ON ESTIMATING THE PARAMETER OF A TRUNCATED GEOMETRIC DISTRIBUTION BY THE METHOD OF MOMENTS

by

C. H. Kapadia and R. L. Thomasson

Technical Report No. 95
Department of Statistics ONR Contract

January 7, 1971

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

DEPARTMENT OF STATISTICS
Southern Methodist University

and

$$E(Y) = [1-dq^{d-1} + (d-1)q^{d}][(1-q^{d-1})(1-q)]^{-1} \quad . \tag{1.1}$$

Let equation (1.1) equal to the sample mean and this will give the following equation for the estimator $\hat{q}_m$

$$\bar{y} = [1-d\hat{q}_m^{d-1} + (d-1)\hat{q}_m^{d}][1-\hat{q}_m - \hat{q}_m^{d-1} + \hat{q}_m]^{-1} \quad .$$

This equation is identical to equation (2.1) given by the authors [4] which was used in constructing Table I [4]. Therefore, the method of moment estimator is identical to the maximum likelihood estimator.

2. Derivation of an Estimator Based on Sample Moments

Let f_y denote the frequency with which the value y occurs in the sample before truncation as suggested by Rider [2, 3]. The expected value of f_y is $Nf(y)$, where N is the (unknown) total size of the untruncated sample and $f(y)$ is the geometric probability mass function evaluated at the point $Y = y$. Define the following functions:

$$T_0 = \sum_{y=1}^{d-1} f_y \ , \ T_1 = \sum_{y=1}^{d-1} yf_y \ , \ T_2 = \sum_{y=1}^{d-1} y^2 f_y \ ,$$

$$T_0' = T_0 + \sum_{y=d}^{\infty} Npq^{y-1} \ , \ T_1' = T_1 + \sum_{y=d}^{\infty} yNpq^{y-1} \ , \ T_2' = T_2 + \sum_{y=d}^{\infty} y^2 Npq^{y-1} \quad .$$

The ratio T_1'/T_0' is an estimator of the first moment of the complete distribution, namely $1/p$. The ratio T_2'/T_1' is an estimator of the second non-central moment $(q + 1)/p^2$, divided by the first moment, this quotient being $(q + 1)/p$. The following two equations in two unknowns N and p can be derived by setting $T_0 = pT_1'$ and $pT_2' = (q + 1)T_1'$:

$$T_0 + N \sum_{y=d}^{\infty} pq^{y-1} = pT_1 + \sum_{y=d}^{\infty} ypq^{y-1} \tag{2.1}$$

$$pT_2 + Np \sum_{y=d}^{\infty} y^2 pq^{y-1} = (q+1)T_1 + N(q+1) \sum_{y=d}^{\infty} ypq^{y-1} \quad . \tag{2.2}$$

From the above two equations, the modified method of moments estimator p^* is

$$p^* = \frac{dT_0 - 2T_1}{(d-1)T_1 - T_2}$$

$$= (dn - 2 \sum_{i=1}^{n} y_i)/[(d-1) \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i^2] \quad ,$$

where $n = T_0, \sum_{i=1}^{n} y_i = T_1$ and $\Sigma y_i^2 = T_2$.

In terms of the sample mean and variance, p^* is

$$p^* = (d-2\bar{y})/[(d-2)\bar{y} - \bar{y}^2 - s^2] \tag{2.3}$$

where

$$s^2 = [\sum_{i=1}^{n} (y_i - \bar{y})^2]/n \quad .$$

The Asymptotic Variance of the Moment Estimator

Cramér [1, pp. 366-67] has proved that any sample characteristic based on moments is asymptotically normally distributed about the corresponding population characteristic. Cramér's theorems 27.7.3 and 28.4 give the variance of the asymptotic distribution. Let $H_{\bar{y}^2}$ and H_{s^2} denote the first order partial derivative of p^*, given in equation (2.3), evaluated at the point $\bar{y} = \mu$ and $s^2 = \text{var}(Y) = \mu_2$ where var means variance and cov will be used for covariance. The asymptotic variance of p^* is given by the leading terms of the equation

$$\mathrm{var}(p^*) = \mathrm{var}(\bar{y})H_{\bar{y}^2} + 2\mathrm{cov}(\bar{y},\ s^2)H_{\bar{y}}H_{s^2} + \mathrm{var}(s^2)H_{s^2}^2 \quad . \tag{3.1}$$

