MOMENTS OF MAXIMUM LIKELIHOOD ESTIMATORS IN THE DISCRETE CASE*

K.O. BOWMAN[1] and L.R. SHENTON[2]

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, P.O.Box 2008, 4500N, MS-6191, Oak Ridge, TN 37831-6191, USA,
bowmanko@ornl.gov

[2]Department of Statistics, University of Georgia, Athens, Georgia 30602, USA

September 7, 2005

Abstract

The distributions considered are the geometric, zero-truncated Poisson, logarithmic series, doubly truncated binomial and negative binomial distribution. For the doubly truncated binomial, an application is recalled. Maple program for the one parameter discrete distribution is included in the Appendix.

Key words and phrases: bias, kurtosis, low order moments, maximum likelihood estimation, negative binomial distribution, sample mean, skewness, variance.

1 Introduction

1.1 The basic structure

In Bowman and Shenton (2005) we have given Maple programs for the bias, variance, skewness and kurtosis of maximum likelihood estimators for probability functions involving s parameters. Briefly for categorized data

	Categories						
	1	2	$\cdots$	j	$\cdots$	w	$\sum_j$
Probability	p_1	p_2	$\cdots$	p_j	$\cdots$	p_w	1
Observed	n_1	n_2	$\cdots$	n_j	$\cdots$	n_w	N

and $p_j = p_j(\theta_1, \theta_2, \cdots, \theta_s) = p_j(\underline{\theta})$. Returning to the one parameter case, we have for the maximum likelihood estimator $\hat{\theta}$.

$$\hat{\theta} = f(n_1^*, n_2^*, \cdots, n_w^*) \tag{1}$$

where $n_j^* = n_j/N$, so that

$$E(n_j^* - p_j) = 0. \qquad (j = 1, 2, \cdots, w)$$

As described in Shenton and Bowman (1977, §3.3) we can set up a Taylor expansion for $\hat{\theta}$ in (1) and its powers to the fourth order. In the general case the expansion for $\hat{\theta}$ in (1) becomes an expansion for $\hat{\theta}_a$, where a relates to some s parameters.

1.2 A note on notation

We shall use μ_{sr} for the coefficient of N^{-r} in the sth central moment of a random variable. Thus, our interest is in $\mu_{11}(\hat{\theta})$, $\mu_{12}(\hat{\theta})$, $\sqrt{\beta_1(\hat{\theta})}$ and $\beta_2(\hat{\theta})$ where

$$\begin{cases} \mu_1'(\hat{\theta}) \sim \theta + \frac{\mu_{11}'}{N} + \frac{\mu_{12}'}{N^2} + \cdots & (N \to \infty) \\ \mu_2(\hat{\theta}) \sim \frac{\mu_{21}}{N} + \frac{\mu_{22}}{N^2} + \cdots \\ \mu_3(\hat{\theta}) \sim \frac{\mu_{32}}{N^2} + \cdots \\ \mu_4(\hat{\theta}) \sim \frac{3\mu_{21}^2}{N^2} + \frac{\mu_{43}}{N^3} + \cdots \\ \sqrt{\beta_1(\hat{\theta})} \sim \frac{\mu_{32}}{\mu_{21}^{3/2}}/\sqrt{N} + \cdots \\ \beta_2(\hat{\theta}) = \frac{\mu_4}{\mu_2^2} \sim 3 + \frac{K}{N} \cdots, \quad (K = \frac{\mu_{43}}{\mu_{21}^2} - 6\frac{\mu_{22}}{\mu_{21}}). \end{cases}$$

In general these moments refer to the random variable itself or to the maximum likelihood estimator $\hat{\theta}$.

1.3 Modified discrete distributions

A detailed account of modifications is given in Johnson, Kotz and Kemp (1992, §4 and §8 in particular).

(i) Left and right truncated

$$Pr(X = x) = p(x, \theta) / \left\{ \sum_{x=r_1}^{r_2} p(x, \theta) \right\} \qquad (x = r_1, r_1 + 1, \cdots, r_2; r_2 > r_1 \geq 0).$$

(ii) Some left truncated or some right truncated discrete distributions

$$Pr(X = x) = p(x, \theta) / \left\{ \sum_{x=0}^{r} p(x, \theta) \right\} \qquad (x = 0, 2, \cdots, r; r > x \geq 0).$$

$$Pr(X = x) = p(x, \theta) / \left\{ \sum_{x=r}^{\infty} p(x, \theta) \right\} \qquad (x = r, \cdots, \infty; r \leq x < \infty).$$

Tippett (1932) assumed a modified Poisson distribution of this kind for particle counts in a cloud chamber.

(iii) Misrecorded discrete distributions

This descriptive is used in Johnson, Kotz and Kemp (1992). Particular values of the frequency of a random variable may be recorded in error.

We now give some details concerning the geometric distribution and several modified distributions (Poisson, logarithmic series, and several truncated binomial). The case of the geometric distribution provides more support for the validity of the formula for a maximum likelihood estimator given in Shenton and Bowman (1977). With regard to the corresponding Maple symbolic programs, accuracy problems should receive most careful scrutiny; losses of accuracy are always possible for numerical computations. One check on this is to increase the length of the digits used in the program.

2 The geometric distribution

Th probability function is

$$Pr(X = x) = p(1-p)^x \qquad (0 < p < 1, x = 0, 1, \cdots, \infty)$$

or $Pr(X = x) = pq^x$, $p + q = 1$. For a random sample $X_1, X_2, \cdots, X_N$, the maximum likelihood estimator $\hat{p}$ of p is given by

$$\hat{p} = 1/(1 + \bar{x})$$

where $\bar{x}$ is the mean of the sample. It follows that (Shenton and Bowman, 2001, p114)

$$\hat{p} - p = -p^2 y/(1 + py), \tag{2}$$

where $y = \bar{x} - \mu_1'(x)$ with $\mu_1'(x) \equiv q/p$. We need the moments of the variate X. We have

$$\begin{aligned}
\mu_2(x) &= q/p^2, \\
\mu_3(x) &= (q + q^2)/p^3, \\
\mu_4(x) &= 3\mu_2^2(x) + (q + 4q^2 + q^3)/p^4.
\end{aligned}$$

From the moments of the mean $\bar{x}$ we readily set up, using expression (2), low order moments of $\hat{p}$ are found. For example,

$$\begin{aligned}
\mu_{11}'(\hat{p}) &= pq, \\
\mu_{12}'(\hat{p}) &= (1 - 2p)pq, \\
\mu_{21}(\hat{p}) &= p^2 q, \\
\mu_{22}(\hat{p}) &= 2(2 - 3p)p^2 q, \\
\mu_{32}(\hat{p}) &= (5q - 1)p^3 q, \\
\sqrt{\beta_{11}(\hat{p})} &= \frac{5q - 1}{\sqrt{q}}, \\
\mu_{43}(\hat{p}) &= (54 - 138p + 85p^2)p^4 q,
\end{aligned}$$

from which there is the kurtosis factor

$$K(\hat{p}) = \frac{30 - 78p + 49p^2}{q} \quad \text{or} \quad K(\hat{p}) = \frac{1}{q} - 20 + 49q.$$

Comment. This approach to the low order maximum likelihood moments of $\hat{p}$ is quite independent of the treatment in the Maple symbolic program. The results from the two processes agree exactly. The skewness of the distribution of $\hat{p}$, measured by the dominant asymptotic, will be large if q is small; strangely enough it will be zero when $q = 1/5$.

We note a generalized form for the geometric defined by the probability generating function

$$G(p,t) = p^k(1-qt)^{-k} \quad (0 < p < 1, p+q = 1, k \text{ a known positive integer})$$

when $k = 1$ it degenerates to the usual geometric probability function.

$$\begin{aligned} \text{Mean}: &\quad kq/p, \\ \text{Max.Lik.Estimator}: &\quad \hat{p} = k/(k+\bar{x}), \end{aligned}$$

so that

$$\hat{p} - p = -(p^2y/k)/(1+py/k).$$

Factorial moment generating function: $(1 - q\alpha/p)^{-k}$.
Moments are

$$\begin{aligned} \mu_{21}(\hat{p}) &= qp^2/k, \\ \mu_{32}(\hat{p}) &= p^3q(5q-1)/k^2, \\ \sqrt{\beta_{11}(\hat{p})} &= \frac{5q-1}{\sqrt{kq}} \end{aligned}$$

Comments: The asymptotic skewness will in general be small if k is large and q not small.

