ProbStat Forum, Volume 08, October 2015, Pages 157–164
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

An asymptotic comparison of MLE and MME of the parameter in a new discrete distribution analogous to Burr distribution

G. Nanjundan[a], T Raveendra Naika[b]

[a]*Department of Statistics, Bangalore University, Bangalore - 560 056, India.*
[b]*Maharani's Science College for Women, Bangalore 560 001, India.*

Abstract. Estimation of parameter in a new discrete distribution which is analogous to a Burr distribution is discussed in this paper. The maximum likelihood and the method of moment estimators are obtained. The asymptotic normality of the moment estimator is established. The asymptotic relative efficiency of the maximum likelihood estimator over the moment estimator is computed. It is illustrated that the new distribution fits better than the Poisson distribution to a clinical trial data set. A simulation study has been carried out to demonstrate the asymptotic normality of the estimators.

1. Introduction

Sreehari (2010) has characterized a class of discrete distributions which turns out to be an analogue of Burr (1942) family. The probability mass function (pmf) of the random variable X of the d-th distribution of the class characterized by Sreehari (2010) is

$$p(x,\theta) = \begin{cases} (x+\theta-1)\frac{\theta^x}{(x+1)!}, x = 0,1,2,\ldots \\ 0, \qquad \text{otherwise,} \end{cases} \tag{1}$$

$0 < \theta < 1$. The distribution function of X is

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \frac{\theta^{[x+1]}}{([x+1])!}, & x \geq 0. \end{cases} \tag{2}$$

We refer to this distribution as $S(d)$- distribution.

The mean of the distribution is given by

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} xp(x) \\ &= \sum_{x=0}^{\infty} \theta \frac{\theta^{x-1}}{(x-1)!} - \sum_{x=0}^{\infty} (x+1)\frac{\theta^{x+1}}{(x+1)!} + \sum_{x=0}^{\infty} \frac{\theta^{x+1}}{(x+1)!} \\ E(X) &= \theta e^{\theta} - \theta(e^{\theta}-1) + \left(e^{\theta} - (1-\theta)\right). \end{aligned}$$

2010 *Mathematics Subject Classification.*
Keywords. Maximum likelihood and moment estimators, Fisher information, asymptotic relative efficiency.
Received: 31 May 2014; Accepted: 25 October 2015
Email addresses: nanzundan@gmail.com (G. Nanjundan), raviisec@yahoo.co.uk (T Raveendra Naika)

Therefore $E(X) = e^{\theta} - 1.$

Now,

$$E[X(X+1)] = \sum_{x=0}^{\infty} x(x+1)p(x)$$
$$= \sum_{x=0}^{\infty} (x-1)\frac{\theta^x}{(x-1)!} + 2\theta \sum_{x=0}^{\infty} \frac{\theta^{x-1}}{(x-1)!} - \theta^2 \sum_{x=0}^{\infty} \frac{\theta^{x-1}}{(x-1)!}$$
$$= \theta^2 \sum_{x=0}^{\infty} \frac{\theta^{x-2}}{(x-2)!} + 2\theta e^{\theta} - \theta^2 \sum_{x=0}^{\infty} \frac{\theta^{x-1}}{(x-1)!}$$
$$= \theta^2 e^{\theta} + 2\theta e^{\theta} - \theta^2 e^{\theta}.$$

$$E[X(X+1)] = 2\theta e^{\theta}.$$

Hence $E\left[X^2\right] = 2\theta e^{\theta} - e^{\theta} + 1.$

In turn, $V(X) = e^{\theta}\left(2\theta - e^{\theta} + 1\right)$. We get $E(X) - V(X) = e^{2\theta} - 2\theta e^{\theta} - 1$.
Consider $e^{2\theta} - 1 = \sum_{r=1}^{\infty} \frac{(2\theta)^r}{r!} = 2\theta \sum_{r=1}^{\infty} \frac{2^{(r-1)}\theta^{(r-1)}}{r!} > 2\theta \sum_{r=1}^{\infty} \frac{r\theta^{(r-1)}}{r!} = 2\theta \sum_{r=1}^{\infty} \frac{\theta^{(r-1)}}{(r-1)!}$.
Therefore $e^{2\theta} - 1 > 2\theta e^{\theta}$. Hence $E(X) - V(X) > 0$.

