Estimating the Size of a Truncated Sample

LALITHA SANATHANAN*

Consider N (unknown) values of a variable X (discrete or continuous) independently sampled from a distribution with density $f(x, \theta)$, where θ is an unknown vector parameter. Suppose that the values of X belonging to a certain region R are not observable. This article deals with the problem of estimating N and θ in such situations which arise, for instance, in life testing and capture-recapture census. Asymptotic theory for maximum likelihood estimation of N and θ is presented here and is shown to yield as corollaries both some existing results and a new result pertaining to the truncated negative binomial distribution.

KEY WORDS: Truncated sampling; Truncated negative binomial; Sample size estimation; Maximum likelihood; Conditional maximum likelihood; Asymptotic theory.

1. INTRODUCTION AND SUMMARY

The problem of estimating the size of a truncated sample has been dealt with in various articles for several specific cases. The following life testing situation is considered by Blumenthal and Marcus (1975): Suppose that a known number M of items are put on a life test, and that out of these M items there is an unknown number N having a defect of a particular type, which is identifiable only after the item fails. (Failure can occur among the $N - M$ other items as well as the M items of interest, although diagnosis is possible only after the item fails.) Suppose further that the test is terminated after a burn-in period T. Assuming that the life time of an item with this particular defect is the variable of interest, the observations here form a truncated sample in that N, and hence the number of missing observations of life times greater than T, are unknown. Under the assumption of an exponential distribution for life time, Blumenthal and Marcus (1975) give results pertaining to the estimation of N and also deal with second order asymptotic properties of the estimators considered. Estimation of N is discussed by Dahiya and Gross (1973), who consider a truncated Poisson sample, where all the observations assuming the zero value are missing. The same problem is considered by Blumenthal, Dahiya, and Gross (1974), with emphasis on second order asymptotic properties of the estimators of N. The general problem of estimating N from truncated multinomial data is dealt with in Sanathanan (1972a), while other articles by Sanathanan (1972b, 1973) focus on specific truncated multinomial models for visual scanning experiments. Problems involving truncated multinomial data of the multiple recapture census type are considered by Fienberg (1972), who uses the contingency table approach to solve such problems. Other articles dealing with the estimation of N are by Johnson (1962) and Marcus and Blumenthal (1974) for the continuous case and by Draper and Guttman (1971) and Feldman and Fox (1968) for the discrete case.

In this article, we develop the asymptotic theory for the unconditional maximum likelihood (uml) and conditional maximum likelihood (cml) estimation of N, both of which are defined in Section 2. Section 2 also contains some notation and other preliminaries. Section 3 contains results applicable to any multiparameter family of distributions considered for X (which may be discrete or continuous). Section 4 contains a discussion of special cases. Besides providing a unifying theory to cover the various specific situations which have already been dealt with in the literature, this article gives a basis for deriving new results, such as the one pertaining to the truncated negative binomial distribution given in Section 4.

2. LIKELIHOOD

Consider N (unknown) values of a variable X (discrete or continuous) independently sampled from a distribution with density $f(x, \theta)$ (with respect to a σ-finite measure), where θ is an unknown vector parameter. Suppose there is a region R such that X is not observable if $X \in R$. Each of the X values is thus observed with probability $p(\theta) = P(X \in \bar{R}) > 0$. This filtering process results in a random number $n \leq N$ of observations, which may be labeled $x_1, x_2, \ldots, x_n$. The likelihood function is given by

$$L(N; \theta) = L_1(N; p(\theta))L_2(\theta) \ ,$$

where

$$L_1(N; p(\theta)) = \frac{N!}{n!(N-n)!}(p(\theta))^n(1 - p(\theta))^{N-n}$$

and

$$L_2(\theta) = \prod_{i=1}^{n} g(x_i, \theta) \ ,$$

with

$$p(\theta) = P(X \in \bar{R})$$

and

$$g(x_i, \theta) = f(x_i, \theta)/p(\theta) \ .$$

As in the case of the truncated multinomial studied in Sanathanan (1972a), both the uml and cml estimates of

* Lalitha Sanathanan is Associate Professor, Department of Quantitative Methods, University of Illinois at Chicago Circle, Chicago, IL 60680. The author wishes to thank the referees and the editor for helpful suggestions.

© Journal of the American Statistical Association
September 1977, Volume 72, Number 359
Theory and Methods Section

(N, θ), denoted by $(\hat{N}_U, \hat{\theta}_U)$ and $(\hat{N}_C, \hat{\theta}_C)$, respectively, are considered here.

The uml and cml estimates are defined in such a way that $(\hat{N}_U, \hat{\theta}_U)$ maximizes $L(N; \theta)$ simultaneously with respect to N and θ, whereas $(\hat{N}_C, \hat{\theta}_C)$ is obtained by first obtaining $\hat{\theta}_C$ as the value of θ which maximizes $L_2(\theta)$ and then maximizing $L_1(N; p(\hat{\theta}_C))$ with respect to N. Since $L_2(\theta)$ is the likelihood based on the conditional distribution of the x's given n, and since $\hat{\theta}_C$ involves only maximizing $L_2(\theta)$, $\hat{\theta}_C$ is referred to as the cml estimate of θ. Also, $\hat{N}_C$ is based on n and $\hat{\theta}_C$ and is hence referred to as the cml estimate of N. Given θ, $L_1(N; p(\theta))$ is maximized with respect to N by $N = [n/p(\theta)]$ (a standard result), where the notation $[y]$ is used to denote the greatest integer less than y. When $p(\theta) = n/N'$ for some integer N', N can be either N' or $N' - 1$. Thus $\hat{N}_U = [n/p(\hat{\theta}_U)]$ and $\hat{N}_C = [n/p(\hat{\theta}_C)]$.

The basic difference between the cml and uml approaches is as follows: $\hat{\theta}_C$ is derived by solely maximizing $L_2(\theta)$ with respect to θ, whereas $\hat{\theta}_U$ maximizes $L_1(N; p(\theta))L_2(\theta)$ with respect to θ, N being subject only to the constraint that it is an integer. Further, since $(N, p(\theta)) = (n, 1)$ maximizes $L_1(N; p(\theta))$, and since $\theta = \hat{\theta}_C$ maximizes $L_2(\theta)$, it follows that $p(\hat{\theta}_C) < p(\hat{\theta}_U) < 1$ and hence $\hat{N}_U < \hat{N}_C$. However, $\hat{N}_U$ and $\hat{N}_C$ are asymptotically equivalent, as can be seen from the asymptotic distributions of $(\hat{N}_U - N)/\sqrt{N}$ and $(\hat{N}_C - N)/\sqrt{N}$ given by Theorem 2, which is stated in the next section.

3. ASYMPTOTIC THEORY

Let $(\hat{N}, \hat{\theta})$ be an estimate of (N, θ), and let r be the dimension of the vector θ. In Theorem 1, we give a set of conditions for the $(r + 1)$-dimensional random vector $A(\hat{N}, \hat{\theta}) = (N^{\frac{1}{2}}(\hat{\theta} - \theta), N^{-\frac{1}{2}}(\hat{N} - N))$ to have a limiting normal distribution $N(0, \Sigma)$, with mean 0 and covariance matrix Σ. Theorem 2 states that both $(\hat{N}_U, \hat{\theta}_U)$ and $(\hat{N}_C, \hat{\theta}_C)$ satisfy the conditions specified in Theorem 1 and hence that $A(\hat{N}_U, \hat{\theta}_U)$ and $A(\hat{N}_C, \hat{\theta}_C)$ have limiting distributions specified by that theorem.

Theorems 1 and 2 are generalizations of the corresponding theorems for the truncated multinomial given in Sanathanan (1972a). The assumptions and notation used in the sequel are similar to those in the previous article and are stated again. The proofs of the more general results given here involve a suitable modification of the methods used in Sanathanan (1972a) and are omitted.

Let $A1$ be the assumption that at every admissible value of θ, $p(\theta)$ admits continuous first-order partial derivatives and $f(x, \theta)$ admits continuous first and second order partial derivatives almost everywhere with respect to x.

Let the partial derivatives of $p(\theta)$, $\log f(x, \theta)$, $\log g(x, \theta)$, $\log L(N; \theta)$, and $\log L_2(\theta)$, with respect to θ_j (the jth component of θ), be denoted by $p_j(\theta)$, $f_j(x, \theta)$, $g_j(x, \theta)$, $L_j(N; \theta)$, and $L_{2,j}(\theta)$, respectively.

Let $A2$ be the assumption that $L_2(\theta)$ is a one-to-one function of θ almost everywhere with respect to $(x_1, x_2, \ldots, x_n)$, for any given n.

We now define the following quantities which appear in the statement of Theorem 1.

Let A be the $r \times r$ matrix whose (j, m)th element a_{jm} is given by

$$a_{jm} = p_j(\theta)p_m(\theta)/[p(\theta)(1 - p(\theta))] + p(\theta)E(g_j(x, \theta)g_m(x, \theta)) \ ,$$

or equivalently, by

$$a_{jm} = p_j(\theta)p_m(\theta)/(1 - p(\theta)) + p(\theta)E(f_j(x, \theta)f_m(x, \theta)) \ .$$

(In each case, expectation is taken with respect to the truncated distribution of X, namely $g(x, \theta)$.)

Let a_o be the r-dimensional row vector whose mth element a_{om} is given by

$$a_{om} = p_m(\theta)/(1 - p(\theta)) \ .$$

Finally, let a_{oo} be the scalar given by

$$a_{oo} = p(\theta)/(1 - p(\theta)) \ .$$

Theorem 1: Assume $A1$ and let $\hat{N}$ and $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_r)$ be estimates of N and θ, respectively, such that, as $N \to \infty$,

(i) $\hat{\theta} \to \theta$ a.s.
(ii) $N^{-\frac{1}{2}}(\hat{N} - n/p(\hat{\theta})) \to 0$ a.s.
(iii) $N^{-\frac{1}{2}}L_j(\hat{N}; \hat{\theta}) \to 0$ a.s. for $j = 1, \ldots, r$.

Then $A(\hat{N}, \hat{\theta}) = (N^{\frac{1}{2}}(\hat{\theta} - \theta), N^{-\frac{1}{2}}(\hat{N} - N))$ is asymptotically normal $N(0, \Sigma)$, where Σ^{-1} is the $(r + 1) \times (r + 1)$ matrix given by

$$\Sigma^{-1} = \begin{bmatrix} A & a_o' \\ a_o & a_{oo} \end{bmatrix},$$

(A, a_o, and a_{oo} as previously defined), and a_o' is the transpose of a_o.

In the case of a discrete (finite or infinite) distribution with a missing category, A is the usual information matrix for the corresponding complete or untruncated distribution. In the case of a truncated continuous distribution, A is the information matrix corresponding to the situation where we have a censored sample (i.e., N is known, although only observations of X outside R are available).

Theorem 2: Assume $A1$ and $A2$. Then, as $N \to \infty$,

(i) $(\hat{N}_C | N, \hat{\theta}_C, p(\hat{\theta}_C)) \to (1, \theta, p(\theta))$ a.s.
(ii) $(\hat{N}_C | N, \hat{\theta}_U, p(\hat{\theta}_U)) \to (1, \theta, p(\theta))$ a.s.
(iii) $A(\hat{N}_U, \hat{\theta}_U)$ and $A(\hat{N}_U, \hat{\theta}_C)$ are both asymptotically $N(0, \Sigma)$, where Σ is as defined in Theorem 1.

Theorem 2 is a corollary of Theorem 1, and its proof reduces to showing that both $(\hat{N}_C, \hat{\theta}_C)$ and $(\hat{N}_U, \hat{\theta}_U)$ satisfy the conditions specified in Theorem 1.

4. SPECIAL CASES

Specific cases of truncated sampling involving the exponential, Poisson, and multinomial distributions have already been cited in Section 1. The relevant asymptotic distributions in these situations can be derived from

Theorem 2. This will be demonstrated through the example of the truncated multinomial. In addition, a new result pertaining to the negative binomial distribution is derived here.

4.1 The Truncated Multinomial

Consider the multinomial distribution, with a missing cell. Suppose there are $l+1$ categories labeled 0, 1, ..., l and that the 0 category is not observable. Let $p_k(\theta)$ be the probability of an observation belonging to the kth category. Then

$$f(x_i, \theta) = p_k(\theta)$$

if the ith observation falls into the kth category. Here,

$$R = \{0\}$$

and

$$p(\theta) = 1 - p_0(\theta) \ .$$

By partially differentiating $\log f(x_i, \theta)$ with respect to θ_j, we have

$$f_j(x_i, \theta) = p_{k,j}(\theta)/p_k(\theta)$$

if the ith observation falls into the kth category ($p_{k,j}(\theta)$ is the partial derivative of $p_k(\theta)$ with respect to θ_j). Expectation of $f_j(x, \theta)f_m(x, \theta)$, with respect to the truncated distribution of X, is thus given by

$$\frac{1}{p(\theta)} \sum_{k=1}^{l} p_{k,j}(\theta)p_{k,m}(\theta)/p_k(\theta) \ ,$$

implying that

$$a_{jm} = \sum_{k=0}^{l} p_{k,j}(\theta)p_{k,m}(\theta)/p_k(\theta) \ .$$

Thus A is the usual information matrix for the complete or untruncated multinomial with $l+1$ cells.

4.2 The Truncated Negative Binomial

The negative binomial probability function has the form

$$f(x_i, k, \pi) = \frac{(k + x_i - 1)}{(k-1)!x_i!} \frac{\pi^{x_i}}{(1+\pi)^{k+x_i}} \ , \qquad x_i = 0, 1, \ldots, \quad (4.1)$$

with unknown scalar parameters $k, \pi > 0$. The corresponding truncated distribution $g(x_i, k, \pi)$ is given by dividing (4.1) by

$$p(\theta) = p(k, \pi) = 1 - 1/(1+\pi)^k \ .$$

The maximum likelihood solution $\hat{\theta}_C$ for $\theta = (k, \pi)$, based on n (fixed) observations from a truncated negative binomial, is given in Sampford (1955). This method, however, requires a special aid table. A method proposed by Hartley (1958) uses an iterative approach for the truncated situation based on the method developed by Bliss and Fisher (1953) for estimating k and π from a complete sample. This iterative method is also illustrated in Hartley (1958) with the help of some real data. Once we obtain $(\hat{k}_C, \hat{\pi}_C)$ using Hartley's method, $\hat{N}_C$ is given by

$$\hat{N}_C = [n/(1 - 1/(1+\hat{\pi}_C)^{\hat{k}}C)] \ .$$

To get the asymptotic variance of $N^{-\frac{1}{2}}(\hat{N}_C - N)$, we need to apply the general result for Σ^{-1} in Theorem 1. This involves the calculation of $E(g_j(x, \theta)g_m(x, \theta))$. In practice, however, the actual value of $E(g_j(x, \theta)g_m(x, \theta))$ cannot be calculated, since the values of k and π occurring in it are not known and can only be estimated. Therefore, we must content ourselves with its estimate. Consider the following standard result:

$$E(g_j(x, \theta)g_m(x, \theta)) = -E \frac{\partial^2 \log g(x, \theta)}{\partial\theta_m \partial\theta_j} \cdot$$

It is thus enough to substitute for $E(g_j(x, \theta)g_m(x, \theta))$ an estimate of $-E(\partial^2 \log g(x, \theta)/\partial\theta_m\partial\theta_j)$, which is provided by

$$-\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log g(x_i, \theta)}{\partial\theta_m\partial\theta_j} = -\frac{1}{n}\frac{\partial^2 \log L_2(\theta)}{\partial\theta_m\partial\theta_j} \cdot$$

Specific expressions are given in Sampford (1955) for $\partial^2 \log L_2(\theta)/\partial\theta_m\partial\theta_j$, with j and m ranging from 1 to 2, $\theta_1 = k$, and $\theta_2 = 1/(1+\pi)$. Using those results in terms of θ_2 reparameterized as π, and completing the rest of the calculations involved in getting Σ^{-1}, we can state the following result: Both

$$A(\hat{N}_C, \hat{k}_C, \hat{\pi}_C) \quad \text{and} \quad A(\hat{N}_U, \hat{k}_U, \hat{\pi}_U)$$

are asymptotically normal $N(0, \Sigma)$. An estimate of Σ^{-1} is provided by the matrix (s^{ij}), with

$$s^{11} = \frac{1}{n}\sum_{t=1}^{Y} (\hat{k} + t - 1)^{-2} \sum_{t=1}^{Y} n_t \ ,$$

where Y denotes the highest observed value of X, and n_t is the number of observations with value t. Also,

$$s^{12} = s^{21} = 1/(1+\hat{\pi}) \ ,$$

$$s^{22} = \frac{\hat{k}(\hat{k}-1)}{(1+\hat{\pi})^2} + \frac{a((1+\hat{\pi})^{\hat{k}} - 1)(1+2\hat{\pi})}{\hat{\pi}^2(1+\hat{\pi})^{\hat{k}+2}} \ ,$$

where a denotes the sample mean,

$$s^{13} = s^{31} = \log(1+\hat{\pi}) \ ,$$

$$s^{23} = s^{32} = \hat{k}/(1+\hat{\pi}) \ ,$$

and

$$s^{33} = (1+\hat{\pi})^{\hat{k}} - 1 \ .$$

[*Received January 1976. Revised February 1977.*]

REFERENCES

Bliss, C. I., and Fisher, Ronald A. (1953), "Fitting the Negative Binomial Distribution to Biological Data," *Biometrics*, 9, 176–96.

Blumenthal, Saul, and Marcus, Richard (1975), "Estimating Population Size with Exponential Failure," *Journal of the American Statistical Association*, 28, 913–22.

———, Dahiya, Ram C., and Gross, Alan J. (1974), "Estimating the Complete Sample Size from an Incomplete Poisson Sample," Technical Report No. 79, Department of Statistics, University of Kentucky.

Dahiya, Ram C., and Gross, Alan J. (1973), "Estimating the Zero Class from a Truncated Poisson Sample," *Journal of the American Statistical Association*, 68, 731-3.

Draper, Norman, and Guttman, Irwin (1971), "Bayesian Estimation of the Binomial Parameter," *Technometrics*, 13, 667-73.

Feldman, Dorian, and Fox, Martin (1968), "Estimation of the Parameter n in the Binomial Distribution," *Journal of the American Statistical Association*, 63, 150-8.

Fienberg, Stephen E. (1972), "The Multiple-Recapture Census for Closed Populations and Incomplete 2^k Contingency Tables," *Biometrika*, 59, 591-603.

Hartley, H.O. (1958), "Maximum Likelihood Estimation from Incomplete Data," *Biometrics*, 14, 174-94.

Johnson, Norman L. (1962), "Estimation of Sample Size," *Technometrics*, 4, 59-67.

Marcus, Richard, and Blumenthal, Saul (1974), "A Sequential Screening Procedure," *Technometrics*, 16, 229-34.

Sampford, M.R. (1955), "The Truncated Negative Binomial Distribution," *Biometrika*, 42, 58-69.

Sanathanan, Lalitha (1972a), "Estimating the Size of a Multinomial Population," *Annals of Mathematical Statistics*, 43, 142-52.

——— (1972b), "Models and Estimation Methods in Visual Scanning Experiments," *Technometrics*, 14, 813-30.

——— (1973), "A Comparison of Some Models in Visual Scanning Experiments," *Technometrics*, 15, 67-78.