This algebraic moment approach agrees exactly with the Maple symbolic program.

3 Zero truncated Poisson distribution

3.1 Moments of Zero truncated Poisson distribution

$$Pr(X = x) = \frac{e^{-\theta}\theta^x/x!}{1-e^{-\theta}}. \qquad (x = 1, 2, \cdots, \infty, \theta > 0)$$

Factorial moments are

$$\mu_{[s]} = \theta^s/(1-e^{-\theta})$$

and the mean

$$\mu_1'(x) = \theta/(1-e^{-\theta}).$$

The central moments

$$\begin{aligned}
\mu_2(x) &= \frac{\theta}{1-e^{-\theta}} - \frac{\theta^2 e^{-\theta}}{(1-e^{-\theta})^2}, \\
\mu_3(x) &= \frac{\theta}{1-e^{\theta}} + \frac{\theta^2}{1-e^{-\theta}}\left(3 - \frac{3}{1-e^{-\theta}}\right) + \frac{\theta^3}{1-e^{-\theta}}\left(1 - \frac{3}{1-e^{\theta}} + \frac{2}{(1-e^{-\theta})^2}\right), \\
\mu_4(x) &= \frac{\theta}{1-e^{\theta}} + \frac{\theta^2}{1-e^{-\theta}}\left(7 - \frac{6}{1-e^{-\theta}}\right) + \frac{\theta^3}{1-e^{-\theta}}\left(1 - \frac{4}{1-e^{\theta}} + \frac{6}{(1-e^{-\theta})^2}\right) \\
&\quad + \frac{\theta^4}{1-e^{-\theta}}\left(1 - \frac{4}{1-e^{\theta}} + \frac{6}{(1-e^{-\theta})^2} - \frac{3}{(1-e^{-\theta})^3}\right).
\end{aligned}$$

3.2 Moments of maximum likelihood estimator

$$\mu_{11}'(\hat{\theta}) = -\frac{\theta e^{-\theta}(e^{2\theta}\theta - \theta - 2e^{2\theta} + 4e^{\theta} - 2)}{2(-e^{\theta}+\theta+1)^2},$$

$$\mu_{21}(\hat{\theta}) = -\frac{\theta e^{-\theta}(e^{\theta}-1)^2}{(-e^{\theta}+\theta+1)},$$

$$\mu_{22}(\hat{\theta}) = \frac{e^{-\theta}\theta A}{2(-e^{\theta}+\theta+1)^4}$$

where

$$\begin{aligned}
A = {} & 20 + 2e^{4\theta}\theta^2 - 40\theta - 4\theta^3 - 4e^{-\theta} + 40e^{2\theta} + 8e^{2\theta}\theta^2 + 10e^{-\theta}\theta^2 \\
& -26\theta^2 + 8e^{-\theta}\theta + 2e^{\theta}\theta^3 + 4e^{4\theta} - 40e^{\theta} - 20e^{3\theta} + 16e^{\theta}\theta^2 + 40\theta e^{3\theta} \\
& +80e^{\theta}\theta - 8e^{4\theta}\theta - 10\theta^2 e^{3\theta} - 80e^{2\theta}\theta + 3\theta^3 e^{-\theta} - 4e^{2\theta}\theta^3 + 3e^{3\theta}\theta^3,
\end{aligned}$$

$$\mu_{32}(\hat{\theta}) = \frac{(e^{\theta}-1)^2 e^{-2\theta}(2e^{2\theta}\theta^2 - 2\theta^2 - 3e^{2\theta\theta} + 6e^{\theta}\theta - 3\theta - e^{3\theta} + 1 - 3e^{\theta} + 3e^{2\theta})\theta}{(-e^{\theta}+\theta+1)^3},$$

$$\sigma(\hat{\theta}) = \sqrt{-\frac{\theta e^{-\theta}(e^{\theta}-1)^2}{-e^{\theta}+\theta+1}},$$

$$\sqrt{\beta_{11}(\hat{\theta})} = \frac{-e^{-\theta}(2e^{2\theta}\theta^2 - 2\theta^2 - 3e^{2\theta}\theta + 6e^{\theta} - 3\theta + 3e^{2\theta} - 3e^{\theta} - e^{3\theta} + 1)}{(-e^{\theta}+\theta+1)^2\sigma},$$

$$\mu_{43}(\hat{\theta}) = -\frac{\theta B}{(-e^{\theta}+\theta+1)^5}$$

where

$$\begin{aligned} B = {} & 196e^{-2\theta}\theta - 56 - 264e^{-2\theta}\theta^2 - 42e^{4\theta}\theta^2 + 980\theta + 18\theta^4 e^{3\theta} - 54\theta^4 e^{2\theta} - 225\theta^3 \\ & +28e^{-\theta} + 18e^{-3\theta}\theta^4 - 56e^{2\theta} - 852e^{2\theta}\theta^2 + 807e^{-\theta}\theta^2 - 1370\theta^2 - 8e^{-2\theta} + 54e^{-\theta}\theta^4 \\ & -255e^{-2\theta}\theta^3 - 54e^{-2\theta}\theta^4 - 588e^{-\theta}\theta - 105e^{\theta}\theta^3 - 8e^{4\theta} + 70e^{\theta} + 28e^{3\theta} + 1395e^{\theta}\theta^2 \\ & -196te^{3t} - 980e^{t}t + 9e^{4t}t^3 + 54t^4e^{t} - 36\theta^4 + 28e^{4\theta}\theta + 289\theta^2 e^{3\theta} + e^{5\theta} + 588e^{2\theta}\theta \\ & +411\theta^3 e^{-\theta} + 183e^{2\theta}\theta^3 + e^{-3\theta} - 75e^{3\theta}\theta^3 + 57e^{-3\theta}\theta^3 + 37e^{-3\theta}\theta^2 - 28e^{-3\theta}\theta, \end{aligned}$$

$$K = \frac{-C}{\theta(-e^{\theta}+\theta+1)^3(e^{\theta}-1)^4}$$

where

$$\begin{aligned} C = {} & -33e^{5\theta}\theta^3 - 8 - 348e^{4\theta}\theta^2 + 112\theta + 15\theta^4 e^{3\theta} - 117\theta^3 + e^{-\theta} \\ & -56e^{2\theta} - 530e^{2\theta}\theta^2 + 13e^{-\theta}\theta^2 - 96\theta^2 + 9e^{-\theta}\theta^4 - 16e^{-\theta}\theta \\ & +177e^{\theta}\theta^3 - 56e^{4\theta} + 28e^{\theta} + 70e^{3\theta} + 303e^{\theta}\theta^2 - 560\theta e^{3\theta} - 336e^{\theta}\theta \\ & +93e^{4\theta}\theta^3 + 15\theta^4 e^{\theta} - 24\theta^4 + 336e^{4\theta}\theta + 555\theta^2 e^{3\theta} \\ & +28e^{5\theta} + 560e^{2\theta}\theta + 27\theta^3 e^{-\theta} - 75e^{2\theta}\theta^3 - 75e^{3\theta}\theta^3 \\ & +e^{7\theta} - 24e^{4\theta}\theta^4 + 3e^{6\theta}\theta^3 + 121e^{5\theta}\theta^2 - 112e^{5\theta}\theta - 8e^{6\theta} \\ & -18e^{6\theta}\theta^2 + 9e^{5\theta}\theta^4 + 16e^{6\theta}\theta. \end{aligned}$$

