Estimating the Complete Sample Size from an Incomplete Poisson Sample

SAUL BLUMENTHAL, RAM C. DAHIYA, and ALAN J. GROSS*

Maximum likelihood estimators and a modified maximum likelihood estimator are developed for estimating the zero class from a truncated Poisson sample when the available sample size itself is a random variable. All the estimators considered here are asymptotically equivalent in the usual sense; hence their asymptotic properties are investigated in some detail theoretically as well as by making use of Monte Carlo experiments. One modified estimator appears to be best with respect to the chosen criteria. An example is given to illustrate the results obtained.

KEY WORDS: Truncated Poisson; Estimation; Simulation; Estimating sample size.

1. INTRODUCTION

Let X be a random variable which has the Poisson distribution

$$f(x;\lambda) = e^{-\lambda}\lambda^x/x! \ , \quad x = 0, 1, \ldots; \quad \lambda > 0 \ . \tag{1.1}$$

Consider a truncated random sample consisting of n observations from (1.1) where all observations for which $X = 0$ are missing. Let n_x denote the number of sample observations for which $X = x$, and let $n = \sum_{x\geq 1} n_x$.

For the case of n fixed, the truncated sample is assumed to come from a truncated Poisson distribution, and in this case the only concern is the estimation of λ. For a summary of work on this topic, see Johnson and Kotz (1969).

Dahiya and Gross (1973) consider the case for which $N = n + n_o$, the complete sample size, is constant but in which n is assumed to be a random variable. In their article, Dahiya and Gross estimate N and hence n_o and obtain an asymptotic confidence interval for N. Related work for the binomial distribution is that of Feldman and Fox (1968) and Draper and Guttman (1971).

To review the procedure of Dahiya and Gross (1973) for estimating N, let

$$L_1(\lambda, N) = \binom{N}{n} p^n q^{N-n} \ , \quad L_2(\lambda) = \left(\frac{q}{p}\right)^n \prod_{x=1}^{R} \left[\frac{\lambda^x}{x!}\right]^{n_x} ,$$

where $p = 1 - e^{-\lambda}$ and $q = 1 - p$. Their estimators $\hat{\lambda}_c$ and $\hat{N}_c$ (say) are the values of λ and N, respectively, which maximize $L_2(\lambda)$ and $L_1(\lambda, N)$ simultaneously. These values are the solutions of

$$\hat{\lambda}_c/(1 - e^{-\hat{\lambda}_c}) = \bar{X}^* \ , \tag{1.2}$$

and

$$\hat{N}_c = [\sum_{x=1}^{R} x n_x/\hat{\lambda}_c] = [n/\hat{p}] \ , \tag{1.3}$$

where

$$\begin{aligned} \bar{X}^* &= (1/n)S \quad \text{if} \quad n \geq 1 \\ &= \quad 0 \qquad \text{if} \quad n = 0 \ , \end{aligned} \tag{1.4}$$

$S = \sum_{x\geq 1} x n_x$, $[a]$ is the greatest integer $\leq a$, and $\hat{p} = 1 - e^{-\hat{\lambda}_c}$. If $S = n$, or $n = 0$, no solution for $\hat{N}_c$ exists. Note that if $1 - \hat{p} = n/N'$, for an integer N', both N' and $N' - 1$ maximize L_1. As pointed out by Sanathanan (1972) (in a slightly different context) the value $\hat{\lambda}_c$ first maximizes L_2 and is used in L_1 prior to maximizing L_1 with respect to N. Hence the estimator $\hat{N}_c$ is only a conditional maximum likelihood (ML) estimator of N.

In the next two sections of this article we shall treat the problem of obtaining unconditional ML estimators $(\hat{N}_u, \hat{\lambda}_u)$ and a modified ML estimator of N, and investigate their asymptotic properties. We compare them for moderate sample sizes by a Monte Carlo simulation in Section 4. In that section, reasons are given for preferring a particular modified estimator rather than either of the ML estimators.

2. NEW ESTIMATORS OF *N*

The unconditional likelihood L of the sample is

$$L = L_1 L_2 \propto \binom{N}{n} e^{-N\lambda}\lambda^S \ . \tag{2.1}$$

Let $\mathcal{L} = \ln L$. Then, up to an additive constant

$$\mathcal{L} = \ln\binom{N}{n} - N\lambda + S\ln\lambda \ . \tag{2.2}$$

For all N, $\hat{\lambda}_u = \bar{X} = S/N$. Substituting $\hat{\lambda}_u$ for λ in (2.2) we obtain $\mathcal{L}_N^*$, where by definition,

$$\mathcal{L}_N^* \propto \ln\binom{N}{n} - S + S(\ln S - \ln N) \ . \tag{2.3}$$

* Saul Blumenthal is Professor, Department of Statistics, University of Kentucky, Lexington, KY 40506. Ram C. Dahiya is Associate Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01002. Alan J. Gross is Professor, Department of Biometry, The Medical University of South Carolina, Charleston, SC 29401. The work of the first author was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-75-2841. The authors wish to thank an Associate Editor for many useful suggestions.

