UNBIASSED MINIMUM VARIANCE ESTIMATION IN A CLASS OF DISCRETE DISTRIBUTIONS

By JOGABRATA ROY and SUJIT KUMAR MITRA
Indian Statistical Institute, Calcutta

The problem of estimation of parameters involved in discrete distributions like the Binomial, Poisson and Negative Binomial and their truncated forms have been considered by various authors. Most of them, e.g., Fisher (1936, 1941), Haldane (1941), Finney (1949, 1955) David and Johnson (1952) discuss computational aspects of estimation by maximum likelihood while a few others e.g., Plackett (1953), Moore (1952, 1954), Rider (1953, 1955) give other simpler but inefficient methods of estimation which at times lead to unbiassed estimates.

In this paper, for a wide class of discrete distributions involving one unknown parameter the uniformly minimum variance unbiassed (UMVU) estimate of the parameter is derived and the UMVU estimate of its variance is also obtained. It is shown that results for distributions truncated at zero can be obtained from those for the complete disbribution by a difference operation. The negative binomial and the Poisson distributions are treated as special cases. Tables of the UMVU estimate are given for the Poisson distribution truncated at zero, for sample size upto ten.

1. UMVU estimates in a class of discrete distributions

Consider the following discrete probability distribution defined by Noak (1950).

$$\text{prob } \{X = x\} = a(x)\theta^x/f(\theta),\ x = 0, 1, 2, \ldots, \qquad \ldots \quad (1.1)$$

where $\theta > 0$ is an unknown parameter, $a(x) > 0$ does not involve θ and

$$f(\theta) = \sum_{x=0}^{\infty} a(x)\theta^x. \qquad \ldots \quad (1.2)$$

Without loss of generality, we shall assume $a(0) = 1$. The Poisson, Negative Binomial and Logarithmic series distributions occur as special cases of the above.

To derive uniformly minimum variance unbiassed (UMVU) estimate for θ on the basis of a random sample $X_i(i = 1, 2, \ldots n)$ of size n from the distribution (1.1) we require the lemmas stated below without proof.

Lemma 1.1: *If* $t_r(x)$ *is defined as*

$$\left.\begin{aligned} t_r(x) &= 0 \quad \text{for } x < r \\ &= \frac{a(x-r)}{a(x)} \quad \text{for } x \geqslant r \end{aligned}\right\} \qquad \ldots \quad (1.3)$$

then $E\{t_r(X)\} = \theta^r$.

Lemma 1.2: $T = \sum_{i=1}^{n} X_i$ *is a complete sufficient statistic for* θ *in the sense of Lehmann and Scheffé* (1950).

Lemma 1.3: *The probability distribution of T is given by*

$$\text{prob } \{T = t\} = C(t, n)\, \theta^t/[f(\theta)]^n, \; t = 0, 1, 2, \ldots, \quad \ldots \quad (1.4)$$

where $C(t, n)$ is the coefficient of θ^t in the expansion of $\{f(\theta)\}^n$ and it may be expressed alternatively as

$$C(t, n) = \Sigma \prod_{i=1}^{n} a(x_i),$$

Σ denoting summation over non-negative integral values of $x_1, x_2, \ldots, x_n$ subject to, $x_1+x_2+\ldots+x_n = t$.

Defining for any positive integer r,

$$\left.\begin{aligned} u_r(t) &= 0 && \text{for } t < r \\ &= \frac{C(t-r, n)}{C(t, n)} && \text{for } t \geqslant r \end{aligned}\right\} \quad \ldots \quad (1.5)$$

and noting that the probability distribution of T is of the same form as (1.1) it follows from Lemma 1.1 that

$$E\{u_r(T)\} = \theta^r. \quad \ldots \quad (1.6)$$

Incidentally we may note that

$$u_r(t) = \prod_{i=0}^{r-1} u_1(t-i). \quad \ldots \quad (1.7)$$

From (1.6) and the completeness of the sufficient statistic T it follows from Rao-Blackwell Theorem (Blackwell, 1947 and Rao, 1945) that $u_r(T)$ is the UMVU estimate of θ^r. The variance of $u_1(T)$, the UMVU estimate of θ is given by

$$V\{u_1(T)\} = E\{u_1(T)\}^2 - \theta^2. \quad \ldots \quad (1.8)$$

Hence we have the following:

Theorem 1.1: *For the distribution (1.1) the UMVU estimate for θ is $u_1(T)$ and the UMVU estimate of the variance of $u_1(T)$ is*

$$v(T) = u_1^2(T) - u_2(T) \quad \ldots \quad (1.9)$$

which because of (1.7) may be put in the form:

$$v(T) = u_1(T)\{u_1(T) - u_1(T-1)\} \quad \ldots \quad (1.10)$$

where $u_r(T)$ for $r = 1, 2$, are defined by (1.5).

2. UMVU ESTIMATE IN THE CASE OF TRUNCATION ON THE LEFT

Consider the distribution (1.1) truncated on the left at $x = s-1$. The probability distribution can then be written as

$$\text{prob } \{X = x\} = a(x)\theta^x/f_s(\theta), \; x = s, s+1, s+2, \ldots, \quad \ldots \quad (2.1)$$

where

$$f_s(\theta) = \sum_{x=s}^{\infty} a(x)\, \theta^x \quad \ldots \quad (2.2)$$

By an argument exactly similar to the one used in section 1, it follows that on the basis of a random sample $X_i (i = 1, 2, ..., n)$ of size n from the distribution (2.1) the UMVU estimate for θ^r is $u_{r,s}(T)$ where

$$\left.\begin{aligned} u_{r,s}(t) &= 0 \qquad \text{for } t < ns+r \\ &= \frac{C_s(t-r,n)}{C_s(t,n)} \text{ for } t \geqslant ns+r \end{aligned}\right\} \quad \dots \quad (2.3)$$

where $T = \sum_{i=1}^{n} X_i$ and

$$C_s(t, n) = \Sigma_s \prod_{i=1}^{n} a(x_i), \quad \dots \quad (2.4)$$

Σ_s denoting summation over integral values of $x_1, x_2, ..., x_n$ subject to $x_1+x_2 ...+x_n = t$ and $x_i \geqslant s$ $(i = 1, 2, ... n)$.

The following relation connecting $C_s(t, n)$ can be easily established,

$$C_s(t, n) = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} \{a(s-1)\}^j C_{s-1}\{t-j(s-1), n-j\}, \quad \dots \quad (2.5)$$

where we write $C_0(t, n) = C(t, n)$.

For the special case of interest of truncation at zero, that is $s = 1$, we have

$$C_1(t, n) = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} C(t, n-j), \quad \dots \quad (2.6)$$

which may be expressed in the alternative form :

$$C_1(t, n) = \Delta^n C(t, x)]_{x=0} \quad \dots \quad (2.7)$$

where Δ is the ordinary difference operator on $C(t, x)$ regarded as a function of the integer x.

In general, introducing the bivariate operator Δ_k defined by

$$\Delta_k f(x, y) = f(x+k, y+1) - a(k) f(x, y),$$

we can write (2.5) in the alternative form :

$$C_s(t, n) = \Delta_{s-1}^{n} \; C_{s-1}\{t-n(s-1), 0\}. \quad \dots \quad (2.8)$$

3. Illustrations

3.1. *Complete Negative Binomial Distribution.* We shall consider the following form of the complete Negative Binomial distribution:

$$\text{prob } \{X = x\} = \frac{(k+x-1)!}{(k-1)!\, x!} \theta^x/(1-\theta)^{-k}, \; x = 0, 1, 2, ..., \quad (3.1)$$

and take k as a known positive integer as in problems of inverse binomial sampling. It is well-known that the total T of a random sample of size n from this distribution has again the Negative Binomial distribution.

$$\text{prob}\ \{T = t\} = \frac{(kn+t-1)!}{(kn-1)!\,t!}\ \theta^t/(1-\theta)^{-kn},\ t = 0, 1, 2, \ldots, \qquad \ldots\ (3.2)$$

so that here

$$C(t, n) = \frac{(kn+t-1)!}{(kn-1)!\,t!}. \qquad \ldots\ (3.3)$$

Hence for the complete distribution we have the following.