TABLE I. Zero truncated Poisson distribution

θ	μ_{11}	μ_{21}	μ_{22}	μ_{32}	σ	$\sqrt{\beta_{11}}$	μ_{43}	K
0.1	-0.0312	0.1936	-0.1108	0.3505	0.4399	4.12	0.3953	13.99
0.2	-0.0584	0.3750	-0.1831	0.6157	0.6124	2.68	0.2312	4.57
0.3	-0.0819	0.5456	-0.2255	0.8134	0.7386	2.02	-0.2128	1.77
0.4	-0.1023	0.7063	-0.2451	0.9582	0.8404	1.61	-0.7535	0.57
0.5	-0.1197	0.8582	-0.2472	1.0619	0.9264	1.34	-1.2756	-0.0040
0.6	-0.1345	1.0020	-0.2362	1.1341	1.0010	1.13	-1.7105	-0.2893
0.7	-0.1470	1.1386	-0.2157	1.1826	1.0670	0.97	-2.0217	-0.4228
0.8	-0.1573	1.2687	-0.1884	1.2136	1.1264	0.85	-2.1937	-0.4719
0.9	-0.1658	1.3931	-0.1565	1.2323	1.1803	0.75	-2.2251	-0.4727
1.0	-0.1726	1.5122	-0.1217	1.2428	1.2297	0.67	-2.1227	-0.4456
2.0	-0.1795	2.5173	0.2069	1.3606	1.5866	0.34	3.3100	0.0292
3.0	-0.1382	3.3823	0.3617	1.9438	1.8391	0.31	10.1478	0.2455
4.0	-0.0919	4.2434	0.3775	2.9411	2.0600	0.34	14.8704	0.2921
5.0	-0.0554	5.1407	0.3184	4.1468	2.2673	0.36	16.9368	0.2693
6.0	-0.0309	6.0757	0.2333	5.4019	2.4649	0.36	16.8121	0.2250
7.0	-0.0162	7.0386	0.1536	6.6211	2.6530	0.35	15.5014	0.1820
8.0	-0.0081	8.0188	0.0929	7.7779	2.8318	0.34	13.9893	0.1480
9.0	-0.0039	9.0089	0.0526	8.8774	3.0015	0.33	12.8706	0.1236
10.0	-0.0018	10.0041	0.0282	9.9354	3.1629	0.31	12.3362	0.1063

Comments: The skewness of the basic random variable and that of $\hat{\theta}$ both become large when θ is small. If the safe sample size (N^*) is set from $\sqrt{\beta_{11}(\hat{\theta})} = 1/4$, then $N^* = 276$ when $\theta = 0.1$. In Figure 1, graphs of the low order moments of $\hat{\theta}$ are displayed.

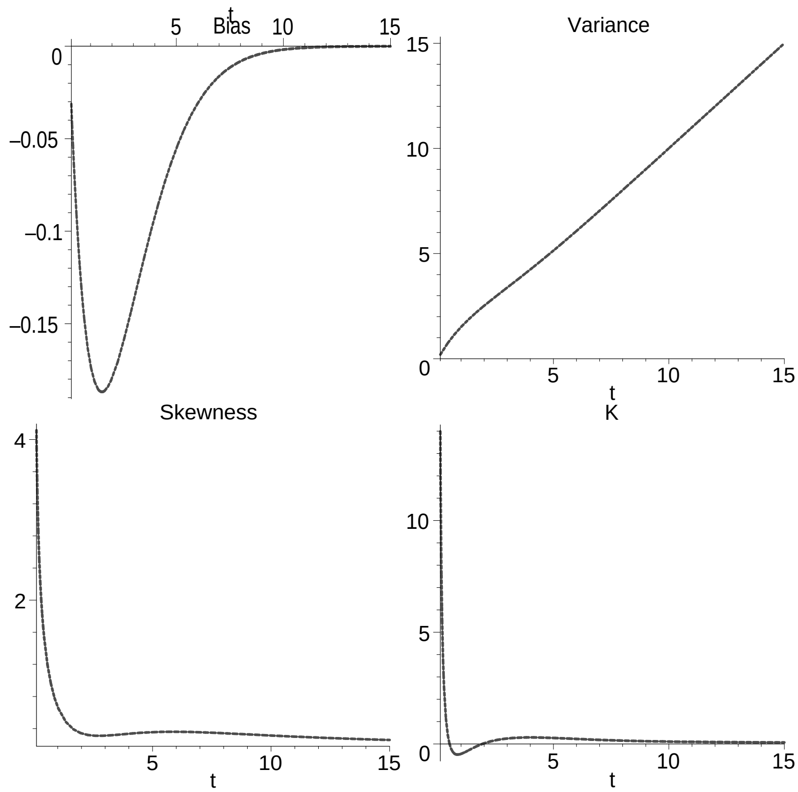

Figure 1: Zero truncated Poisson distribution

4 Logarithmic series distribution

Probability function is

$$Pr(X=x)=\frac{-t^x}{x\ln(1-t)}. \qquad (0<t<1, x=1,2,\cdots,\infty)$$

Asymptotic moments of maximum likelihood estimator $\hat{t}$ are

$$\mu'_{11}(\hat{t})=-\frac{2\,(1-t)^2\ln{(1-t)}^2\,t^2\,(F\,(2,2;3;t)\,A_1+A_2)}{(F\,(2,2;3;t)\,A_3+A_4)^2}$$

where $F(2,2;3;t)=\frac{\Gamma(3)}{\Gamma(2)\Gamma(2)}\sum_{n=0}^{\infty}\frac{\Gamma(2+n)\Gamma(2+n)}{\Gamma(3+n)}\frac{z^n}{n!}$, a hypergeometric function and

$$\begin{aligned}A_1 &= -t^5(1-t)(2-t)-\ln{(1-t)}\,t^4(1-t)(2-t)\\&\quad+(\ln{(1-t)})^2\,t^3(1-t)^2(2-t)(1-2t)+(\ln{(1-t)})^3\,t^2(1-t)^3(2-t^2),\\A_2 &= 2t^4(2-t)+2\ln{(1-t)}\,t^3(4-5t+2t^2)+\ln{(1-t)}^2\,t^3(1-t)(6-5t)\end{aligned}$$

$$+ 2\ln(1-t)^3\, t(1-t)(-4+9t-7t^2+t^3) + 2\ln(1-t)^4\,(1-t)^3(2-t^2),$$
$$A_3 = -t^4(1-t)(2-t) - 2\ln(1-t)\, t^3(1-t)^2(2-t) - \ln(1-t)^2\, t^2(1-t)^2(2-t),$$
$$A_4 = 2t^3(2-t) + 4\ln(1-t)\, t^2(1-t)(3-t) + 4\ln(1-t)^2\, t(1-t)(3-3t+t^2)$$
$$+ 2\ln(1-t)^3\,(1-t)^2(2-t),$$

$$\mu_{21}(\hat{t}) = \frac{2t^3\,(\ln(1-t))^3\,(-1+t)^3}{F(2,2;3,t)B_1 + B_2}$$

where

$$B_1 = \Big\{-(\ln(1-t))^2\,(t^2)(1-t)^2(2-t)$$
$$-\; 2\ln(1-t)\, t^3(1-t)^2(2-t) - t^4(1-t)(2-t)\Big\},$$
$$B_2 = 2\,(\ln(1-t))^3\,(1-t)^2(2-t) + 4\,(\ln(1-t))^2\, t(1-t)(3-3t+t^2)$$
$$+ 4\ln(1-t)\, t^2(1-t)(3-t) + 2t^3(2-t).$$

Maple will compute many mathematical functions including hyper geometric and Zeta functions.

Low order moments of $\hat{t}$ tabulated in Table II.