© Journal of the American Statistical Association
March 1978, Volume 73, Number 361
Theory and Methods Section

Define

$$\Delta_N\mathcal{L}^* = \mathcal{L}_N^* - \mathcal{L}_{N-1}^* = \ln[N/(N-n)] - S\ln[N/(N-1)] \ .$$

Let $D(V) = \ln(1 - n/V)/\ln(1 - 1/V)$, for $V > n$. It is easy to verify that $D(V)$ is a decreasing function of V. Hence the equation $S = D(V)$ has a unique solution, say $\hat{V}_u$. $\hat{N}_u$ is that N for which $\Delta_N\mathcal{L}^* \geq 0$ and $\Delta_{N+1}\mathcal{L}^* < 0$, or equivalently, $\hat{N}_u$ is that integer such that

$$D(\hat{N}_u + 1) < S \leq D(\hat{N}_u) \ . \tag{2.4}$$

Clearly, $D(\hat{N}_u + 1) < D(\hat{V}_u) \leq D(\hat{N}_u)$; thus $\hat{N}_u \leq \hat{V}_u < \hat{N}_u + 1$, i.e., $\hat{N}_u = [\hat{V}_u]$.

For the development in Section 3, it is convenient to define $\hat{V}_c = n/\hat{p}$. Then if $n \geq 1$, it follows that $\hat{V}_c$ is the solution of

$$-V\ln(1 - n/V) = S \ . \tag{2.5}$$

Let us define $G(V)$ as $= V\ln(1 - n/V)$, for $V > n$. Similar to $D(V)$, $G(V)$ is also a decreasing function of V. Hence the equation $S = G(V)$ has a unique solution $\hat{V}_c$, and $\hat{N}_c = [\hat{V}_c]$.

Note that $\hat{N}_c \geq \hat{N}_u$, a fact that is proved in some generality by Sanathanan (1977).

For values of S near n, the estimates $\hat{N}_c$ and $\hat{N}_u$ are large. See Table 1 for some examples. Note that if $S = n$, $\hat{N}_c$ and $\hat{N}_u$ are undefined. Although in practice S and n are not often equal, this can and does occur if all values from the complete Poisson sample are either zero or one. Clearly, for samples of size N, the probability that all observations are equal to zero or one is $\{e^{-\lambda}(1 + \lambda)\}^N$. As an example, if $\lambda = 0.10$ and $N = 25$, this probability is 0.89 while for $N \geq 50$ and $\lambda \geq 0.75$, it is essentially zero (to at least two decimal places).

1. Values of Estimates for $n = 55$

S	$\hat{N}_c$	$\hat{N}_u$	$\hat{N}_m$*
75	114	113	110
60	339	334	287
56	1549	1521	788
55	undefined	undefined	1540

It is thus clear that if λ is less than 0.5 and N is less than 100, estimating N by $\hat{N}_c$ or $\hat{N}_u$ may not be possible. In order to overcome this difficulty we consider a modification of $\hat{N}_u$. A simple way to generate a class of modified estimators which includes $\hat{N}_u$ as a special case is to weight the likelihood function and then maximize the resulting modified likelihood. This is formally equivalent to finding Bayes modal estimators for an appropriate prior distribution. The likelihood function L for λ and N fixed is given by (2.1). To modify L, we multiply it by $b(\lambda)$ (see (2.6)) which is equivalent in Bayesian terms to putting prior density functions on λ and N, where

$$b(\lambda) = r^{\rho+1}\lambda^{\rho}e^{-r\lambda}/\Gamma(\rho + 1) \ , \quad r > 0, \rho > -1 \ , \tag{2.6}$$

$$P(N = j) = 1/N_o \ , \quad j = 1, 2, \ldots, N_o \ , \tag{2.7}$$

and N_o is an upper bound on N. Since the modified estimators derived in the following do not depend on N_o, N_o can be chosen as large as desired but need not be specified. The modified likelihood of λ and N is given by

$$h(\lambda, N) = C(r, \rho, N_o)\binom{N}{n}e^{-(N+r)\lambda}\lambda^{S+\rho} \ , \quad n \leq N \leq N_o \ , \tag{2.8}$$

where $C(r, \rho, N_o)$ depends on neither N nor λ. The value of λ, $\hat{\lambda}_m$ (say), which maximizes h for fixed N is $\hat{\lambda}_m = (S + \rho)/(N + r)$. Substituting $\hat{\lambda}_m$ for λ in (2.8) we obtain $\hat{h}_N$, where

$$\hat{h}_N = C(r, \rho, N_o) \cdot \binom{N}{n}e^{-(S+\rho)}((S + \rho)/(N + r))^{S+\rho} \ . \tag{2.9}$$