Note that the pmf of $S(d)$- distribution is similar to that of Poisson distribution in structure. But the mean and the variance of Poisson distribution are equal. Hence $S(d)$- distribution is a suitable model for the data exhibiting under dispersion.

2. Maximum likelihood estimation

If $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a random sample on X having the pmf specified in (1), then the likelihood function becomes

$$L(\theta|x) = \prod_{j=1}^{n} P(X = x_j)$$
$$L(\theta|x) = \prod_{j=1}^{n} (x_j + \theta - 1)\frac{\theta^{x_j}}{(x_j+1)!}.$$

The log-likelihood function is

$$\log L(\theta|x) = \sum_{j=0}^{n} \log(x_j + 1 - \theta) + \sum_{j=0}^{n} x_j \log\theta - constant.$$

The likelihood equation is

$$\frac{d\log L(\theta|x)}{d\theta} = -\sum_{j=0}^{n} \frac{1}{(x_j + 1 - \theta)} + \frac{\sum_{j=0}^{n} x_j}{\theta} = 0. \tag{3}$$

The maximum likelihood estimate (MLE) of θ is the solution of

$$\sum_{j=0}^{n} \frac{\theta}{(x_j + 1 - \theta)} - n\bar{x} = 0. \tag{4}$$

The likelihood equation does not yield a closed form expression for the MLE of θ. Hence a numerical procedure like Newton-Raphson method can be employed to compute it.

3. Fisher information

When X has the pmf specified in (1),

$$\log p(x,\theta) = \log(x+1-\theta) + x\log\theta - \log(x+1)!$$

Also

$$\frac{d\log p(x,\theta)}{d\theta} = \frac{-1}{x+1-\theta} + \frac{x}{\theta}$$

$$\frac{d^2\log p(x,\theta)}{d\theta^2} = \frac{-1}{(x+1-\theta)^2} - \frac{x}{\theta^2}.$$

Hence

$$E[\frac{d^2\log p(x,\theta)}{d\theta^2}] = -\sum_{x=0}^{\infty} \frac{1}{(x+1-\theta)^2} p(x) - \sum_{x=0}^{\infty} \frac{x}{\theta^2} p(x)$$

$$E[\frac{d^2\log p(x,\theta)}{d\theta^2}] = -\sum_{x=0}^{\infty} \frac{1}{(x+1-\theta)} \frac{\theta^x}{(x+1)!} - \frac{1}{\theta^2} E(X)$$

$$E[\frac{d^2\log p(x,\theta)}{d\theta^2}] = -\sum_{x=0}^{\infty} \frac{1}{(x+1-\theta)} \frac{\theta^x}{(x+1)!} - \frac{1}{\theta^2} (e^\theta - 1).$$

The Fisher information becomes

$$I(\theta) = \sum_{x=0}^{\infty} \frac{1}{(x+1-\theta)} \frac{\theta^x}{(x+1)!} + \frac{1}{\theta^2}(e^\theta - 1).$$

Since $0 < \theta < 1$, $1 - \theta > 0$. Hence

$$\sum_{x=0}^{\infty} \frac{1}{(x+1-\theta)} \frac{\theta^x}{(x+1)!} << \sum_{x=1}^{\infty} \frac{1}{x} \frac{\theta^x}{(x+1)!} << \sum_{x=1}^{\infty} \frac{\theta^x}{(x+1)!} = \frac{1}{\theta}(e^\theta - 1) < \infty.$$

Therefore $\sum_{x=0}^{\infty} \frac{1}{(x+1-\theta)} \frac{\theta^x}{(x+1)!}$ is convergent and it can be evaluated numerically.