In order to evaluate var(p*), use equation $\mathrm{var}(\bar{y}) = \frac{1}{n}\mathrm{var}(Y) = \frac{1}{n}\mu_2$ and Cramér's equations 27.4.2 and 27.4.4 for $\mathrm{var}(s^2)$ and $\mathrm{cov}(\bar{y},\ s^2)$:

$$\mathrm{var}(s^2) = \frac{\mu_4-\mu_2^2}{n} - \frac{2\left(\mu_4 - 2\mu_2^2\right)}{n^2} + \frac{\mu_4-3\mu_2^2}{n^3} \quad ,$$

$$\mathrm{cov}(\bar{y},\ s^2) = \frac{n-1}{n^2}\mu_3 \quad .$$

In these equations μ_i indicate the i^{th} central moment of the truncated geometric distribution.

Differentiating p* with respect to $\bar{y}$ and s^2 and evaluating $\frac{\partial p^*}{\partial y}$ and $\frac{\partial p^*}{\partial s^2}$ at $\bar{y} = \mu$ and $s^2 = \mu_2$, equation (3.1) can be written in the following form:

$$\mathrm{var}(p^*) = \frac{1}{n}\left\{\mu_2(2d\mu-2\mu^2-d^2+d)^2 + 2\mu_3(2d\mu-2\mu^2+2\mu_2-d^2+d) \cdot (d-2\mu) + (\mu_4-\mu_2)^2(d-2\mu)^2\right\} \cdot \left\{(d-1)\mu-\mu^2-\mu_2\right\}^{-4} \quad . \tag{3.2}$$

4. <u>The Efficiency of the Moments Estimator Relative to the Maximum Likelihood Estimator</u>

It is clear that an estimator based on sample moments is very easy to compute and it is suggested that in cases in which computational simplicity is important, the modified moments estimator might be preferable to the maximum likelihood estimator. This section investigates the loss of efficiency incurred when the modified moments estimator is used.

In order to compute the efficiency of the modified moments estimator relative to the maximum likelihood estimator, evaluate the ratio of the asymptotic variance of the two estimators. The asymptotic variance of the maximum likelihood estimator is given by equation (3.1), [4]. Equation (3.2) gives the asymptotic variance of the moments estimator. The desired efficiency is given by the ratio of equation (3.1) of the authors [4] to equation (3.2).

The equations for the asymptotic variances are too complicated to permit any analytic conclusions. Table I presents selected values of the efficiency of the moments estimator relative to the maximum likelihood estimator. The table indicates that the modified moments estimator is highly efficient for all tabulated values of the parameters. Consequently, when computational simplicity is an important factor, the modified moments estimator is a suitable alternative to the maximum likelihood estimator.

TABLE IV

THE EFFICIENCY OF THE MODIFIED MOMENTS ESTIMATOR RELATIVE TO THE MAXIMUM LIKELIHOOD ESTIMATOR

d / p	0.01	0.05	0.10	0.20	0.40	0.60	0.80
5	1.0000	0.9995	0.9978	0.9906	0.9602	0.9226	0.9220
10	0.9999	0.9967	0.9867	0.9517	0.8963	0.9285	0.9784
15	0.9995	0.9919	0.9697	0.9146	0.9121	0.9689	0.9921
20	0.9994	0.9854	0.9501	0.8951	0.9434	0.9843	0.9959
25	0.9990	0.9776	0.9313	0.8933	0.9646	0.9906	0.9975
30	0.9985	0.9687	0.9152	0.9026	0.9766	0.9938	0.9983
35	0.9981	0.9594	0.9032	0.9166	0.9836	0.9956	0.9988
40	0.9975	0.9498	0.8957	0.9312	0.9879	0.9967	0.9991
45	0.9968	0.9404	0.8921	0.9440	0.9907	0.9974	0.9993
50	0.9960	0.9314	0.8921	0.9545	0.9926	0.9980	0.9994

REFERENCES

[1] Cramér, Harold (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

[2] Rider, P. R. (1953). "Truncated Poisson Distribution," Journal of the American Statistical Association, 48, 826-30.

[3] Rider, P. R. (1955). "Truncated Binomial and Negative Binomial Distributions," Journal of the American Statistical Association, 50, 877-83.

[4] Thomasson, R. L. and Kapadia, C. H. (1968). "On Estimating the Parameter of a Truncated Geometric Distribution," Annals of the Institute of Statistical Mathematics, 20, 519-23.