The value of N which maximizes (2.9) is $\hat{N}_m$, where

$$D(\hat{N}_m + 1, r) < S + \rho \leq D(\hat{N}_m, r) \ , \tag{2.10}$$

and

$$D(V, r) = \ln(1 - n/V)/\ln(1 - 1/(V + r)) \ .$$

The solution of (2.10) for $\hat{N}_m$ does not diverge for $S = n$ as long as $\rho > 0$. Note that when $r = 0$, the modified estimator has the appealing form $S + \rho = D(V)$, which is the type of modification that is suggested often in practice to prevent an estimator from blowing up. Clearly, $\hat{N}_u$ is $\hat{N}_m$ with $r = \rho = 0$. The weight function (2.6) was chosen for two reasons. First, it leads to a "natural" looking class of modified estimators. Second, as shown in Blumenthal (1976), a weight of this form or nearly this form is needed to minimize the maximum asymptotic bias, as discussed in the next section, where the choice of ρ and r is examined.

We might note in passing, that $\hat{N}_m$ increases with r for fixed ρ, and decreases with ρ for fixed r. If $\hat{N}_{m,i}$ corresponds to (r_i, ρ_i), $i = 1, 2$ with $r_1 < r_2$, $\rho_1 < \rho_2$, then we can verify that $\hat{N}_{m,1} < \hat{N}_{m,2}$ when S is large and $\hat{N}_{m,1}$ is close to n, with the inequality being reversed for S near n, when $\hat{N}_{m,1}$ is of the order n^2. In particular then, $\hat{N}_m < \hat{N}_u$ where both are large, but not for all S, unless $r = 0$.

As with $\hat{V}_c$ and $\hat{V}_u$, define $\hat{V}_m$ to be that value of V, such that $D(\hat{V}_m, r) = S + \rho$. Again we note that $\hat{N}_m = [\hat{V}_m]$.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

In this section we derive the asymptotic properties of our three estimators, $\hat{V}_c$, $\hat{V}_u$, and $\hat{V}_m$. The approach used here is to express $\hat{V}_i$ as $N + a_i\sqrt{N} + b_i + O_p(1/\sqrt{N})$ $(i = c, u, m)$ where $O_p(1/\sqrt{N})$ is a term which is stochastically of order $(1/\sqrt{N})$, and a_i, b_i are random variables to be determined in what follows. To contrast this expansion with the limit distribution result of Dahiya and Gross (1973), in our terms they showed that $\hat{V}_c = N + a_c\sqrt{N} + O_p(1)$. We shall see that $a_c = a_u = a_m = a$

(say) and that a is asymptotically $N(0, \sigma^2)$, where

$$\sigma^2 = q/(p - \lambda q) \ . \quad (3.1)$$

Thus in the standard asymptotic approach, $(\hat{V}_i - N)/\sqrt{N}$ is asymptotically $N(0, \sigma^2)$, and all the $\hat{V}_i$ are asymptotically equivalent. We shall see that the b_i differ, allowing us to make some comparisons among the estimators. We refer to $E(b_i)$ as the asymptotic bias of V_i, and in Section 4, we show its connection with finite sample bias.

Using the same notation as Dahiya and Gross, we define $T_N = (n - Np)/\sqrt{N}$ and $Y_N = \sqrt{N}(\bar{X} - \lambda)$. It then follows that $n = Np + \sqrt{N}T_N$ and $S = N\lambda + \sqrt{N}Y_N$.

3.1 Asymptotic Properties of $\hat{V}_c$

We write $\hat{V}_c = N + a_c\sqrt{N} + b_c + O_p(1/\sqrt{N})$, where a_c and b_c are to be determined. In the new notation we can then rewrite (2.5) as

$$N\lambda + \sqrt{N}Y_N = -(N + a_c\sqrt{N} + b_c) \cdot \ln\{1 - ((Np + \sqrt{N}T_N)/(N + a_c\sqrt{N} + b_c))\} \ . \quad (3.2)$$