TABLE II. Logarithmic series distribution

t	μ_{11}	μ_{21}	μ_{22}	μ_{32}	σ	$\sqrt{\beta_{11}}$	μ_{43}	K
0.05	-0.0806	0.0918	-0.2733	0.1389	0.3030	4.99	-0.0687	9.71
0.1	-0.1559	0.1677	-0.4371	0.1768	0.4096	2.57	-0.6408	-7.14
0.2	-0.2890	0.2754	-0.5001	0.0609	0.5248	0.42	-1.6184	-10.44
0.3	-0.3974	0.3300	-0.3131	-0.1616	0.5744	-0.85	-1.3233	-6.46
0.4	-0.4786	0.3391	0.0057	-0.3617	0.5823	-1.83	0.0086	-0.03
0.5	-0.5296	0.3109	0.3462	-0.4654	0.5576	-2.68	1.4301	8.11
0.6	-0.5461	0.2548	0.6099	-0.4484	0.5048	-3.49	2.1034	18.03
0.7	-0.5218	0.1812	0.7165	-0.3313	0.4257	-4.30	1.7750	30.33
0.8	-0.4469	0.1024	0.6152	-0.1700	0.3200	-5.19	0.8665	46.59
0.9	-0.3016	0.0340	0.3135	-0.0398	0.1844	-6.34	0.1471	71.80
0.95	-0.1872	0.0104	0.1268	-0.0077	0.1021	-7.23	0.0182	94.69
0.99	-0.0534	0.0006	0.0108	-0.0001	0.0241	-8.84	0.0001	143.47

Remarks: (i) The sensitive parts of the domain of $\hat{t}$ are $t = 0+\varepsilon$, and $t = 1-\varepsilon^*$, where ε and ε^* are small and positive. For example, first take $t = 0.1$ and set $\sqrt{\beta_{11}(0.1)} = \frac{1}{4}$. Then N^* the sample size is 106; however for $t = 0.9$, at the same setting, $N^* = 643$. For more critical values the reader may refer to the Maple program.

(ii) Note in this case the comparison of the variance terms, $\mu_{21}(\hat{t})$, $\mu_{22}(\hat{t})$.

(iii) Some checks on μ'_{11}, μ_{21}, μ_{22}, μ_{32}, μ_{43} are given in Table III.

The agreement between the Maple results and a completely independent algebraic approach is most gratifying. See Bowman and Shenton (1970) and Shenton and Bowman (1977, §5.2).

TABLE III Comparison of Maple and previous results $t = 0.1$

	μ'_{11}	μ_{21}	μ_{22}	μ_{32}	μ_{43}
Maple	-0.1559	0.1677	-0.4371	0.1768	-0.6408
Table	-0.15586	0.16774	-0.4371	0.1768	-0.641

We must point out that we are surprised by these coincidences, since the table values (5.1c) are based on a sample size of $N = 10$, but still can be used for other sample sizes.

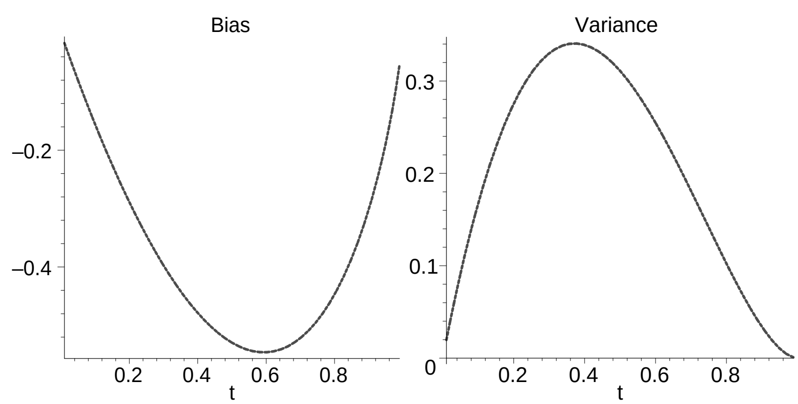

Figure 2: Logarithmic series distribution

5 Doubly truncated binomial distribution

5.1 Probability function

The binomial distribution with $n = 4$

$$Pr(X = x) = \binom{n}{x} p^x q^{n-x}$$

and doubly truncated with 1, the distribution is

$$\begin{aligned} Pr(X = x) &= 4pq^3/L(p) \quad x = 1, \\ &= 6p^2q^2/L(p) \quad x = 2, \\ &= 4p^3q/L(p) \quad x = 3, \end{aligned}$$

where $0 < p < 1$, $p + q = 1$, and $L(p) = 1 - p^4 - q^4$.

$$E(\bar{x}) = np + \frac{pq(q^{n-1} - p^{n-1})}{L(p)}.$$

Asymptotic moments of maximum likelihood estimator are

$$L = -\frac{xt^2 - 2t^2 - xt - 2t - 2 + 2x}{t\,(t-1)\,(t^2 - t + 2)}, \qquad (t = p)$$

$$\mu'_{11} = 1/2\,\frac{(t-1)\,(t^2 - t + 2)\,(2\,t - 1)\,(t^2 - t - 5)\,t}{(2\,t^2 - 2\,t - 3)^2},$$

$$\mu'_{12} = \frac{(t-1)\,t\,(2\,t - 1)\,(t^2 - t + 2)\,A}{4\,(2\,t^2 - 2\,t - 3)^5}$$

where

$$A = 12\,t^{10} - 60\,t^9 + 29\,t^8 + 244\,t^7 - 196\,t^6 - 518\,t^5 - 266\,t^4 + 1800\,t^3 - 2973\,t^2 + 1928\,t - 318,$$

$$\mu_{21} = 1/2\,\frac{t\,(t^2 - t + 2)^2\,(t-1)}{2\,t^2 - 2\,t - 3},$$

$$\mu_{22} = \frac{(t^2 - t + 2)^2 t(t-1)B}{2(2t^2 - 2t - 3)}$$

where

$$B = 16t^8 - 64t^7 + 5t^6 + 209t^5 - 46t^4 - 331t^3 + 454t^2 - 243t + 30,$$

$$\mu_{32} = \frac{(t-1)\,(2\,t-1)\,\left(t^2-t+2\right)^3\,(6\,t^4-12\,t^3-13\,t^2+19\,t-6)\,t}{4\,(2\,t^2-2\,t-3)^3},$$

$$\sigma = \sqrt{\frac{t\,\left(t^2-t+2\right)^2(t-1)}{2(2\,t^2-2\,t-3)}},$$

$$\sqrt{\beta_1} = \frac{(6\,t^4-12\,t^3-13\,t^2+19\,t-6)\,(t^2-t+2)\,(2\,t-1)}{2\,(2\,t^2-2\,t-3)^2\sqrt{\frac{t(t^2-t+2)^2(t-1)}{2(2\,t^2-2\,t-3)}}},$$

$$\mu_{43} = \frac{t\,\left(t^2-t+2\right)^4(t-1)\,C}{8\,(2\,t^2-2\,t-3)^5}$$

where

$$C = -1104\,t + 17261\,t^4 - 16722\,t^3 + 7141\,t^2 - 3544\,t^5 - 9272\,t^6 + 6080\,t^7 + 2080\,t^8 + 480\,t^{10} - 2400\,t^9 + 36,$$

$$K = \frac{D}{2t\,(t-1)\,(2\,t^2-2\,t-3)^3}$$

where

$$D = -744\,t + 7841\,t^4 - 8358\,t^3 + 3865\,t^2 - 124\,t^5 - 6212\,t^6 + 3632\,t^7 + 1252\,t^8 + 288\,t^{10} - 1440\,t^9 + 36.$$

5.2 Newell's example

This example is due to Newell (1965). As an outstanding example of double truncation it is mentioned in Johnson, Kotz and Kemp (1992). Newell considers a medical situation in which a sputum sample from an individual is divided into four sub-samples, each sample being tested for cancer malignancy. A sample is used for statistical consideration if it contains

(i) one positive and three negatives,

(ii) two positives and two negatives,

(iii) three positives and one negative.

Here in which one to three positives relate to malignancy. Newell's practical example is for a hundred persons in which

$$\begin{cases} \text{45 specimens had one positive,} \\ \text{27 specimens had two positives,} \\ \text{15 specimens had three positives,} \\ \text{13 specimens had four positives.} \end{cases}$$

The maximum likelihood estimator is a solution of the equation

$$p^2(2-\bar{x}) + p(2+\bar{x}) + 2(1-\bar{x}) = 0$$

where $\bar{x}$ is the mean of a random sample $X_1, X_2, \cdots, X_N$ taken from the basic truncated binomial. Newell states the solution for his application to be $\hat{p} = 0.347 \pm 0.034$, the range apparently being in terms of the standard error.

Briefly in this case

$$Pr(X = x) = \binom{4}{x} p^x q^{4-x}/L(p). \quad (x = 1, 2, 3, 0 < p < 1)$$

Also from the likelihood expression there is the form

$$\frac{x}{p} - \frac{n-x}{q} - \frac{dL(p)}{dp}/L(p),$$

summed over the sample, leads to the equation

$$\bar{x} = np + \frac{pqdL(p)}{dp}/L(p); \quad (p = \hat{p})$$

moreover we have

$$L(p) = 2pq(2q^2 + 3pq + 2p^2) = 2pq(p^2 - p + 2).$$

Hence

$$\frac{dL(p)}{dp} = 4(q^3 - p^3),$$

and

$$\frac{dL(p)}{dp}/L(p) = \frac{2(q^3 - p^3)}{pq(p^2 - p + 2)}$$

so that after simplification

$$\bar{x} = \frac{2(p^2 + p + 1)}{p^2 - p + 2}. \quad (0 < p < 1)$$

This agrees with Newell's expression

$$p^2(2 - \bar{x}) + p(2 + \bar{x}) + 2(1 - \bar{x}) = 0. \quad (0 < p < 1, p = \hat{p})$$

5.3 There is an obvious generalization

Take

$$Pr(X = x) = \binom{n}{x} p^x q^{n-x}/L_{n,r}(p). \quad (x = r, r+1, \cdots, n-r, 0 < p < 1, p+q = 1) \quad (3)$$

where

$$L_{n,r}(p) = \sum_{x=r}^{n-r} \binom{n}{r} p^x q^{n-x}.$$

Moreover if $n = 2m$, $(m = 1, 2, 3, \cdots)$ then $0 \leq r < m$; if $n = 2m+1$, then $0 \leq r \leq m$ $(m = 1, 2, \cdots)$.

For the maximum likelihood estimator $\hat{p}$, in terms of the mean $\bar{x}$,

$$\frac{\bar{x}}{p} - \frac{n - \bar{x}}{q} - \frac{dL_{n,r}(p)}{dp}/L_{n,r}(p) \quad (p = \hat{p})$$

from which

$$\bar{x} = np + pq\left(\frac{dL_{n,r}(p)}{dp}/L_{n,r}(p)\right). \quad (p = \hat{p}) \quad (4)$$

But

$$\begin{aligned} \frac{dL_{n,r}(p)}{dp} &= \sum_{x=r}^{n-r} \binom{n}{x} \frac{d}{dp}(p^x q^{n-x}) \\ &= r(pq)^{r-1}(q^{n-2r+1} - p^{n-2r+1}). \end{aligned}$$

Hence from (3),

$$\bar{x} = np + \frac{r\binom{n}{r}(q^{n-2r+1} - p^{n-2r+1})}{\sum_{x=r}^{n-r}\binom{n}{x} p^{x-r} q^{n-x-r}}, \quad (5)$$

with the relation between n and r set out under equation (2). This expression may be used to give an approximate solution for $\hat{p}$, plotting $\bar{x}$ against p for $p = 0.1(0.1)0.9$ or more extended range if necessary. Table IV, display the Maple results for a range of p and the cases (4,1), (10,1), (10,2), and (10,3), where (n^*, r) refers to a doubly truncated binomial with index n^*, and double truncation r.

TABLE IV Newell's doubly truncated binomial distribution

p	μ'_{11}	μ'_{12}	$\bar{x}$	μ_{21}	μ_{22}	μ_{32}	$\sqrt{\beta_{11}}$	μ_{43}	K
$n = 4,\ x = 1, 2, 3,$ (4,1)									
0.1	-0.0346	0.0162	1.1623	0.0516	-0.0159	0.0165	1.41	-0.0083	-1.27
0.2	-0.0413	0.0041	1.3478	0.0816	0.0069	0.0115	0.49	-0.0040	-1.10
0.3	-0.0335	-0.0030	1.5531	0.0984	0.0267	0.0053	0.17	0.0126	-0.33
0.347	-0.0269	-0.0039	1.6549	0.1032	0.0329	0.0033	0.10	0.0196	-0.08
0.4	-0.0183	-0.0033	1.7727	0.1068	0.0377	0.0017	0.05	0.0254	0.12
0.5	-0.0000	-0.0000	2.0000	0.1094	0.0410	-0.0000	0.00	0.0299	0.25
0.6	0.0183	0.0033	2.2273	0.1068	0.0377	-0.0017	-0.05	0.0254	0.12
0.7	0.0335	0.0030	2.4469	0.0984	0.0267	-0.0053	-0.17	0.0126	-0.33
0.8	0.0413	-0.0041	2.8377	0.0816	0.0069	-0.0115	-0.49	-0.0040	-1.10
0.9	0.0346	-0.0162	0.0516	-0.0159	-0.0165	0.2272	-1.41	-0.0083	-1.27
$n = 10,\ x = 1, 2, \cdots, 9,$ (10,1)									
0.1	-0.0201	0.0022	1.0533	0.0145	-0.0008	0.0007	0.40	-0.0003	-0.98
0.2	-0.0187	-0.0026	2.0241	0.0204	0.0031	0.0000	0.01	0.0004	-0.05
0.3	-0.0116	-0.0031	3.0087	0.0233	0.0037	-0.0000	-0.01	0.0006	0.18
0.4	-0.0050	-0.0020	4.0024	0.0249	0.0029	0.0000	0.01	0.0005	0.13
0.5	0.0000	0.0000	5.0000	0.0254	0.0023	0.0000	0.00	0.0004	0.07
0.6	0.0050	0.0020	5.9976	0.0249	0.0029	-0.0000	-0.01	0.0005	0.13
0.7	0.0116	0.0031	6.9913	0.0233	0.0037	0.0000	0.01	0.0006	0.18
0.8	0.0187	0.0026	7.9759	0.0204	0.0031	-0.0000	-0.01	0.0004	-0.05
0.9	0.0201	-0.0022	8.9465	0.0145	-0.0008	-0.0007	-0.40	-0.0003	-0.98

p	μ'_{11}	μ'_{12}	$\bar{x}$	μ_{21}	μ_{22}	μ_{32}	$\sqrt{\beta_{11}}$	μ_{43}	K
				$n = 10,\ x = 2, 3, \cdots, 8,\ (10{,}2)$					
0.1	-0.0408	0.0102	2.3212	0.0236	-0.0053	0.0011	0.30	-0.0021	-2.42
0.2	-0.0455	-0.0049	2.7741	0.0303	0.0076	-0.0021	-0.39	0.0014	-0.02
0.3	-0.0343	-0.0088	3.3833	0.0305	0.0115	-0.0024	-0.45	0.0029	0.86
0.4	-0.0178	-0.0054	4.1435	0.0292	0.0109	-0.0014	-0.27	0.0027	0.91
0.5	0.0000	0.0000	5.0000	0.0286	0.0102	0.0000	0.00	0.0024	0.85
0.6	0.0178	0.0054	5.8565	0.0292	0.0109	0.0014	0.27	0.0027	0.91
0.7	0.0343	0.0088	6.6167	0.0305	0.0115	0.0024	0.45	0.0029	0.86
0.8	0.0455	0.0049	7.2259	0.0303	0.0076	0.0021	0.39	0.0014	-0.02
0.9	0.0408	-0.0102	7.6788	0.0236	-0.0053	-0.0011	-0.30	-0.0021	-2.42
				$n = 10,\ x = 3, 4, \cdots, 7,\ (10{,}3)$					
0.1	-0.0622	0.0283	3.2078	0.0368	-0.0172	0.0029	0.41	-0.0093	-4.06
0.2	-0.0750	-0.0032	3.4985	0.0469	0.0157	-0.0058	-0.57	0.0041	-0.14
0.3	-0.0592	-0.0205	3.8970	0.0459	0.0291	-0.0069	-0.70	0.0115	1.65
0.4	-0.0312	-0.0149	4.4096	0.0427	0.0280	-0.0038	-0.43	0.0103	1.73
0.5	0.0000	0.0000	5.0000	0.0413	0.0260	0.0000	0.00	0.0091	1.56
0.6	0.0312	0.0149	5.5904	0.0427	0.0280	0.0038	0.43	0.0103	1.73
0.7	0.0592	0.0205	6.1030	0.0459	0.0291	0.0069	0.70	0.0115	1.65
0.8	0.0750	0.0032	6.5015	0.0469	0.0157	0.0058	0.57	0.0041	-0.14
0.9	0.0622	-0.0283	6.7922	0.0368	-0.0172	-0.0029	-0.41	-0.0093	-4.06
				$n = 10,\ x = 4, 5, 6,\ (10{,}4)$					
0.1	-0.0672	0.0316	4.1379	0.0577	-0.0226	0.0141	1.01	-0.0200	-3.66
0.2	-0.0735	-0.0131	4.3119	0.0836	0.0261	0.0016	0.07	0.0056	-1.08
0.3	-0.0561	-0.0238	4.5192	0.0946	0.0507	-0.0042	-0.14	0.0320	0.36
0.4	-0.0296	-0.0151	4.7525	0.0989	0.0585	-0.0035	-0.11	0.0434	0.88
0.5	0.0000	0.0000	5.0000	0.1000	0.0600	0.0000	0.00	0.0460	1.00
0.6	0.0296	0.0151	5.2475	0.0989	0.0585	0.0035	0.11	0.0434	0.88
0.7	0.0561	0.0238	5.4808	0.0946	0.0507	0.0042	0.14	0.0320	0.36
0.8	0.0735	0.0131	5.6881	0.0836	0.0261	-0.0016	-0.07	0.0056	-1.08
0.9	0.0672	-0.0316	5.8621	0.0577	-0.0226	-0.0141	-1.01	-0.0200	-3.66

5.4 Remarks on the Tabulations

(i) The tabulations are consistent with the symmetry involved in the exchange p for q. For example $\mu_{32}(0.5) = 0$ and $\sqrt{\beta_{11}(0.5)} = 0$. The case of the doubly truncated binomial (10,1) requires more accuracy than the case of (4,1), which is Newell's example. This is also the case for (10,2), (10,3) and (10,4).

(ii) For the tabulated cases, $p = 0.1(0.1)0.9$, large samples are not anticipated for $4 \leq n < 10$; for $n > 10$, caution is advised.

5.5 Limiting case of double truncation of the binomial

For $n = 3$, the binomial doubly truncated by 1 becomes

$$3q^2p + 3qp^2,$$

so that

$$Pr(X = 1) = q, \quad Pr(X = 2) = p.$$

There is a similar characteristic for (5,2), (7,3), and in general $(2n + 1, n)$, $n = 1, 2, \cdots$. It turns out that $\mu_{21}(\hat{p})$ remains unaltered for the cases (5,2), (7,3) and (9,4). Similarly the moments $\mu_{22}(\hat{p})$, $\mu_{32}(\hat{p})$, $\sqrt{\beta_{11}(\hat{p})}$, and K remain unchanged as the n of the binomial changes.

We now look for an independent check. From expression (4) for the mean of the random variables $X_1, X_2, \cdots, X_N$, we have

$$\bar{x} = n + \hat{p}$$

for the doubly truncated case $(2n+1, n)$, $n = 1, 2, \cdots$. Hence except for the mean the central moments of $\hat{p}$, the maximum likelihood estimator, are those of $\bar{x}$. Consider then the moments of X itself. We have for two components

$$Pr(X = n) = \binom{2n+1}{n} p^n q^{n+1} / L(p),$$

$$Pr(X = n+1) = \binom{2n+1}{n+1} p^{n+1} q^n / L(p),$$

so that

$$L(p) = \binom{2n+1}{n} p^n q^n$$

leading to

$$Pr(X = n) = q, \quad Pr(X = n+1) = p,$$

the mean value of X being $n + p$. We have then for moments of X about its mean given by

	x^*	
	$n - (n+p) = -p$	$n + 1 - (n+p) = q$
Probability	q	p

Moments of X now follow

$$\mu_2(x) = q(-p)^2 + p(q)^2 = pq,$$
$$\mu_3(x) = q(-p)^3 + p(q)^3 = pq(1 - 2p),$$
$$\mu_4(x) = qp^4 + pq^4 = pq(1 - 3p + 3p^2).$$

Now for moments of $\hat{p} = \bar{x} - n$,

$$\mu_{21}(\hat{p}) = \mu_2(x) = pq,$$
$$\mu_{32}(\hat{p}) = \mu_3(x) = pq(1 - 2p),$$
$$\mu_{43}(\hat{p}) = \mu_4(x) + 3\mu_2^2(x) = pq(1 - 6p + 6p^2),$$

so that K (an element in the Maple output) is given by

$$K = (6p^2 - 6p + 1)/(pq).$$

We see that the moments are independent of n, $n \geq 1$. These moments agree exactly with those derived from the Maple program in three special cases. This is a very satisfactory outcome, and provides an independent check on the Maple (Bowman and Shenton, 2005) approach.

For all the examples included (geometric, Poisson, etc) the sample mean is a function of the estimators. We include further examples involving the doubly truncated binomial distributions $(2n, n-1)$.

For the sample mean $\bar{x} = \sum X_i/N$,

$$\bar{x} = \frac{n\left\{(n-1)\hat{p}^2 - (n-3)\hat{p} + n - 1\right\}}{(n-1)\hat{p}^2 - (n-1)\hat{p} + n}. \quad (0 < \hat{p} < 1, n = 1, 2, \cdots)$$

Note that the polynomials in the numerator and the denominator do not have real zeros. Similarly, when the probability function reduces to three terms, we have for the doubly truncated binomial $(2n+1, n-1)$,

$$\bar{x} = \frac{2(n-1)\hat{p}^3 + 2(n-1)^2\hat{p}^2 - n(2n-5)\hat{p} + n(n-1)}{(2n-2)\hat{p}^2 - (2n-2)\hat{p} + n}. \quad (n = 1, 2, \cdots, 0 < p < 1)$$

6 Negative binomial distribution

6.1 Basic formula

The pgf is

$$G(t) = (1 + p - pt)^{-k}, \qquad (p > 0, k > 0)$$

and probability function

$$Pr(X = x) = \frac{p^x}{x!(p+1)^{k+x}} \frac{\Gamma(k+x)}{\Gamma(k)}. \qquad (x = 0, 1, \cdots) \tag{6}$$

The fmgf is

$$b(\alpha, p, k) = (1 - p\alpha)^{-k},$$

with moments

$$\mu'_{[1]} = pk,$$
$$\mu'_{[2]} = p^2 k(k+1),$$

with $\mu_2 = kp(p+1)$. The distribution is called the negative binomial distribution since it is derived from the binomial, parameters p, n, by the transformation $-p$ by p, k for $-n$.

In Bowman and Shenton (1965) contains variance and covariance for the maximum likelihood estimators $\hat{p}$, $\hat{k}$ for a set of parameters $0.1 \leq p \leq 50$, $0.1 \leq k \leq 190$. The N^{-1} and N^{-2} coefficients are given (N the sample size). A year later (Bowman and Shenton, 1966) bias to order N^{-2} was reported. These two reports serve as a checking background for the new Maple program. Note that in the first report the efficiency, introduced by Fisher, is studied.

6.2 Maximum likelihood estimation

From (6), the maximum likelihood equations are

$$\hat{p}\hat{k} = \bar{x},$$

$$\frac{n_1}{\hat{k}} + n_2\left(\frac{1}{\hat{k}} + \frac{1}{\hat{k}+1}\right) + \cdots + n_j \sum_{s=1}^{j} \frac{1}{\hat{k}+s-1} + \cdots = n \ln\left(1 + \frac{\bar{x}}{\hat{k}}\right) \tag{7}$$

where the variate value x_j, $j = 0, 1, \cdots$ occurs n_j times, and where $\bar{x}$ is the mean of a random sample $X_1, X_2, \cdots, X_N$ from (6). Clearly the equation (7) has no simple solution for $\hat{k}$, assuming a solution exists; it is also multivariate in the sample set values $(n_1, n_2, \cdots, n_j)$. From this point of view there is a marked contrast with the geometric, zero truncated Poisson and other discrete one parameter distributions considered. For examples:

Geometric	$\hat{p} = 1/(1+\bar{x})$,
Z.t.P.	$\bar{x} = \hat{\theta}/(1 - e^{-\hat{\theta}})$,
L.S.D.	$\bar{x} = -\hat{\theta}/\left\{(1-\hat{\theta})\ln(1+\hat{\theta})\right\}$.

We assume that the interested reader has noted our previous paper (Bowman and Shenton, 2005) on the Maple symbolic program. It contains details of two parameter cases in particular. An earlier study (Bowman and Shenton, 1965) gives complete details for the two parameter case; see pages 18 to 26.

6.3 Low order maximum likelihood moment using Maple

In Table V a set of 25 points in the parameter space are given for the N^{-1} and N^{-2} terms in the bias and the variance. There is also μ_{32} and $\sqrt{\beta_{11}}$ for the skewness and parameter K for the kurtosis. Computing time considerations linked the parameter range to p and k in (0.1, 10).

We are able to say that the tabulated values for bias (N^{-1} and N^{-2}), and variance (N^{-1} and N^{-2}) agree with these computed 40 years ago in the ORNL reports.

TABLE V Negative binomial distribution

k	p	$\mu'_{11}(\hat{k})$	$\mu'_{12}(\hat{k})$	$\mu_{21}(\hat{k})$	$\mu_{22}(\hat{k})$	$\mu_{32}(\hat{k})$	$\sqrt{\beta_{11}(\hat{k})}$	$K(\hat{k})$
.1	.1	247.55	1263052	25.07	367702	30500	243	153001
.1	.5	14.69	4251	1.55	1180	100	52	8202
.1	1.0	5.37	546	0.58	146	13	28	2783
.1	5.0	1.04	18	0.12	4	0.4	9	449
.1	10.0	0.67	7	0.08	2	0.2	7	272
.5	.1	344.26	536500	172.61	757907	304645.6	134	47867
.5	.5	21.84	1955	10.99	2678	1093.5	30	2683
.5	1.0	8.52	272	4.28	365	151.0	17	944
.5	5.0	2.14	13	1.02	17	7.1	7	172
.5	10.0	1.58	7	0.72	8	3.5	6	112
1.0	.1	466.87	522839	464.22	1448441	1140451	114	34844
1.0	.5	31.30	2059	30.41	5547	4392	26	2017
1.0	1.0	12.85	309	12.16	818	649	15	729
1.0	5.0	3.82	20	3.22	51	40	7	146
1.0	10.0	3.05	12	2.41	29	22	6	99
5.0	.1	1468.22	1134121	7133.22	15184256	57476665	95	24623
5.0	.5	112.84	5387	504.55	70435	263214	23	1520
5.0	1.0	51.47	948	215.89	12227	45300	14	576
5.0	5.0	19.77	98	71.20	1207	4376	7	130
5.0	10.0	16.92	66	58.47	802	2884	6	93
10.0	.1	2730.11	1988664	26349.57	52939367	397594891	93	23377
10.0	.5	217.28	9887	1905.03	257490	1905475	23	1462
10.0	1.0	101.22	1801	829.40	46248	339415	14	558
10.0	5.0	40.15	197	285.58	4887	35374	7	127
10.0	10.0	34.52	134	237.38	3289	23726	6	92

k	p	$\mu'_{11}(\hat{p})$	$\mu'_{12}(\hat{p})$	$\mu_{21}(\hat{p})$	$\mu_{22}(\hat{p})$	$\mu_{32}(\hat{p})$	$\sqrt{\beta_{11}(\hat{p})}$	$K(\hat{p})$
.1	.1	3.15	190.48	26.17	4694.38	7327.83	54.74	37851
.1	.5	3.83	65.33	46.13	2050.64	6341.17	20.24	3222
.1	1.0	4.59	43.57	78.29	2110.32	10833.86	15.64	1490
.1	5.0	9.37	23.93	605.95	6799.60	166735.75	11.18	477
.1	10.0	14.06	18.59	1907.14	16251.89	857961.89	10.30	356
.5	.1	0.19	-55.17	7.12	279.07	429.67	22.60	10668
.5	.5	0.15	-0.48	12.49	119.54	412.85	9.35	888
.5	1.0	0.07	0.80	21.12	120.72	733.75	7.56	406
.5	5.0	-0.94	0.57	162.09	357.34	11934.50	5.78	127
.5	10.0	-2.64	-0.11	509.89	805.51	62465.06	5.43	95
1.0	.1	-0.26	-53.94	4.75	107.78	154.89	14.95	7297
1.0	.5	-0.44	-1.98	8.35	43.97	164.61	6.82	595
1.0	1.0	-0.69	0.19	14.16	42.30	304.00	5.71	268
1.0	5.0	-3.02	0.61	110.45	96.96	5298.84	4.56	80
1.0	10.0	-6.36	0.59	351.41	171.50	28489.35	4.32	58
5.0	.1	-0.83	-14.16	2.88	22.25	28.98	5.94	4654
5.0	.5	-1.19	0.02	5.20	5.01	45.13	3.81	359
5.0	1.0	-1.66	0.30	9.04	0.70	95.04	3.50	153
5.0	5.0	-5.53	-0.16	77.20	-54.52	2136.46	3.15	38
5.0	10.0	-10.46	-0.59	255.87	-209.72	12604.17	3.08	26
10.0	.1	-0.95	-4.52	2.65	11.24	19.14	4.45	4332
10.0	.5	-1.34	0.13	4.84	0.28	35.39	3.33	330
10.0	1.0	-1.83	0.09	8.49	-3.98	78.23	3.16	138
10.0	5.0	-5.80	-0.24	74.40	-66.04	1910.09	2.98	33
10.0	10.0	-10.78	-0.38	248.38	-232.20	11521.60	2.94	23

TABLE VI Ratios of μ'_{12}/μ'_{11} and μ_{22}/μ_{21}

k	p	$\mu'_{12}(\hat{k})/\mu'_{11}(\hat{k})$	$\mu_{22}(\hat{k})/\mu_{21}(\hat{k})$	$\mu'_{12}(\hat{p})/\mu'_{11}(\hat{p})$	$\mu_{22}(\hat{p})/\mu_{21}(\hat{p})$
.1	.1	5102.18	14666.94	60.47	179.38
.1	.5	289.44	763.68	17.08	44.45
.1	1.0	101.74	249.86	9.49	26.96
.1	5.0	17.39	34.60	4.03	11.60
.1	10.0	10.50	19.57	1.32	8.52
.5	.1	1558.44	4390.92	-287.07	39.17
.5	.5	89.53	243.64	-3.23	9.57
.5	1.0	31.95	85.30	11.17	5.72
.5	5.0	6.09	16.41	-0.60	2.20
.5	10.0	4.14	11.48	0.04	1.58
1.0	.1	1119.89	3120.13	204.12	22.68
1.0	.5	65.77	182.38	4.47	5.26
1.0	1.0	24.05	67.26	0.28	2.99
1.0	5.0	5.33	15.60	-4.83	2.38
1.0	10.0	3.88	12.17	-0.09	0.49
5.0	.1	772.45	2128.67	17.03	7.74
5.0	.5	47.74	139.60	-0.01	0.96
5.0	1.0	18.43	56.63	-0.18	0.08
5.0	5.0	4.94	16.96	-0.03	-0.71
5.0	10.0	3.90	13.72	0.06	-0.82
10.0	.1	728.42	2009.12	4.75	4.25
10.0	.5	45.51	135.16	-0.10	0.06
10.0	1.0	17.79	55.76	-0.05	-0.47
10.0	5.0	4.92	17.11	0.04	-0.89
10.0	10.0	3.90	13.86	0.03	-0.93

6.4 Sample size questions

A brief tabulation of low order moments of the distributions of $\hat{p}$ and $\hat{k}$ are given in Table V, and the ratios of N^{-2} to N^{-1} terms of μ'_1 and variance are given in Table VI. Contours for the bias and variance are given in Shenton and Bowman (1977, pp82-87). The skewness of both $\hat{k}$ and $\hat{p}$ is large for $0 < p < 1$, $0 < k < 1$; for $p > 1$, or $k > 1$ there is a marked decrease.

For an estimator $\hat{t}$ of t, we consider for example,

$$E(\hat{t}) \sim t + \frac{\mu'_{11}}{N} + \frac{\mu'_{12}}{N^2} + \cdots \quad (N \to \infty)$$

$$Var(\hat{t}) \sim \frac{\mu_{21}}{N} + \frac{\mu_{22}}{N^2} + \cdots \quad (N \to \infty)$$

and for "convergence" set up the ratios

$$\text{Bias}: \quad N_b(\hat{t}) = 10|\mu'_{12}/\mu'_{11}| \quad (\mu'_{11} \neq 0)$$

$$\text{Variance}: \quad N_v(\hat{t}) = 10|\mu_{22}/\mu_{21}|.$$

the multiplying factor 10 being just one possibility. But for the skewness we only have the $1/\sqrt{N}$ term, so that a comparison is not possible; however, experience of statistical applications suggests taking a value such as a half for skewness, so this leads to a third measure

$$\text{skewness}: \quad N_s(\hat{t}) = 4(\sqrt{\beta_{11}(\hat{t})})^2.$$

Thus we have three potential sample size possibilities,

$$N_b(bias), \quad N_v(variance), \quad N_s(skewness).$$

Perhaps a possibility for "safe" sample size is to use $Max(N_b, N_v, N_s)$. It is not possible to say much in the general case; for each distribution and estimator will display completely different characteristics, for example, N_b (and N_v) might be small, but N_s large. Remember too, that N_s is not a ratio, it is based on the dominant skewness. Moreover although we could use the kurtosis ratio (μ_{43}/μ_{42}), in general one would expect this to be large since β_2 itself has a range 1 to 15 or so in applied statistics. What confronts us in an unusual choice of options, a kind of risk analysis. We have three disparate entries; can the concepts be integrated decision-wise.

7 Conclusion

Some details of maximum likelihood low order moments for one parameter discrete distributions are given using a Maple symbolic program. Question of the safe sample size is mentioned. Independent processes are sometimes discussed.

The paper is a supplement to Bowman and Shenton (2005).

Appendix

A Maple program for one parameter discrete distribution

We introduced a general program for finding low order moments of maximum likelihood estimators for s parameter in Bowman and Shenton (2005). For the one parameter case, the program is much simpler so here, is an example with the geometric distribution.

```
#Find bias, variance, skewness, kurtosis of ml estimators of
# geometric distribution
# To change distribution to other one parameter discrete distribution
# change "pf" and the ranges "lim", "Llim".
#number of parameters=1

lim :=infinity;
Llim :=0;
pf :=t*(1-t)^x; LL :=log(pf);
   D1 :=simplify(diff(LL,t));
   SOLx :=solve(D1=0,x);
   D2 :=simplify(diff(D1,t));
   d2 :=sum(D2*pf,x=Llim..lim);
   D11 :=simplify(D1*D1);
   f11 :=sum(D11*pf,x=Llim..lim);
   D3 :=simplify(diff(D2,t));
   f3 :=sum(D3*pf,x=Llim..lim);
   D21 :=simplify(D2*D1);
   f21 :=sum(D21*pf,x=Llim..lim);
   D111 :=simplify(D11*D1);
   f111 :=sum(D111*pf,x=Llim..lim);
   D4 := simplify(diff(D3,t));
```

```
 f4 :=sum(D4*pf,x=Llim..lim);
 D31 :=simplify(D3*D1);
 f31 :=sum(D31*pf,x=Llim..lim);
 D22 :=simplify(D2*D2);
 f22 :=sum(D22*pf,x=Llim..lim);
 D211 :=simplify(D21*D1);
 f211 :=sum(D211*pf,x=Llim..lim);
 D1111 :=simplify(D111*D1);
 f1111 :=sum(D1111*pf,x=Llim..lim);
 D5 := simplify(diff(D4,t));
 f5 :=sum(D5*pf,x=Llim..lim);
 D41 :=simplify(D4*D1);
 f41 :=sum(D41*pf,x=Llim..lim);
 D32 :=simplify(D3*D2);
 f32 :=sum(D32*pf,x=Llim..lim);
 D311 :=simplify(D31*D1);
 f311 :=sum(D311*pf,x=Llim..lim);
 D221 :=simplify(D22*D1);
 f221 :=sum(D221*pf,x=Llim..lim);
 U11 := factor(simplify((f3/2+f21)/f11^2));
 U12 := factor(simplify(-U11+(4*f311+8*f221+4*f41+12*f32+f5)/8/f11^3
       +(2*f4*f111+18*f3*f211+15*f3*f4+36*f3*f22+36*f3*f31+
       36*f21*f31+36*f21*f22+24*f4*f21)/12/f11^4
       +(4*f3^2*f111+15*f3^3+48*f3*f21^2+60*f3^2*f21)/8/f11^5));
 U21 := factor(simplify(1/f11));
 U22 :=factor(simplify(-U21
       +U21^3*(3*f22+3*f31+2*f211+f4)
       +U21^4*(5*f21^2+11*f3*f21+7*f3^2/2+f3*f111)));
U32 :=factor(simplify(U21^3*(f111+3*f3+6*f21)));
SIG := factor(simplify(sqrt(U21)));
RB1 := factor(simplify(U32/U21/SIG));
U43 := factor(simplify(-9*U21^2
```

```
        +U21^4*(f1111+24*f211+30*f31+30*f22+10*f4)
        +U21^5*(45*f3^2+12*f21*f111+18*f3*f111+90*f21^2+150*f3*f21)));
  K := factor(simplify(U43/U21^2-6*U22/U21));
save LL,D1,SOLx,U11,U12,U21,U22,U32,SIG,RB1,U43,K,geoans;
quit;
```

The saved file "geoans" could be used to compute any parameter value t, see Table I.

References

Bowman, K.O. and Shenton, L.R. (1965). *Asymptotic Covariances for the Maximum Likelihood Estimators of the Parameters of a Negative Binomial Distribution*, K-1643, Union Carbide Corporation Nuclear Division.

Bowman, K.O. and Shenton, L.R¿ (1966). *Biases of Estimator for the Negative Binomial Distribution*, ORNL-4005, Oak Ridge National Laboratory.

Bowman, K.O. and Shenton, L.R. (1970). Properties of the maximum likelihood estimator for the parameter of the logarithmic series distribution, *Random Counts in scientific Work*, 1,ed., G.P. Patil, Pennsylvania State University Press, 127-150.

Bowman, K.O. and Shenton, L.R. (2005). Bias, variance, skewness, and kurtosis of maximum likelihood estimator using Maple, *Far East Journal of Theoretical Statistics*, 17(2), 137-195.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1993). *Univariate Discrete Distributions, Sec. Ed* , John Wiley & Sons, New York.

Newell, D.J. (1965). Unusual frequency distributions, *Biometrics*, 21, 159-168.

Shenton, L.R. and Bowman, K.O. (1963). Higher moments of a maximum-likelihood estimator, *J. Roy. Statist. Soc., Ser. B.*, 25, 305-317.

Shenton, L.R. and Bowman, K.O. (1977). *Maximum Likelihood Estimation in Small Samples*, Charles Griffin and Co., Ltd.

Shenton, L.R. and Bowman, K.O. (2001). The geometric distribution, cumulants, Eulerian numbers, and the logarithmic function, *Far East Journal of Theoretical Statistics*, 5(1), 113-142.

Tippett, L.H.C. (1932). A modified method of counting particles, *Proceedings of the Royal Society of London, Series A*, 137, 434-446.