Result 3.1: The UMVU estimate of θ on the basis of a random sample of size n from the Negative Binomial distribution (3.1) is

$$T/(kn+T-1), \qquad \ldots\ (3.4)$$

and the UMVU estimate of the variance of this estimate is

$$\frac{(kn-1)\,T}{(kn+T-1)^2(kn+T-2)}, \qquad \ldots\ (3.5)$$

where T is the total of the sample.

3.2. *Poisson distribution truncated on the left at zero.* For the complete Poisson distribution

$$\text{prob}\ \{X = x\} = \frac{1}{x!}\ \theta^x/e^\theta, \quad x = 0, 1, 2, \ldots, \qquad \ldots\ (3.6)$$

it is well known that the distribution of the total T of a random sample of size n is given by

$$\text{prob}\ \{T = t\} = \frac{n^t}{t!}\,\theta^t/e^{n\theta}, \qquad \ldots\ (3.7)$$

which is again of the Poisson form. Hence,

$$C(t, n) = \frac{n^t}{t!} \qquad \ldots\ (3.8)$$

Therefore for a Poisson distribution truncated on the left at zero, with probability law given by

$$\text{prob}\ \{X = x\} = \frac{1}{x!}\ \theta^x/(e^\theta-1),\ x = 1, 2, \ldots, \qquad \ldots\ (3.9)$$

we have from (3.8) and (2.7) $\quad C_1(t, n) = \left.\frac{\Delta^n x^t}{t!}\right]_{x=0}$

which may be written in the usual form:

$$C_1(t, n) = \frac{\Delta^n 0^t}{t!}. \qquad \ldots\ (3.10)$$

Using (2.3) and the results of Theorem 1.1 we thus have the following:

Result 3.2: On the basis of a random sample of size n from a Poisson distribution truncated on the left at zero as given by (3.9), the UMVU estimate for θ is $u_1(T)$ and the UMVU estimate for the variance of $u_1(T)$ is $V(T) = u_1(T)\{u_1(T) - u_1(T-1)\}$ where

$$\left.\begin{aligned} u_1(t) &= 0 \qquad \text{for } t < n+1 \\ &= \frac{t\Delta^n 0^{t-1}}{\Delta^n 0^t} \text{ for } t \geqslant n+1 \end{aligned}\right\}$$

T being the sample total.

Table 2 gives values of $u_1(t)$ correct to three places of decimals for $n = 2, 3, \ldots, 10$. The values of the differences of zero tabulated by Fisher and Yates (1933) may be used for $n > 10$ if T happens to be less than 25.

On the basis of a sample $x_1, x_2, \ldots x_n$ from a Poisson distribution truncated at zero Plackett (1953) suggested an unbiassed estimate

$$\theta = \frac{1}{n} \sum_{i=1}^{n} y_i$$

where

$$\left.\begin{aligned} y_i &= 0 \quad \text{if } x_i = 1 \\ &= x_i \ \text{ if } x_i > 1. \end{aligned}\right\}$$

The variance of this estimate is

$$V(\theta) = \frac{1}{n}\left\{\theta + \frac{\theta^2}{e^\theta - 1}\right\}.$$

The following table (Table 1) shows the efficiency of Plackett's estimate relative to the UMVU estimate for $n = 10$ and several values of θ.

TABLE 1. EFFICIENCY OF PLACKETT'S ESTIMATES RELATIVE TO UMVU ESTIMATE BASED ON A SAMPLE OF SIZE 10

	variance of		efficiency (3)/(2)
	Plackett's estimate	UMVU estimate	
(1)	(2)	(3)	(4)
1.0	0.158198	0.151703	0.9589
2.0	0.262607	0.252182	0.9603
3.0	0.347150	0.338585	0.9753
4.0	0.439853	0.424500	0.9875

This shows that Plackett's estimate is highly efficient. For larger sample sizes there is a slight fall in efficiency.