With a bit of algebra it then follows that

$$\begin{aligned} \ln\{1 - (Np + \sqrt{N}T_N)/(N + a_c\sqrt{N} + b_c)\} &= \ln q + \ln(1 + (a_c - T_N)/q\sqrt{N}) + (b_c/qN) \\ &\quad - \ln(1 + (a_c/\sqrt{N}) + (b_c/N)) \\ &= -\lambda + (1/q\sqrt{N})(a_c p - T_N) + (1/2q^2N) \\ &\quad \cdot\{2qb_c p - (a_c - T_N)^2 + a_c^2q^2\} + O_p(N^{-\frac{3}{2}}) \ , \quad (3.3) \end{aligned}$$

via the series expansion of $\log(1 + x)$ where $|x| < 1$. Substituting (3.3) into (3.2), we then have

$$\begin{aligned} N\lambda = \sqrt{N}Y_N &= N\lambda + \sqrt{N}\{\dot{a}_c\lambda - (a_c p - T_N)/q)\} \\ &\quad - (1/2q^2)\{2b_c pq - (a_c - T_N)^2 + a_c^2q^2 \\ &\quad + 2qa_c(a_c p - T_N) - 2b_c\lambda q^2\} + O_p(1/\sqrt{N}) \ . \quad (3.4) \end{aligned}$$

Equating coefficients of $\sqrt{N}$, we find that

$$\begin{aligned} a_c &= (T_N/q - Y_N)/(p/q - \lambda) \ , \\ b_c &= (T_N - a_c p)^2/2q(p - \lambda q) \ . \quad (3.5) \end{aligned}$$

Noting that $(T_N - a_c p) = q(pY_N - \lambda T_N)/(p - \lambda q)$, it follows that

$$b_c = q(pY_N - \lambda T_N)^2/2(p - \lambda q)^3 \ . \quad (3.6)$$

Using an argument similar to Dahiya and Gross (1973, pp. 732–733) it can be shown that the asymptotic distribution of a_c is $N(0, \sigma^2)$, and that of b_c is $(\lambda pq/2(p - \lambda q)^2)\chi^2_1$, where χ^2_1 has the chi-squared distribution with one degree of freedom.

3.2 Asymptotic Properties of $\hat{V}_m$ and $\hat{V}_u$

As with $\hat{V}_c$, write

$$\hat{V}_m = N + a_m\sqrt{N} + b_m + O_p(1/\sqrt{N}) \ .$$

Write the equation $S + \rho = D(\hat{V}_m, r)$ as

$$\begin{aligned} N\lambda + (NY_N)^{\frac{1}{2}} + \rho &= \ln\{1 - (Np + \sqrt{N}T_N)/(N + a_m\sqrt{N} + b_m)\}/ \\ &\quad \ln\{1 - 1/(N + a_m\sqrt{N} + b_m + r)\} \ . \quad (3.7) \end{aligned}$$

Note that

$$\begin{aligned} \ln\{1 - 1/(N + a_m\sqrt{N} + b_m + r)\} &= -\{1/(N + a_u\sqrt{N} + b_u + r) \\ &\quad + (\tfrac{1}{2}(N + a_u\sqrt{N} + b_u)^{-2}) + \ldots\} \ . \quad (3.8) \end{aligned}$$

Inverting (3.8) it then follows that

$$\begin{aligned} 1/\ln\{1 - 1/(N + a_m\sqrt{N} + b_m + r)\} &= -\{N + a_m\sqrt{N} + (b_m + r - \tfrac{1}{2}) + O_p(1/\sqrt{N})\} \ . \quad (3.9) \end{aligned}$$

Again if we multiply the two series together, collect like terms, and set the resulting value equal to the left side of (3.7) we obtain

$$\begin{aligned} N\lambda + \sqrt{N}Y_N + \rho &= N\lambda + \sqrt{N}\{a_m\lambda - ((a_m p - T_N)/q) \\ &\quad - (1/2q^2)\{2b_m pq - (a_m - T_N)^2 + a_m^2q^2 \\ &\quad + 2qa_m(a_m p - T_N) - 2q^2\lambda(b_m + r - \tfrac{1}{2})\} \\ &\quad + O_p(1/\sqrt{N}) \ . \end{aligned}$$

The values a_m and b_m are obtained as in Section 3.1, and we have $a_m = a_c$ and

$$b_m = b_c - \lambda q/2(p - \lambda q) + q(\lambda r - \rho)/(p - \lambda q) \ . \quad (3.10)$$

Clearly, b_u is obtained by setting $r = \rho = 0$ in (3.10).

3.3 Asymptotic Comparisons

In order to compare conditional and unconditional ML estimators, we have

$$\hat{V}_c - \hat{V}_u = \lambda q/2(p - \lambda q) + O_p(1/\sqrt{N}) \ . \quad (3.11)$$

This difference approaches zero in probability for large λ, but is approximately $1/\lambda$ for small λ and hence is large for λ near zero. Note that except for the term which converges stochastically to zero, there is a systematic, nonrandom difference between these estimators. The example in Table 1 indicates the possible magnitude of this difference.

