THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ON ESTIMATING THE PARAMETER OF A TRUNCATED GEOMETRIC DISTRIBUTION

by

R. L. THOMASSON and C. H. KAPADIA

Technical Report No. 5
Department of Statistics THEMIS Contract

Department of Statistics
Southern Methodist University
Dallas, Texas 75222

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ON ESTIMATING THE PARAMETER OF A TRUNCATED GEOMETRIC DISTRIBUTION

by

R. L. THOMASSON and C. H. KAPADIA

Technical Report No. 5
Department of Statistics THEMIS Contract

September 6, 1968

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

On Estimating the Parameter of a
Truncated Geometric Distribution

by

R. L. Thomasson
Texas Instruments, Inc., Dallas

and

C. H. Kapadia*
Southern Methodist University

1. Introduction and Summary

This paper concerns the use of the method of maximum likelihood in estimating the parameter of the geometric distribution from samples which are truncated at arbitrary points in either or both tails of the distribution. It is shown that the maximum likelihood estimator is the solution of a polynomial of high degree, and a table is given for solving the maximum likelihood estimating equation. The asymptotic variance of the estimator is presented and this result is used in studying the effect of truncation on the asymptotic efficiency of maximum likelihood estimation. The efficiency of the maximum likelihood estimator for truncated samples relative to the maximum likelihood estimator for complete samples is derived, and it is shown that for any given sample size the estimator for truncated samples always provides less information about the parameter than the one for complete samples.

The geometric probability mass function is given by

*This research was accomplished while under partial support from NIH Grant GM-951 and ONR Grant #N0014-68-A-0515.

$$f(x) = p(1-p)^{x-1} = pq^{x-1} \quad ; \; x = 1,\ 2,\ \ldots$$
$$= \qquad 0 \qquad , \text{ otherwise} \qquad (1.1)$$

and the cumulative distribution function F(x) is

$$F(x) = 1-q^x,\ x=1,\ 2,\ 3,\ldots$$

where $q = 1-p$. The mean and the variance of the random variable X having this probability mass function are $1/p$ and q/p^2 respectively. The maximum likelihood estimator of p based on a complete random sample of size n say $X_1, X_2, \ldots, X_n$ is $\left(\sum_{i=1}^{n} X_i/n\right)^{-1}$.

Let X be a random variable having probability function (1.1) and if the truncation points on the left and the right are at a and b respectively then the truncated probability function of X from (1.1) is

$$f(x) = \begin{cases} pq^{x-1}/(q^a-q^{b-1}) & , \ x = a+1,\ a+2,\ \ldots,\ b-1 \\ 0 & , \text{ otherwise.} \end{cases} \qquad (1.2)$$

The transformations $Y = X - a$ and $d = b - a$ reduce (1.2) to

$$g(y) = \begin{cases} \dfrac{pq^{y-1}}{1-q^{d-1}} & , \ y=1,\ 2,\ 3,\ \ldots,\ d-1 \\ 0 & , \text{ otherwise} \end{cases} . \qquad (1.3)$$

In effect, because of the lack of memory property enjoyed by this distribution [1], (1.3) is a geometric distribution singly truncated on the right at the point $Y = d$. In the derivation of the maximum likelihood estimator the transformed variable Y and the mass function (1.3) will be used.

2. Maximum Likelihood Estimation

For a random sample of size n, the likelihood function is

$$L = (1-q^{d-1})^{-n} p^n \, q^{\sum_{i=1}^{n} y - n} .$$

The following equation (2.1), which is a dth degree polynomial in $\hat{q}$, gives the maximum likelihood estimator of $q(\hat{p} = 1-\hat{q}$ by the invariance property).

$$\left(\sum_{i=1}^{n} y_i - nd + n\right)\hat{q}^d + \left(nd - \sum_{i=1}^{n} y_i\right)\hat{q}^{d-1} - \left(\sum_{i=1}^{n} y_i\right)\hat{q} + \sum_{i=1}^{n} y_i - n = 0. \quad (2.1)$$

Given values of d, n, and $\sum_{i=1}^{n} y_i = n\bar{y}$, one can compute the value of $\hat{q}$ using an iterative technique such as the Newton-Rhapson method to solve equation (2.1). It can be shown that there is only one root in the range $0 < \hat{q} < 1$.

In order to avoid the necessity for an iterative solution to equation (2.1), a table which can be used to obtain approximate solutions for $\hat{p}$ is constructed. From equation (2.1),

$$\bar{y} = \{(d-1)\hat{q}^d - d\hat{q}^{d-1} + 1\}/(\hat{q}^d - \hat{q}^{d-1} - \hat{q} + 1). \qquad (2.2)$$

In Table 1 values for d in the range $3 \le d \le 27$ are given in the left hand column of the page. The selected values for $\hat{p}$ are given across the top of the page. For each pair of values of d and $\hat{p}$, the value of $\bar{y}$ computed from equation (2.2) is given in the table.