TABLE 2. UMVU ESTIMATE OF THE PARAMETER IN POISSON DISTRIBUTION TRUNCATED AT ZERO (BASED ON A SAMPLE OF SIZE n WHEN THE TOTAL OF THE OBSERVATIONS IS t)

t \ n	2	3	4	5	6	7	8	9	10
2	0	—	—	—	—	—	—	—	—
3	1.000	0	—	—	—	—	—	—	—
4	1.714	0.667	0	—	—	—	—	—	—
5	2.333	1.200	0.600	0	—	—	—	—	—
6	2.903	1.667	0.923	0.400	0	—	—	—	—
7	3.444	2.093	1.300	0.750	0.333	0	—	—	—
8	3.969	2.493	1.646	1.067	0.632	0.286	0	—	—
9	4.482	2.874	1.970	1.360	0.930	0.545	0.250	0	—
10	4.990	3.242	2.278	1.635	1.159	0.786	0.480	0.222	0
11	5.493	3.601	2.574	1.896	1.399	1.011	0.694	0.419	0.200
12	5.997	3.953	2.860	2.146	1.627	1.224	0.9 6	0.622	0.387
13	6.498	4.299	3.139	2.388	1.846	1.427	1.088	0.805	0.564
14	6.999	4.642	3.412	2.623	2.057	1.622	1.272	0.980	0.731
15	7.500	4.983	3.680	2.852	2.262	1.810	1.448	1.147	0.892
16		5.321	3.945	3.076	2.461	1.993	1.618	1.309	1.046
17		5.658	4.206	3.297	2.656	2.170	1.783	1.464	1.194
18		5.994	4.466	3.514	2.847	2.343	1.944	1.615	1.338
19		6.329	4.723	3.728	3.034	2.513	2.101	1.762	1.477
20		6.664	4.979	3.940	3.219	2.679	2.254	1.906	1.613
21		6.998	5.233	4.150	3.401	2.843	2.404	3.046	1.746
22		7.332	5.487	4.358	3.582	3.004	2.552	2.184	1.875
23		7.666	5.740	4.565	3.760	3.163	2.697	2.319	2.002
24		7.999	5.992	4.771	3.937	3.320	2.840	2.451	2.127
25		8.333	6.244	4.976	4.112	3.476	2.982	2.582	2.249
26			6.495	5.180	4.286	3.630	3.121	2.711	2.370
27			6.746.	5.384	4.460	3.783	3.259	2.838	2.489
28			6.997	5.586	4.632	3.934	3.396	2.964	2.606
29			7.248	5.789	4.803	4.085	3.532	3.088	2.722
30			7.498	5.991	4.974	4.235	3.666	3.121	2.836
31			7.749	6.192	5.145	4.384	3.800	3.334	2.950
32			7.999	6.394	5.314	4.532	3.933	3.334	2.950
33			8.249	6.595	5.484	4.679	4.065	3.573	3.173
34			8.499	6.796	5.652	4.826	4.196	3.694	3.284
35			8.750	6.996	5.821	4.973	4.327	3.813	3.393
36				7.197	5.990	5.119	4.457	3.931	3.502
37				7.398	6.158	5.265	4.586	4.049	3.610
38				7.598	6.326	5.41ó	4.715	4.166	3.717
39				7.798	6.494	5.555	4.844	4.282	3.824
40				7.999	6.661	5.700	4.972	4.398	3.930

* For values of t beyond this use t/n.

TABLE 2 (*Contd.*)

t \ n	5	6	7	8	9	10
41	8.199	6.829	5.845	5.100	4.513	4.030
42	8.399	6.996	5.989	5.228	4.628	4.141
43	8.599	7.163	6.133	5.355	4.743	4.246
44	8.799	7.331	6.277	5.482	4.857	4.351
45	9.000	7.498	6.421	5.609	4.971	4.455
46		7.665	6.565	5.736	5.085	4.557
47		7.832	6.709	5.862	5.199	4.662
48		7.999	6.852	5.989	5.312	4.765
49		8.165	6.996	6.115	5.425	4.868
50		8.332	7.139	6.241	5.538	4.971
51		8.499	7.282	6.367	5.651	5.073
52		8.666	7.426	6.493	5.763	5.175
53		8.833	7.569	6.619	5.867	5.278
54		8.999	7.712	6.744	5.988	5.379
55		9.166	7.855	6.870	6.101	5.481
56			7.998	6.996	6.213	5.583
57			8.141	7.121	6.325	5.684
58			8.284	7.246	6.437	5.786
59			8.427	7.372	6.549	5.887
60			8.570	7.497	6.660	5.988
61			8.713	7.623	6.772	6.089
62			8.856	7.748	6.884	6.190
63			8.999	7.873	6.995	6.291
64			9.142	7.998	7.107	6.392
65			9.285	8.123	7.218	6.492
66				8.249	7.330	6.593
67				8.374	7.441	6.694
68				8.499	7.553	6.794
69				8.624	7.664	6.895
70				8.749	7.776	6.995
71				8.784	7.887	7.096
72				8.999	7.998	7.196
73				9.124	8.109	7.296
74				9.249	8.221	7.397
75				9.375	8.332	7.497
76					8.443	7.570
77					8.554	7.697
78					8.666	7.798
79					8.777	7.898
80					8.888	7.998

TABLE 2 (*Contd.*)

t \ n	9	10
81	8.990	8.098
82	9.110	8.198
83	9.222	8.299
84	9.333	8.399
85	9.444	8.499
86		8.599
87		8.699
88		8.799
89		8.899
90		8.990
91		9.099
92		9.199
93		9.299
94		9.399
95		9.500
96		

References

Blackwell, D. (1947): Conditional expectation and unbiased sequential estimation. *Ann. Math. Stat.*, 18, 105-110.

David, F. N. and Johnson, N. L. (1952): The truncated Poisson. *Biometrics*, 8, 175-185.

Finney, D. J. (1949). The truncated Binomial distribution. *Ann. Eug.*, 14, 319 328.

Finney, D. J. and Varley, G. C. (1955): An example of the truncated Poisson distribution. *Biometrics*, 11, 387—394.

Fisher, R. A. (1936): The effects of method of ascertainment upon estimation of frequencies. *Ann. Eug.* 6, 13-25.

——— (1941): The Negative Binomial Distribution. *Ann. Eug.* 11, 182-187.

Fisher R A. and Yates F. (1953): *Statistical Tables for Biological, Agricultural and Medical Research* Fourth Edition, Oliver & Boyd, Edinburgh, 1953.

Haldane. J. B. S. (1941): The fitting of Binomial distribution *Ann. Eug.* 11, 179-181.

Lehmann, E. L. and Scheffé, H. (1950): Completeness, Similar Regions and Unbiassed estimation. Part I. *Sankhyā*, 10, 305-340.

Moore, P. G. (1952): The estimation of the Poisson parameter from a truncated distribution. *Biometrika*, 39, 247-251.

——— (1954): A note on truncated Poisson distribution. *Biometrics*, 10, 402-406.

Noak, A. (1950): A class of random variables with discrete distributions. *Ann. Math. Stat.*, 21, 127-132.

Plackett, R. L. (1953): The truncated Poisson distribution. *Biometrics*, 9, 485—488.

Rao, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37, 81-91.

Rider, P. R. (1953): Truncated Poisson distributions. *J. Amer. Stat. Ass.*, 48, 826-830.

——— (1955): Truncated Binomial and Negative Binomial distributions. *J. Amer. Stat. Ass.*, 50,

Paper received : January, 1957.