Note that

i) the support $S = \{x : p(x,\theta) > 0\}$ of $p(x,\theta)$ does not depend on the parameter θ

ii) the parameter space (0, 1) is an open interval

iii) $\log p(x,\theta)$ can be differentiated thrice w..r.t. θ

iv) $\sum_{x=0}^{\infty} p(x,\theta) = 1$ is twice differentiable under the summation sign

v) there exists a function $M(x)$ such that $\left|\frac{d^3 \log p(x,\theta)}{d\theta^3}\right| \le M(x)$ and $E[M(X)] < \infty$.

In our case,
$\frac{d^3 \log p(x,\theta)}{d\theta^3} = \frac{-2}{(x+1-\theta)^3} + \frac{2x}{\theta^3}$. Therefore, $\left|\frac{d^3 \log p(x,\theta)}{d\theta^3}\right| \leq \left|\frac{2}{(x+1-\theta)^3}\right| + \left|\frac{2x}{\theta^3}\right|$. Since $0 < \theta < 1, 0 < 1-\theta < 1$. Hence $\left|\frac{d^3 \log p(x,\theta)}{d\theta^3}\right| \leq \left|\frac{2}{x^3}\right| + \left|\frac{2x}{\theta^3}\right| \leq \frac{2}{x} + \frac{2x}{\theta^3} \leq 2 + \frac{2x}{(\theta_0-\epsilon)^3} = M(x)$, $\forall \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon), \epsilon > 0, x \geq 1$, where θ_0 is the true value of the parameter. Since the parameter space is an open interval such a neighborhood of θ_0 exists. Evidently $E[M(X)] < \infty$.

Therefore $p(x,\theta)$ satisfies the regularity conditions of Cramer (1966) and it belongs to Cramer family. Hence if $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a random sample on X having the pmf specified in (1) and $\hat{\theta}_{mle}$ is the MLE of θ, then

$$\sqrt{n}(\hat{\theta}_{mle} - \theta) \xrightarrow{L} N(0, \frac{1}{I(\theta)}), \quad as \quad n \to \infty.$$

That is $\hat{\theta}_{mle}$ is consistent and asymptotically normal (CAN) for θ with the asymptotic variance $\frac{1}{I(\theta)}$.

4. Method of moment estimation

When $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a random sample on X having the pmf specified in (1), the moment estimator of θ is the solution of the equation $e^\theta - 1 = \bar{X}_n$. Hence the moment estimator of θ is

$$\hat{\theta}_{mme} = \log(\bar{X}_n + 1). \tag{5}$$

Since the pmf of X does not belong to exponential family, we need to establish the asymptotic normality of the moment estimator separately. The following theorem states the asymptotic normality of the moment estimator.

Theorem: If $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a random sample on X having the pmf specified in (1), then

$$\sqrt{n}(\hat{\theta}_{mle} - \theta) \xrightarrow{L} N\left(0, e^\theta(2\theta - e^\theta + 1)\frac{1}{(\theta+1)^2}\right), \quad \text{as} \quad n \to \infty.$$

Proof: Since $X_1, X_2, \ldots, X_n$ are i.i.d. with $E(X) = e^\theta - 1$ and $Var(X) = e^\theta(2\theta - e^\theta + 1) < \infty$, by Levy- Lindeberg central limit theorem

$$Z_n = \sqrt{n}(\bar{X} - (e^\theta - 1)) \xrightarrow{L} N(0, e^\theta(2\theta - e^\theta + 1)), \quad \text{as} \quad n \to \infty.$$

Take $g(x) = \log(x+1)$. Then $g'(x) = \frac{1}{(x+1)}$ is non-vanishing and continuous for $0 < x < 1$. hence stated in Mann-Wald (2000) theorem,

$$\sqrt{n}(\log(\bar{X}+1) - \theta)) \xrightarrow{L} N(0, e^\theta(2\theta - e^\theta + 1)\frac{1}{(\theta+1)^2}), \quad \text{as} \quad n \to \infty.$$

That is $\hat{\theta}_{mme}$ is CAN with the asymptotic variance $e^\theta(2\theta - e^\theta + 1)\frac{1}{(\theta+1)^2}$.

The asymptotic relative efficiency of the MLE over the MME is given by

$$ARE = \frac{Asymptotic \ variance \ of \ MME}{Asymptotic \ variance \ of \ MLE}.$$

Since the asymptotic variance $\frac{1}{I(\theta)}$ of the MLE does not have a closed form expression, it is computed for various values of θ and the ARE is shown in the following table.

The ARE of the MLE over the MME is uniformly greater than unity and therefore the MLE is asymptotically more efficient than the MME.

Table 1:

ARE of MLE over MME of the parameter

θ	0.1	0.2	0.3	0.4	0.5
ARE	1.3405	1.7759	2.3291	3.0257	3.8895
θ	0.6	0.7	0.8	0.9	0.95
ARE	4.934	6.1455	7.4479	8.6441	9.0807

5. Illustration

Sreehari (2010) has compared the fit of $S(d)$ and Poisson models for a clinical trial data set. A bioequivalence study was conducted for a test drug (T) and a reference drug (R) by administering them to 144 individuals using a two period, two sequence and two treatment crossover design. The time until maximum concentration (T_{max}) was one of the characteristics observed. Let T_{it} and T_{ir} respectively denote the T_{max} values corresponding to the i-th individual for the test and the reference drugs. Take $D_i = |T_{it} - T_{ir}|$. The individuals were administered the drugs and their blood concentrations were measured just prior to medication and at time points (in hours) 1, 2, 4, 6, 8, 10, 12, 14, 16, 20, 22, 24, 30, 48, 60, 72, 96, 120 after medication. Of the 144 individuals 7 did not complete the course of treatment. The observed values of D_i for the 137 individuals are shown in the following table.

Table 2:

The observed values of $D_i = |T_{it} - T_{ir}|$

x	0	1	2	3	Total
Frequency	75	46	14	2	137

For this observed distribution, mean = 0.583942 > variance = 0.534925 and it exhibits under dispersion. And $\hat{\theta}_{mme} = \log_e(\bar{x}_n + 1) = 0.4599$.

Table 3:

Clinical trial data set

x	Frequency	Expected Frequency	
		$S(d)$	Poisson
0	75	73.9915	76.4043
1	46	48.5192	44.6157
2	14	12.268	13.0265
3	2	2.22129	2.95354
Chi-square value		**0.41111**	**0.64342**

It is evident that $S(d)$ fits better than the Poisson model to this observed distribution.

But $S(d)$ model cannot be fitted to all under dispersed data. We have $0 < \theta < 1$ and hence $0 < E(X) < (e^\theta - 1)$. If $0 < \bar{x}_n < (e^\theta - 1)$ is violated, then it leads to $\hat{\theta}_{mme} = \log(\bar{x}_n + 1) > 1$ which is not meaningful. For example, consider the data (shown in Table 4) on the number of scintillations from a radioactive decay of Polonium reported by Rutherford et. al (1910). This data set has been reproduced in Santner et. al (1989).

For the observed data, mean = 3.871549 > variance = 3.694773. This exhibits under dispersion. But $\hat{\theta}_{mme} = \log(\bar{x_n} + 1) = 1.58341 (> 1)$, is not an appropriate estimate of θ.

Table 4:
Data on the number of Scintillations from a radioactive decay of Polonium

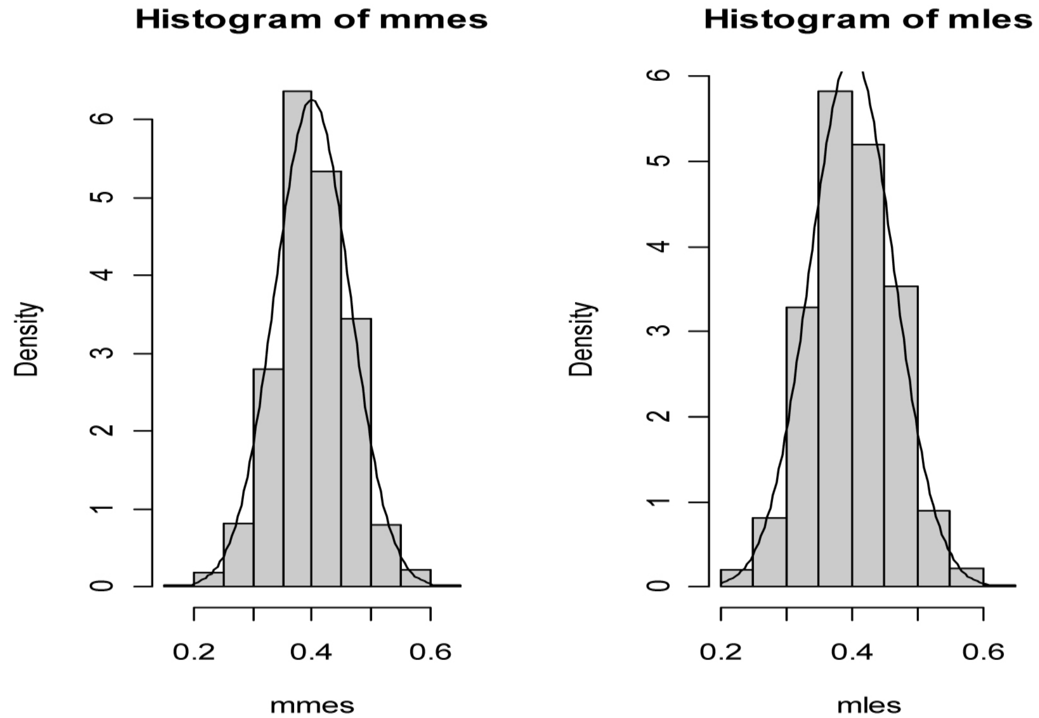
x	0	1	2	3	4	5	6	
f	57	203	383	525	532	408	273	
x	7	8	9	10	11	12	13	14
f	139	45	27	10	4	0	1	1

(x: number of scintillations, f: frequency)

6. Empirical study

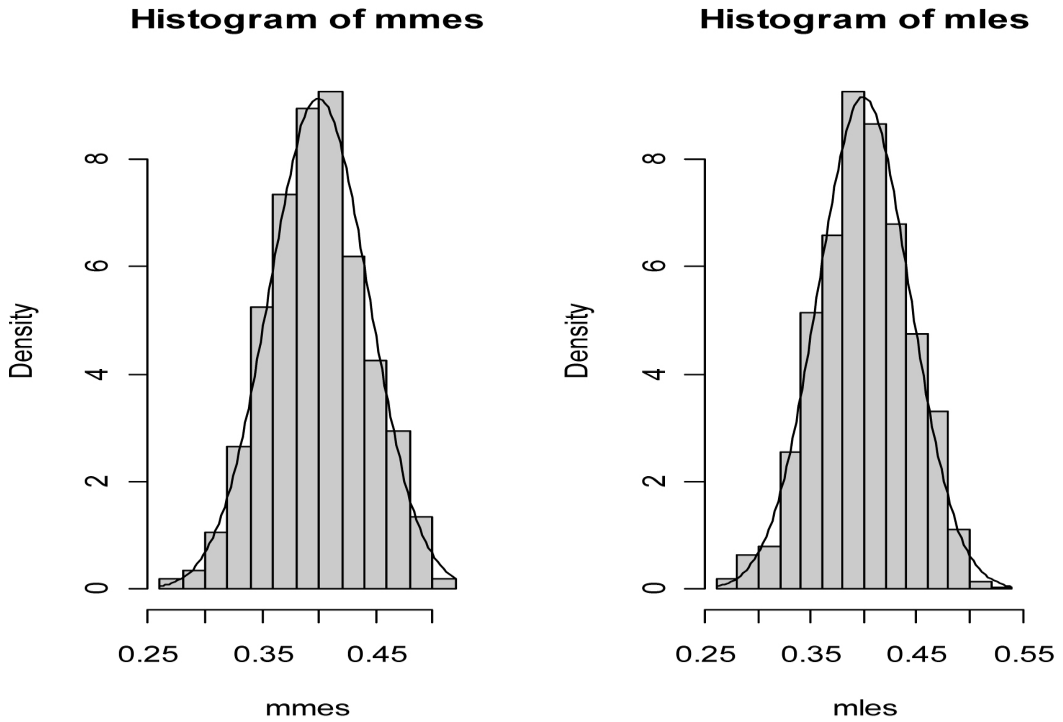
Random observations on X can be simulated using the distribution function (2). A modest simulation study has been carried out to study the performance of the MLE and the MME of θ. Using the R software, 1000 samples of size 50, 100, 150 were simulated for the specified value of $\theta = 0.4$ and the estimates were computed. The MLEs were obtained by solving the likelihood equation by Newton-Raphson method and the MMEs were taken as the initial estimates. The histograms of the estimates are displayed in the following Figures.

Figure 1: Histogram of the MMEs and the MLEs of θ under $S(d)$ model based on 1000 sample of size 50 for $\theta = 0.4$



These histograms give graphical evidence for the asymptotic normality of both the estimates. But the MLE approaches normality faster than the MME.

Figure 2: Histogram of the MMEs and the MLEs of θ under $S(d)$ model based on 1000 sample of size 100 for $\theta = 0.4$



7. Discussion and Conclusion

The $S(d)$-distribution is an appropriate alternate to Poisson model when the observed data exhibit under dispersion. Though the maximum likelihood estimator of the parameter of $S(d)$- distribution has no closed form expression, it can easily be computed by Newton-Raphson method. The moment estimator of the parameter has a closed form expression and easy to compute. Both the estimators are asymptotically normal. When computing facility is available, the MLE can be preferred to the MME, since the former approaches normality faster than the latter.

Acknowledgement

The authors are grateful to Prof. M. Sreehari for sending a softcopy of his paper and to Professors K. Suresh Chandra and B. Chandrasekar for useful comments.The authors are thankful to the referee for valuable suggestions which improved the presentation of the paper.

References

[1] I. W. Burr. (1942). Cumulative frequency functions, Annals of Mathematical Statistics, USA, 215-232.
[2] H. Cramer. (1966). Mathematical Methods of Statistics, Princeton, 11th Print, Princeton University Press, USA.
[3] B. K. Kale. (1999). A First Course on Parametric Inference, Narosa Publishing House, New Delhi.
[4] N. Mukhopadhyay. (2000). Probability and Statistical Inference, Marcel Dekker,New York.
[5] C. R. Rao. (1973). Linear Statistical Inference and Its Applications , John Wiley, New York .
[6] E. Rutherford and H. Geiger. (1910). The probability variations in the distribution of α particles , Phil, Mag. Sixth Ser., 20, pp. 698 - 704.
[7] T. J. Santner and D.E. Duffy. (1989). The Statistical Analysis of Discrete Data, New York, Springer Verlag.
[8] M. Sreehari. (2010). On a class of discrete distributions analogous to Burr family, Journal of Indian Statistical Association, Pune, 1, 48.

Figure 3: Histogram of the MMEs and the MLEs of θ under $S(d)$ model based on 1000 sample of size 150 for $\theta = 0.4$