To use Table 1 in solving for $\hat{p}$, use the row of values for $\bar{y}$ corresponding to a specified value for the parameter d. Once $\bar{y}$ is located in the table, the value of $\hat{p}$ can be read from the top line of the page. The rate of change of $\bar{y}$ with respect to $\hat{p}$ seems to be sufficiently constant to permit linear interpolation.

3. The Asymptotic Variance and Efficiency

For the random variable Y with the probability function (1.3), after some algebraic manipulation, and using the fact that

$$E(Y) = [(1-dq^{d-1} + (d-1)q^{d}]/(1-q)(1-q^{d-1}),$$

it can be shown by following the method [2, p. 236] that

$$\sigma^2_{\hat{p}} = (1/n)[q^2(1-q)^2(1-q^{d-1})^2]/[q-(d-1)^2q^{d-1}+2d(d-2)q^{d}-(d-1)^2q^{d+1}+q^{2d-1}]. \quad (3.1)$$

where E is the expected value operator and $\sigma^2_{\hat{p}}$ denotes the asymptotic variance of $\hat{p}$.

The asymptotic efficiency of one estimator relative to a second estimator is defined as the ratio of their asymptotic variances. The asymptotic variance of the maximum likelihood estimator for truncated samples is given by (3.1). For the complete sample estimator ($\hat{p}_c = 1/\bar{x}$), the asymptotic variance is given by

$$\sigma^2_{\hat{p}_c} = q(1-q)^2/n.$$

The equation (3.1) can be written in the form

$$\sigma^2_{\hat{p}} = \sigma^2_{\hat{p}_c}\left[1/[1-(d-1)^2q^{d-2}\{(1-q)(1-q^{d-1})^{-1}\}^2] .\right] \quad (3.2)$$

Hence the efficiency ϵ of $\hat{p}$ relative to $\hat{p}_c$ is given by

$$\epsilon = \sigma^2_{\hat{p}_c}/\sigma^2_{\hat{p}} = 1-(d-1)^2q^{d-2}\{(1-q)(1-q^{d-1})^{-1}\}^2 . \quad (3.3)$$

The second term on the right is always positive and varies between 0 and 1, depending on the value of q:

$$\lim_{q \to 0} \epsilon = 1 , \qquad \lim_{q \to 1} \epsilon = 0 .$$

Consequently, the maximum likelihood estimator for truncated samples is in general less efficient than the maximum likelihood estimator for complete samples.

If Fisher's terminology is used and the information function I is defined as

$$I = -E(d^2 \log L/dq^2) = [\sigma^2_{\hat{p}}]^{-1} ,$$

then it can be seen from equation (3.3) that an estimator based on a truncated sample of size

$$n\left[1/[1-(d-1)^2 q^{d-2}\{(1-q)(1-q^{d-1})^{-1}\}^2]\right]$$

provides as much information about the parameter as an estimate based on a complete sample of size n.

Table II presents values of ϵ computed for selected values of p and d. The table indicates that the truncated sample estimator is very inefficient if either p or d is small. In these cases it is advisable to use relatively large sample sizes in order to reduce the variance of the estimator. On the other hand, as p takes on larger values the estimator becomes highly efficient. Consequently, in cases in which it is reasonable to assume that the value of the parameter is large (near unity), it may be economically advantageous to use truncated samples to estimate the parameter, rather than going to the additional expense of obtaining a complete sample.