We find $E(b_u)$ and show that it is positive. Since $(\hat{N} - N)/N \xrightarrow{P} 0$ for all of the $\hat{N}$'s, they are all consistent in the usual sense. A reasonable definition of second-order consistency would be that $(\hat{N} - N) \xrightarrow{P} 0$, or a weaker requirement might be that $(\hat{N} - N) \xrightarrow{P} W$, where $EW = 0$. In light of (3.11) and $E(b_u) > 0$, we see that $\hat{N}_c$ and $\hat{N}_u$ do not satisfy these consistency criteria. We now obtain the asymptotic biases of $\hat{V}_u$ and $\hat{V}_c$.

From (3.5) and (3.10), we have

$$b_u = \{(T_N - ap)^2 - \lambda q^2\}/\{2q(p - \lambda q)\} \ . \quad (3.12)$$

From Dahiya and Gross (1973), it follows that

$$E(T_N - ap)^2 = p\lambda q^2/(p - \lambda q) \ , \quad (3.13)$$

and hence after simplification,

$$E(b_u) = \lambda^2/2(e^\lambda - 1 - \lambda)^2 > 0 \ . \quad (3.14)$$

It is obvious from (3.14) that for small λ,

$$E(b_u) = 2/\lambda^2 \ , \quad (3.15)$$

which is large for λ near zero.

From (3.10) and (3.15) it is obvious that the asymptotic biases of both $\hat{V}_c$ and $\hat{V}_u$ are large for values of λ near zero.

Now we obtain the asymptotic bias of $\hat{V}_m$, namely $E(b_m)$. From (3.10), (3.12), and (3.13), we find

$$E(b_m) = \lambda q\{\lambda q + 2(p - \lambda q)(r - \rho/\lambda)\}/2(p - \lambda q)^2 . \quad (3.16)$$

By series expansions, it is not difficult to see that

$$\lambda q/(p - \lambda q) = (2/\lambda)[1 - \lambda/3 + \lambda^2/36 + O(\lambda^3)] , \quad (3.17)$$

and

$$(\lambda q)^2/2(p - \lambda q)^2 = (2/\lambda^2)[1 - (2/3)\lambda + \lambda^2/6 + O(\lambda^3)] . \quad (3.18)$$

Thus for small λ, we have approximately,

$$E(b_m) = (2/\lambda^2)(1 - \rho) - (2/\lambda)(2/3 - r - \rho/3) + (1/3) - (2/3)r - (1/18)\rho . \quad (3.19)$$

It is clear that $\hat{N}_m$ is not asymptotically consistent in the second-order sense and can have large asymptotic biases for small λ with arbitrary choice of ρ and r. However, (3.19) shows that the maximum asymptotic bias occurs as $\lambda \to 0$, and that it is possible to choose ρ and r to minimize this maximum asymptotic bias. Inspection of (3.19) yields $\rho = 1$ and $r = 1/3$ as the unique choice of parameters in the weight function (2.6) which stabilizes the bias for small λ (i.e., prevents $E(b_m)$ from diverging as $\lambda \to 0$). For $\rho = 1$, $r = 1/3$, $E(b_m) \to (1/18)$ as $\lambda \to 0$. Thus $\hat{N}_m$ is very close to being second-order consistent for this choice of ρ, r. Let $\hat{N}_m{}^*$ denote $\hat{N}_m$ with $\rho = 1$, $r = 1/3$.

Asymptotic biases of $\hat{N}_u$, $\hat{N}_c$, and $\hat{N}_m{}^*$ are given in Table 2 where 0.5 is subtracted to account for the difference between $\hat{V}$ and $\hat{N}$. $E(b_u)$ and $E(b_m)$ are given by (3.14) and (3.16), respectively, while $E(b_c)$ can be obtained from (3.10) and (3.14). Notice that $\hat{N}_m{}^*$ has an almost constant asymptotic bias over the entire range of $\lambda < 5$ and that the difference reflected in (3.11) is apparent for λ in the range (0.10, 0.40).

2. Asymptotic Bias and Variance of the Estimators

λ	σ^2	Bias $\hat{N}_u$	Bias $\hat{N}_c$	Bias $\hat{N}_m{}^*$
0.10	193.389	186.50	196.17	−0.446
0.20	46.723	43.16	47.83	−0.447
0.30	20.057	17.60	20.61	−0.449
0.40	10.890	8.99	11.17	−0.450
0.50	6.724	5.15	6.83	−0.452
0.60	4.502	3.15	4.50	−0.453
0.75	2.725	1.59	2.61	−0.455
1.00	1.392	0.47	1.17	−0.459
1.25	0.806	0.008	0.51	−0.462
1.50	0.505	−0.21	0.16	−0.466
1.75	0.333	−0.33	−0.039	−0.469
2.00	0.228	−0.40	−0.168	−0.472
3.00	0.062	−0.48	−0.390	−0.483
5.00	0.007	−0.50	−0.480	−0.490

4. FINITE SAMPLE PROPERTIES AND COMPARISON OF ESTIMATORS BY A MONTE CARLO SIMULATION

As indicated in Section 2, $\hat{N}_u$ and $\hat{N}_c$ both have postive probability of diverging. Thus neither of these estimators has finite moments of any order for finite sample sizes. In practice, in the cases which lead to divergent $\hat{N}_c$ and $\hat{N}_u$, either $\hat{N}$ would be defined arbitrarily or else no estimate would be made. Asymptotic properties are unaffected by arbitrary definition on a set whose probability converges to zero. Noting that when $S = n + k$ $(k = 1, 2, \ldots)$, $\hat{V}_i$ $(i = c, u, m)$ is approximately $(n^2/2k)$, for large n, we shall define $V_c{}^*(V_u{}^*)$ to be n^2 if $S = n$, and to agree with $\hat{V}_c(\hat{V}_u)$ otherwise. In Theorem 1, we show that $E(V_i{}^*)$ is $E(b_i) + o(1)$ as N increases. Hence the asymptotic bias is a legitimate large-sample approximation to the actual bias of $V_i{}^*$ and of $\hat{N}_i$ as encountered in practice.

Theorem 1: For the Poisson distribution,

$$\lim_{N \to \infty} E(V_i{}^* - N - a\sqrt{N}) = Eb_i , \quad i = c, u, m; \; V_m{}^* = \hat{V}_m .$$

The proof is given in the Appendix.

Details of the Monte Carlo simulation are as follows:

(i) A random sample from a Poisson distribution was generated for each of three values of N, $N = 25, 50, 100$. The details of obtaining a random Poisson sample can be found in Tocher (1967, p. 36).

(ii) For each sample with a given N and λ, the estimates $\hat{V}_u$, $\hat{V}_c$, and $\hat{V}_m{}^*$, were computed and corresponding to these estimates $\hat{N}_u$, $\hat{N}_c$, and $\hat{N}_m{}^*$ were obtained.

(iii) This process was repeated 1,000 times for each pair of values N and λ. The simulated average bias $\bar{B} = \sum_1^{1,000} (\hat{N}_i - N)/1{,}000$, and its mean square error (MSE) divided by N, i.e., MSE $= \sum_1^{1,000} (\hat{N}_i - N)^2/1{,}000N$, were then computed. This simulated average bias, is the estimated bias for each given estimator $\hat{N}$. It is computed so that it can be compared with the asymptotic bias in Table 2. The simulated MSE computed here estimates the MSE of $\sqrt{N}(\hat{N}/N - 1)$ rather than $\hat{N}$, for each estimator $\hat{N}$. This is because as $\hat{N}$ becomes large $\sqrt{N}(\hat{N}/N - 1)$ approaches a normal variable with mean zero and variance σ^2. Thus the computed MSE should be compared with σ^2 in Table 2.

(iv) Complete sets of estimates were obtained for all combinations of values $N = 25, 50, 100$; $\lambda = 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 3.00$, and 5.00. However, for values of $\lambda < 0.75$, occasionally all the sample values obtained were zero or one giving rise to $S = n$ in some samples in which case $\hat{V}_c$ and $\hat{V}_u$ are infinite, and hence the bias and MSE of these estimators could not be estimated. The only meaningful estimator is $\hat{V}_m{}^*$ for these small values of λ.

The estimated bias and MSE of the different estimators appear in Tables 3 and 4. It is obvious that $\hat{N}_m{}^*$ under-estimates N. Furthermore, the bias of $\hat{N}_u$ and $\hat{N}_c$ is positive for small values of λ and then becomes negative for larger values of λ. This, however, is not in contradiction to the fact mentioned earlier that $Eb_c > Eb_u > 0$, since $\hat{N}_c$ and $\hat{N}_u$ are smaller than $\hat{V}_c$ and $\hat{V}_u$, respectively, by 0.5 on the average. For larger values of λ, all estimators have bias and MSE of the same magnitude. It is obvious that the modified estimator $\hat{N}_m{}^*$ is most attractive from its overall performance of simulated bias and MSE for finite sample sizes. Furthermore, the modified estimator is the only feasible estimator for values of $\lambda < 0.75$, for the reasons mentioned earlier. Estimated bias and MSE for this estimator in case of small λ are given in Table 4. It is obvious that both bias and MSE are very high for small values of λ which is not surprising since $N - n$, the missing part of the sample, has large expectation for small values of λ.

Comparing the entries in Tables 3 and 4 with Table 2, we conclude that the $O_p(N^{-\frac{1}{2}})$ error term in the bias and the $O_p(N^{-1})$ error term in the MSE go to zero fairly rapidly when λ is large, but for small λ even when N is 100, the agreement between the simulated and limiting values is quite poor. The simulations indicate that for finite N and small λ, the modified estimator has smaller MSE than predicted by σ^2.

To summarize the implications of the simulations, the tabulated values show $\hat{N}_c$ and $\hat{N}_u$ at their best, since infinite values were omitted, yet even in this situation these estimators look no better relative to $N_m{}^*$ than they do in the asymptotic analysis. Since $N_m{}^*$ never blows up, shows no indication of systematically underestimating, and is as easy to compute as its competitors, it is the apparent best estimator among those considered. Another factor to consider is the relative instability of the ML estimators. Since D_N and G_N have asymptotes and are relatively flat for large N, a small change in S when S is near n causes a large change in $\hat{N}$, even when $\hat{N}$ is finite. The modified estimator is not subject to this problem since $D(N, r)$ is unbounded for $r > 0$.

3. Estimated Biases and Mean Squared Errors, $\lambda \geq 0.75$

λ	Bias $\hat{N}_u$	Bias $\hat{N}_c$	Bias $\hat{N}_m{}^*$	MSE/N $\hat{N}_u$	MSE/N $\hat{N}_c$	MSE/N $\hat{N}_m{}^*$
			a. N = 25			
0.75	2.438	3.763	−1.114	7.255	8.625	3.006
1.00	0.964	1.804	−0.460	2.555	2.922	1.640
1.25	0.026	0.567	−0.677	1.255	1.387	0.926
1.50	0.104	0.564	−0.251	0.617	0.670	0.554
1.75	−0.244	0.124	−0.428	0.365	0.387	0.348
2.00	−0.447	−0.181	−0.532	0.253	0.255	0.244
3.00	−0.464	−0.357	−0.468	0.084	0.081	0.084
5.00	−0.152	−0.152	−0.152	0.007	0.007	0.007
			b. N = 50			
0.75	1.148	2.262	−1.430	3.640	3.956	2.626
1.00	0.964	1.712	−0.193	1.904	2.025	1.606
1.25	−0.004	0.516	−0.566	0.934	0.971	0.874
1.50	−0.319	0.092	−0.581	0.553	0.562	0.532
1.75	−0.425	−0.086	−0.558	0.362	0.369	0.336
2.00	−0.340	−0.117	−0.432	0.244	0.247	0.237
3.00	−0.477	−0.392	−0.481	0.067	0.067	0.066
5.00	−0.336	−0.336	−0.336	0.009	0.009	0.009
			c. N = 100			
0.75	1.968	3.045	−0.363	3.176	3.316	2.740
1.00	−0.147	0.563	−1.124	1.554	1.587	1.477
1.25	−0.118	−0.436	−0.585	0.874	0.890	0.852
1.50	−0.416	−0.036	−0.644	0.524	0.533	0.519
1.75	0.017	0.324	−0.127	0.331	0.335	0.329
2.00	−0.558	−0.338	−0.625	0.241	0.241	0.240
3.00	−0.435	−0.352	−0.435	0.065	0.064	0.065
5.00	−0.598	−0.390	−0.598	0.010	0.010	0.010

4. Estimated Bias and Mean Squared Error: Bayes Modal Estimator, $\lambda < 0.75$

λ	Bias, N = 50	Bias, N = 100	MSE/N, N = 50	MSE/N, N = 100
0.10	−35.733	−57.050	27.916	39.399
0.20	−17.377	−12.999	13.768	28.097
0.30	−6.013	−1.720	12.550	27.135
0.40	−1.644	−1.106	14.542	13.900
0.50	−1.171	−1.535	9.368	7.787
0.60	−0.842	−1.040	5.410	5.191

Note that all of the $\hat{N}$'s can be written as

$$\hat{N} - N = a\sqrt{N} + (b_c - \alpha) + (\beta/\sqrt{N}) + O_p(N^{-1}) \quad (4.1)$$

with a and b_c given by (3.5) and α and β are chosen approriately. Since $Ea^2 = \sigma^2$ and it can be shown that $E(ab_c) = O(N^{-\frac{1}{2}})$, it is easily verified that the asymptotic MSE of $\hat{N}$ is $N\sigma^2 + O(1)$, where the first order correction $O(1)$ depends on $E(a(b_c - \alpha))$, $E(b_c - \alpha)^2$, and $E(\alpha\beta)$. If the β term in (4.1) were obtained so that the value of the $O(1)$ term in the MSE oculd be computed, then the estimates could be compared on the basis of their deficiency, as defined by Hodges and Lehmann (1970). Lacking this β term, we must be satisfied with comparisons based on asymptotic bias and on simulated MSE's.

5. AN EXAMPLE

Finally we consider the example given by Dahiya and Gross (1973) referring to an epidemic of cholera in a village in India:

x	1	2	3	4	*Total*
n_x	32	16	6	1	55

where x is the number of cholera cases in a household and n_x denotes the number of houses with x cases of cholera. In addition to the 55 households having at least one case, 168 other households had no cases. The problem is to estimate n_o, the total number of these 168 households that were infected. Using the conditional ML estimate, Gross and Dahiya showed that $\hat{N}_c = 89$ and hence $\hat{n}_{oc} = \hat{N}_c - n = 34$. We find $\hat{N}_u = \hat{N}_m{}^* = 87$. Thus the other estimates of n_o are both 32. More dramatic differences for the same n but smaller S values are exhibited in Table 1.

APPENDIX

Proof of Theorem 1: For simplicity, we outline the proof for V_c^* only. $(V_c^* - N - a_c\sqrt{N}) = \Delta_c^*$ can be rewritten as $\{V_c^* - ((n - qS)/(p - \lambda q))\}$. On the set, $S = \mu n + \xi\sqrt{n}$, where μ is (λ/p), the expectation of the truncated Poisson distribution, a Taylor expansion gives V_c^* as $(n/p) - (\xi q\sqrt{n}/(p - \lambda q)) + (qp^3\xi^2/2(p - \lambda q)^3) + O(\xi^3/\sqrt{n})$. Thus $\Delta_c^* = qp^3\xi^2/2(p - \lambda q)^3 + O(\xi^3/\sqrt{n})$ on this set. Let $\xi = (S - \mu)/\sqrt{n}$, and split $E(\Delta_c^*|n)$ into two integrals, the first over the region $|\xi| < A(\log n)^{\frac{1}{2}}$ for an arbitrary A, and the second over the complement of this region. On the complement, it is easily checked that $\Delta_c^* < c_1 n^2$ for some fixed c_1, and by Corollary 2 of Nagaev (1965),

$$P(|\xi| > A(\log n)^{\frac{1}{2}}) < (\alpha_k n/(A^2 n \log n)^{k/2})$$

provided that $E|X|^k < \infty$. Since all moments of the truncated Poisson are finite, we may choose $k = 6$ to see that the contribution of the second integral is $O((\log n)^{-3})$. In the first integral, the contribution of the $O(\xi^3/\sqrt{n})$ term is bounded by $O((\log n)^{\frac{3}{2}}/n^{\frac{1}{2}})$. Finally, the contribution of the ξ^2 term to the first integral is bounded by the integral over the entire space which is just $E(b_c)$. Hence $E(\Delta_c^*|n)$ converges to $E(b_c)$ as n increases. To get convergence of $E(\Delta_c^*)$ as N increases, note that $n > Np\delta$, for abitrary positive δ except on a set whose probability converges exponentially fast to zero, $E(\Delta_c^*|n)$ is bounded by n^2, hence by N^2 on the exceptional set, and elsewhere, convergence can be established as we have just discussed. This completes the proof.

[*Received May 1975. Revised July 1977.*]

REFERENCES

Blumenthal, Saul (1977), "Estimating Population Size with Truncated Sampling," *Communications in Statistics*, A6, 297–308.

Dahiya, Ram C., and Gross, Alan J. (1973), "Estimating the Zero Class from a Truncated Poisson Sample," *Journal of the American Statistical Association*, 68, 731–733.

Draper, N., and Guttman, I. (1971), "Bayesian Estimation of the Binomial Parameter *n*," *Technometrics*, 13, 667–673.

Feldman, Dorian, and Fox, Martin (1968), "Estimation of the Parameter *n* of the Binomial Distribution," *Journal of the American Statistical Association*, 63, 150–158.

Hodges, J.L., Jr., and Lehmann, E.L. (1970), "Deficiency," *Annals of Mathematical Statistics*, 41, 783–801.

Johnson, Norman L., and Kotz, Samuel (1969), *Discrete Distributions*, New York: John Wiley & Sons.

Nagaev, S.V. (1965), "Some Limit Theorems for Large Deviations," *Theory of Probability and its Applications*, 10, 214–235.

Sanathanan, Lalitha (1972), "Estimating the Size of a Multinomial Population," *Annals of Mathematical Statistics*, 43, 142–152.

——— (1977), "Estimating the Size of a Truncated Sample," *Journal of the American Statistical Association*, 72, 669–672.

Tocher, K.D. (1972), *The Art of Simulation*, London: The English Universities Press, Ltd.