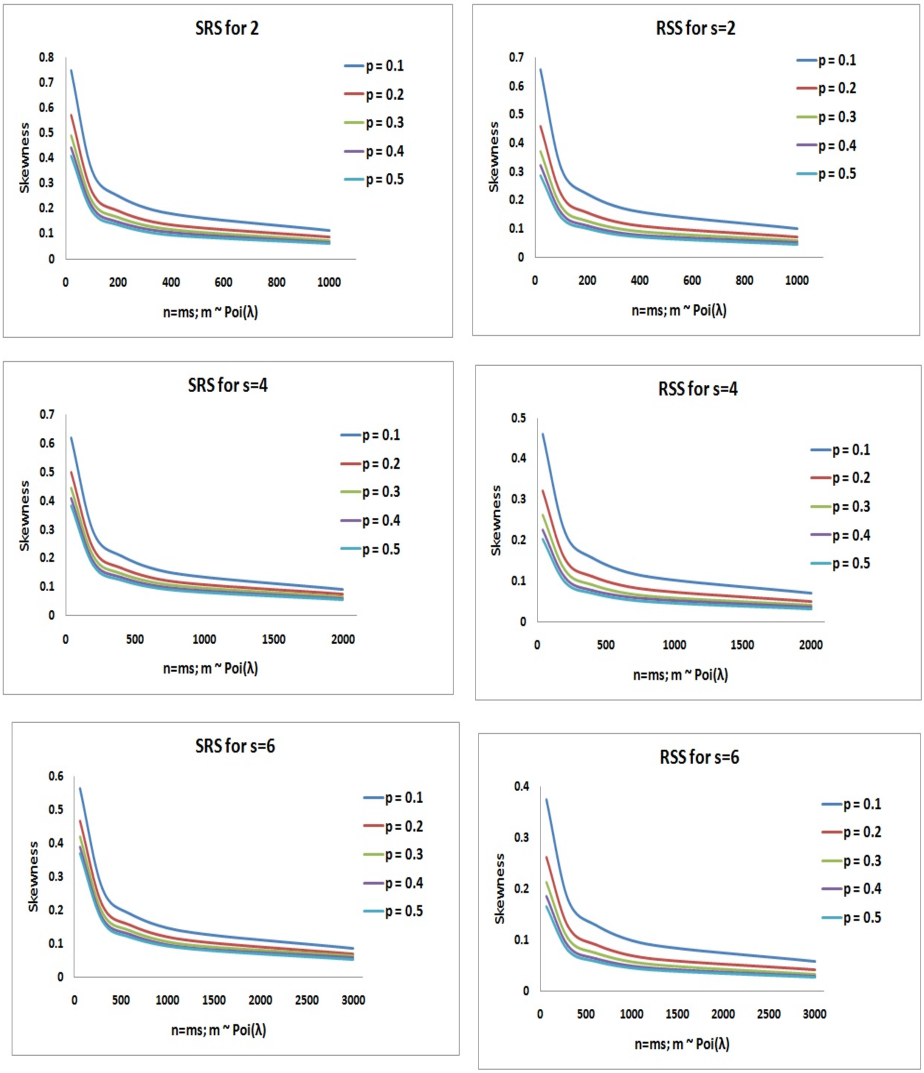

Figure 3: Marginal skewness pattern of sum of Binomial variable under SRS and RSS for set size $s = 2, 4, 6$ and sample size $n = ms$, where $m \sim Poisson(\lambda)$, $\lambda = \{10, 50, 100, 200, 500\}$, and for fixed p.

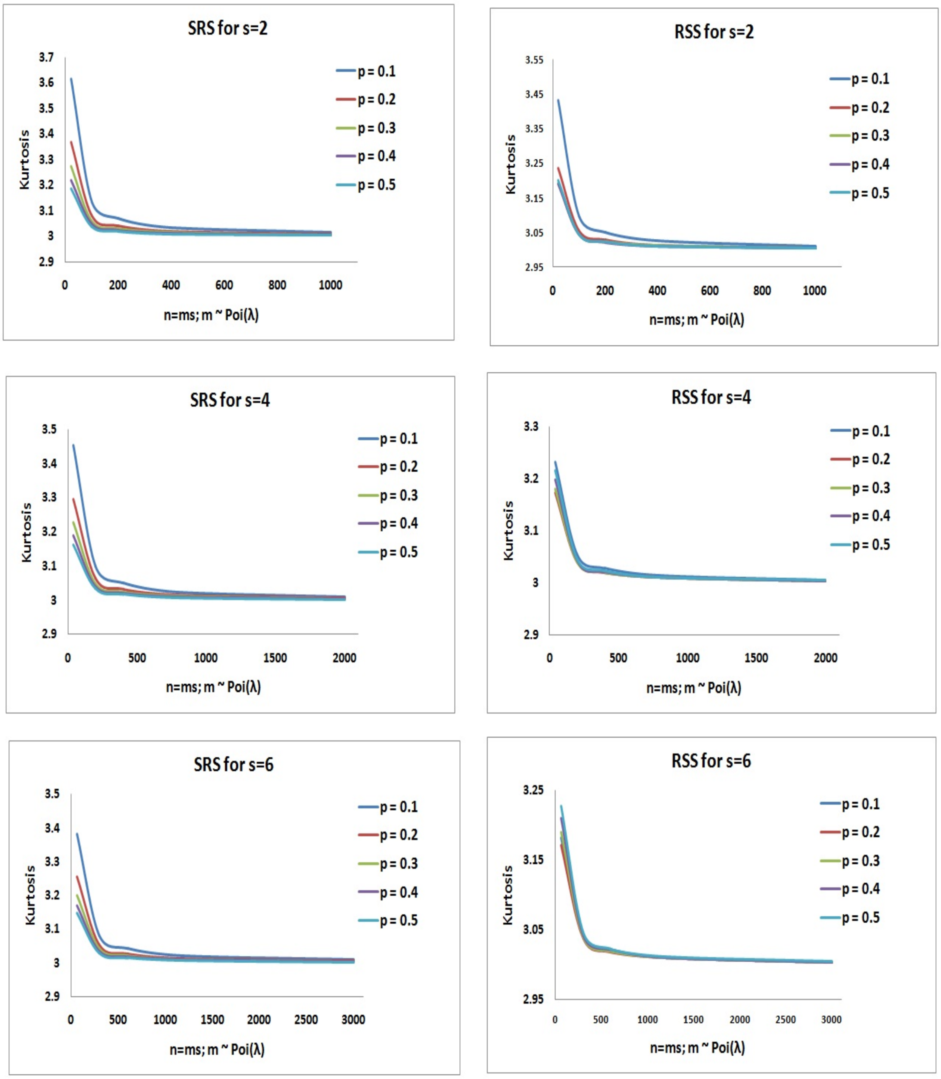

Figure 4: Marginal kurtosis pattern of sum of Binomial variable under SRS and RSS for set size $s = 2,4,6$ and sample size $n = ms$, where $m \sim Poisson(\lambda)$, $\lambda = \{10, 50, 100, 200, 500\}$, and for fixed p.