TABLE I

APPROXIMATE SOLUTIONS FOR THE MAXIMUM LIKELIHOOD ESTIMATOR

d / p	0.001	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.999
3	1.500	1.474	1.444	1.412	1.375	1.333	1.286	1.231	1.167	1.091	1.001
4	1.999	1.930	1.852	1.767	1.673	1.571	1.462	1.345	1.226	1.108	1.001
5	2.499	2.369	2.225	2.069	1.904	1.733	1.562	1.396	1.244	1.111	1.001
6	2.998	2.790	2.563	2.323	2.078	1.839	1.615	1.416	1.248	1.111	1.001
7	3.497	3.195	2.868	2.533	2.206	1.905	1.642	1.424	1.250	1.111	1.001
8	3.996	3.582	3.142	2.705	2.298	1.945	1.655	1.427	1.250	1.111	1.001
9	4.495	3.953	3.387	2.844	2.363	1.969	1.661	1.428	1.250	1.111	1.001
10	4.993	4.308	3.605	2.955	2.408	1.982	1.664	1.428	1.250	1.111	1.001
11	5.492	4.647	3.797	3.043	2.439	1.990	1.666	1.429	1.250	1.111	1.001
12	5.990	4.969	3.966	3.111	2.460	1.995	1.666	1.429	1.250	1.111	1.001
13	6.488	5.277	4.115	3.165	2.474	1.997	1.666	1.429	1.250	1.111	1.001
14	6.986	5.569	4.244	3.206	2.483	1.998	1.667	1.429	1.250	1.111	1.001
15	7.484	5.847	4.356	3.238	2.489	1.999	1.667	1.429	1.250	1.111	1.001
16	7.981	6.111	4.453	3.262	2.493	2.000	1.667	1.429	1.250	1.111	1.001
17	8.479	6.360	4.537	3.280	2.495	2.000	1.667	1.429	1.250	1.111	1.001
18	8.976	6.597	4.608	3.294	2.497	2.000	1.667	1.429	1.250	1.111	1.001
19	9.473	6.821	4.670	3.304	2.498	2.000	1.667	1.429	1.250	1.111	1.001
20	9.970	7.033	4.722	3.312	2.499	2.000	1.667	1.429	1.250	1.111	1.001
21	10.467	7.232	4.767	3.317	2.499	2.000	1.667	1.429	1.250	1.111	1.001
22	10.963	7.429	4.805	3.322	2.500	2.000	1.667	1.429	1.250	1.111	1.001
23	11.460	7.597	4.836	3.325	2.500	2.000	1.667	1.429	1.250	1.111	1.001
24	11.956	7.763	4.863	3.327	2.500	2.000	1.667	1.429	1.250	1.111	1.001
25	12.452	7.920	4.886	3.329	2.500	2.000	1.667	1.429	1.250	1.111	1.001
26	12.948	8.066	4.905	3.330	2.500	2.000	1.667	1.429	1.250	1.111	1.001
27	13.444	8.204	4.921	3.331	2.500	2.000	1.667	1.429	1.250	1.111	1.001

TABLE II

THE EFFECT OF TRUNCATION ON THE EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATION

d / p	0.01	0.05	0.10	0.20	0.40	0.60	0.80
5	0.0001	0.0033	0.0138	0.0599	0.2701	0.6117	0.9178
10	0.0007	0.0174	0.0708	0.2748	0.7779	0.9809	0.9999
15	0.0016	0.0417	0.1624	0.5284	0.9590	0.9995	1.0000
20	0.0030	0.0753	0.2757	0.7322	0.9941	1.0000	1.0000
25	0.0048	0.1171	0.3972	0.8627	0.9993	1.0000	1.0000
30	0.0070	0.1655	0.5153	0.9347	0.9999	1.0000	1.0000
35	0.0097	0.2189	0.6220	0.9707	1.0000	1.0000	1.0000
40	0.0127	0.2759	0.7131	0.9874	1.0000	1.0000	1.0000
45	0.0161	0.3347	0.7873	0.9947	1.0000	1.0000	1.0000
50	0.0200	0.3941	0.8455	0.9979	1.0000	1.0000	1.0000

References

(1) Feller, William. <u>An Introduction to Probability Theory and its Application</u>. John Wiley and Sons, 1957.

(2) Mood, A. M., Greyhill, F. A. <u>Introduction to the Theory of Statistics</u>, McGraw-Hill Book Co., New York, 1963.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author)
SOUTHERN METHODIST UNIVERSITY

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP
UNCLASSIFIED

REPORT TITLE
On Estimating the Parameter of a
Truncated Geometric Distribution

DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

AUTHOR(S) (First name, middle initial, last name)
R. L. Thomasson
C. H. Kapadia

REPORT DATE
September 6, 1968

7a. TOTAL NO. OF PAGES
8

7b. NO. OF REFS
2

CONTRACT OR GRANT NO.
N00014-68-A-0515

PROJECT NO.
NR 042-260

9a. ORIGINATOR'S REPORT NUMBER(S)
5.

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

DISTRIBUTION STATEMENT
No limitations

SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research

ABSTRACT

This paper is concerned with the problem of maximum likelihood estimation
r the parameter, of the geometric distribution, from samples which are truncated
arbitrary points in either or both tails of the distribution. It is shown that
e maximum likelihood estimator is the solution of a polynomial of high degree,
d a table is given for solving the maximum likelihood estimating equation. The
rivation of the asymptotic variance of the estimator is presented and the result
used in the study of the asymptotic efficiency of the estimation in truncated
mpling. The efficiency of the maximum likelihood estimator for complete samples
derived and it is shown that the estimator based on truncated samples always
ovide less information about the parameter than that based on complete sample
the same size.

FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification