The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 1078–1082

THE GEOMETRIC DENSITY WITH UNKNOWN LOCATION PARAMETER

By Jerome Klotz

University of Wisconsin

1. Summary. Unbiased estimators are derived for a sample from the geometric density with unknown p and unknown location parameter. Mean square errors are compared with the maximum likelihood estimator and unbiased tests of hypotheses are given.

2. Model and sufficient statistics. Let $X_1, X_2, \cdots, X_n$ have the discrete geometric density

$$P[X_i = x_i] = q^{x_i - \nu} p \qquad (x_i = \nu, \nu+1, \cdots, \infty) \tag{2.1}$$

where the vector parameter $\boldsymbol{\theta} = (\nu, p)$ is unknown, $q = 1-p$, and ν is the location parameter. When p is known, $X_{(1)} = \min X_i$ is sufficient for ν. Further, $X_{(1)}$ is complete and has a distribution given by

$$P[X_{(1)} = x] = {q_n}^{x-\nu} p_n \qquad (x = \nu, \nu+1, \cdots, \infty) \tag{2.2}$$

where $q_n = q^n$, $p_n = 1-q^n$. Using (2.1) and the factorization theorem, we see that $(X_{(1)}, \sum X_i)$ or equivalently $(X_{(1)}, U)$ is sufficient for $\boldsymbol{\theta}$ where $U = \sum(X_i - X_{(1)})$. By Basu's theorem [1], $X_{(1)}$ and U are independent since the distribution of U does not depend on ν.

3. Distribution of U. The joint distribution of the order statistics $X_{(1)} \leqq X_{(2)} \leqq \cdots \leqq X_{(n)}$ can be written

$$\begin{aligned} P[X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}, \cdots, X_{(n)} = x_{(n)}] \\ = \left[\frac{n!}{\prod_k t_k!}\right] q^{n(x_{(1)}-\nu)}(1-q^n) I_{[x_{(1)} \geqq \nu]} \\ \cdot q^{\Sigma(x_{(1)} - X_{(1)})} \frac{p^n}{1-q^n} I_{[x_{(1)} \leqq x_{(2)} \leqq \cdots \leqq x_{(n)}]} \end{aligned} \tag{3.1}$$

where t_k is the number of x_i equal to the value $k = 0, 1, 2, \cdots, \infty$. Thus

$$\begin{aligned} P[X_{(1)} = x_{(1)}, U = u] = q^{n(x_{(1)}-\nu)}(1-q^n) I_{[x_{(1)} \geqq \nu]} \\ \cdot q^u \frac{p^n}{1-q^n} \sum \left(\frac{n!}{\prod t_k!} I_{[x_{(1)} \leqq \cdots \leqq x_{(n)}]}\right) \end{aligned} \tag{3.2}$$

Received August 4, 1969.

[1] Research partially supported by the Math Research Center U.S. Army, under Contract No. DA-31-124-ARO-D-462.

where the sum is over some region that depends only upon n and u using the independence. If we call this sum $g_n(u)$, we have

$$P[U = u] = q^u \frac{p^n}{1-q^n} g_n(u)$$

and we can determine $g_n(u)$ by summing the probabilities to one:

$$\sum_{n=0}^{\infty} q^u \frac{p^n}{1-q^n} g_n(u) = 1 \quad \text{or} \quad \sum_{n=0}^{\infty} g_n(u) q^u = (1-q^n)(1-q)^{-n}.$$

Equating coefficients of the power series we have

$$g_n(u) = \binom{n+u+1}{u} - \binom{u-1}{u-n},$$

with the usual zero convention for negative arguments of binomial coefficients. Hence

$$\text{(3.3)} \qquad P[U = u] = \left(\binom{n+u-1}{u} - \binom{u-1}{u-n}\right) q^u \frac{p^n}{1-q^n}$$
$$= \frac{1}{1-q^n}\binom{n+n-1}{u} q^u p^n - \frac{q^n}{1-q^n}\binom{u-1}{n-n} q^{u-n} p^n.$$

4. Unbiased estimators of $\boldsymbol{\theta}$. Since (3.3) belongs to the exponential family, u is complete for the family with $0 < p < 1$. Therefore $(X_{(1)}, U)$ is jointly sufficient and jointly complete for $\boldsymbol{\theta}$ and the usual theory of minimum variance unbiased estimation works. For the unbiased estimator of p, we solve for $h(u)$ in the equation

$$\text{(4.1)} \qquad \sum_{u=0}^{\infty} h(u)\left(\binom{n+u-1}{u} - \binom{u-1}{u-n}\right) q^u \frac{p^n}{1-q^n} = p,$$

to obtain

$$\text{(4.2)} \qquad h(u) = \left[\binom{n+u-2}{u} - \binom{u-2}{u-n}\right] \Big/ \left[\binom{n+u-1}{u} - \binom{u-1}{u-n}\right].$$

To obtain the minimum variance unbiased estimator of v, we note that

$$\text{(4.3)} \qquad EX_{(1)} = v + q^n/(1-q^n).$$

Thus we similarly derive the unbiased estimator $f(u)$ for $q^n/(1-q^n)$ to be

$$\text{(4.4)} \qquad f(u) = \binom{u-1}{u-n} \Big/ \left[\binom{n+u-1}{u} - \binom{u-1}{u-n}\right],$$

and construct the unbiased estimator of v to be

$$\text{(4.5)} \qquad X_{(1)} - \left(\binom{U-1}{U-n} \Big/ \left[\binom{n+U-1}{U} - \binom{U-1}{U-n}\right]\right).$$

The mean square error for estimator (4.2) is compared with that of the maximum likelihood estimator $\hat{p} = n/(n+U)$ in Table 1, and a similar comparison is given for (4.5) and the m.l.e. $\hat{v} = X_{(1)}$ in Table 2.

The values, believed accurate to within one unit in the last place, were checked by various methods. Probabilities were summed to one to $6\frac{1}{2}$ decimal places, and checks from $Eh(U) = p$, $Ef(U) = q^n/(1-q^n)$ were obtained. In addition, for $n = 2$ the mean square error of $f(u)$ simplifies to give $[q^2(2-p)/(2p^2(1+q))]+q^2/(1-q^2)$. The number of terms used varied between 170 for $n = 2$ to 680 for $n = 20$. The large number of terms was required for the accuracy given because of heavy tails in the distribution for the smallest value of $p = .1$. An additional check was made

TABLE 1
Mean square error comparison of unbiased and m.l. *estimators of p*

M.S.E.		$p = .1$	.3	.5	.7	.9
$n = 2$	unbiased	6.632(−2)[1]	1.482(−1)	1.667(−1)	1.292(−1)	4.909(−2)
	m.l.($\hat{p}$)	6.177(−2)	4.335(−2)	1.378(−2)	2.188(−2)	7.381(−2)
$n = 5$	unbiased	3.769(−3)	1.994(−2)	3.219(−2)	3.311(−2)	1.677(−2)
	m.l.($\hat{p}$)	7.074(−3)	1.714(−2)	1.500(−2)	1.693(−2)	2.277(−2)
$n = 10$	unbiased	1.222(−3)	7.516(−3)	1.384(−2)	1.557(−2)	8.251(−3)
	m.l.($\hat{p}$)	1.689(−3)	7.718(−3)	9.714(−3)	1.125(−2)	9.621(−3)
$n = 15$	unbiased	7.226(−4)	4.653(−3)	8.907(−3)	1.019(−2)	5.469(−3)
	m.l.($\hat{p}$)	8.734(−4)	4.928(−3)	6.953(−3)	8.253(−3)	6.035(−3)
$n = 20$	unbiased	5.124(−4)	3.388(−3)	6.570(−3)	7.569(−3)	4.089(−3)
	m.l.($\hat{p}$)	5.817(−4)	3.590(−3)	5.394(−3)	6.481(−3)	4.392(−3)

[1] The number in parenthesis is the exponent or power of 10 so that 6.632(−2) represents .06632.

TABLE 2
Mean square error comparison of unbiased and m.l. *estimators of v*

M.S.E.		$p = .1$	.3	.5	.7	.9
$n = 2$	unbiased	4.476(1)	3.683(0)	8.333(−1)	1.907(−1)	1.627(−2)
	m.l.($\hat{v}$)	4.061(1)	2.807(0)	5.556(−1)	1.185(−1)	1.031(−2)
$n = 5$	unbiased	4.378(0)	2.897(−1)	3.725(−2)	2.578(−3)	1.018(−5)
	m.l.($\hat{v}$)	5.600(0)	2.837(−1)	3.434(−2)	2.448(−3)	1.000(−5)
$n = 10$	unbiased	9.047(−1)	3.130(−2)	9.898(−4)	5.915(−6)	1.000(−10)
	m.l.($\hat{v}$)	1.109(0)	3.076(−2)	9.794(−4)	5.905(−6)	1.000(−10)
$n = 15$	unbiased	3.454(−1)	4.855(−3)	3.056(−5)	1.435(−8)	1.000(−15)
	m.l.($\hat{v}$)	3.937(−1)	4.816(−3)	3.052(−5)	1.435(−8)	1.000(−15)
$n = 20$	unbiased	1.634(−1)	8.020(−4)	9.538(−7)	3.487(−11)	1.000(−20)
	m.l.($\hat{v}$)	1.767(−1)	7.998(−4)	9.537(−7)	3.487(−11)	1.000(−20)

by computing the probabilities by two methods and the m.s.e. of the m.l. estimator $\hat{v} = X_{(1)}$ was completed from $q^n(1+q^n)/(1-q^n)^2$.

The results indicate roughly that the maximum likelihood estimator of p is better than the unbiased estimator for the middle values of p, while the unbiased is better for extreme values of p. For estimating v, the unbiased is better for small p values with the m.l. estimator better for moderate to large values, although the difference is slight for large n and p.

5. Tests of hypotheses. For simplicity, we shall restrict attention to one-sided hypotheses although they are easily modified for two-sided hypotheses ([3] Chapter 4).

For testing the hypothesis

$$H_v\colon v \leqq 0 \quad \text{against the alternative} \quad A_v\colon v > 0,$$

we construct a u.m.p. unbiased test by selecting the best similar test on the boundary $v = 0, 0 < p < 1$. On this boundary, the statistic $S = \sum X_i$ is sufficient and complete and under the general model $S - nv$ has the negative binomial distribution with parameters n, p. It is easy to show for a fixed value $s \geqq nv$, that the conditional likelihood ratio of the sample given $S = s$ is monotone in $X_{(1)}$, and so the u.m.p. unbiased level α test rejects with probability

$$\begin{aligned} \phi(x_{(1)}) &= 1 \quad &&\text{if} \quad x_{(1)} > C(s) \\ &= \gamma \quad &&\text{if} \quad x_{(1)} = C(s) \\ &= 0 \quad &&\text{if} \quad x_{(1)} < C(s) \end{aligned}$$

where $C(s), \gamma(s)$ are uniquely determined from

$$\textstyle\sum_{x_{(1)}=0}^{\infty} \phi(x_{(1)})\left[\binom{n+s-nx_{(1)}-1}{s-nx_{(1)}} - \binom{s-nx_{(1)}-1}{s-nx_{(1)}-n}\right]\Big/\binom{n+s-1}{s} = \alpha.$$

For testing the hypothesis

$$H_p\colon p \leqq p_0 \quad \text{against the alternative} \quad A_p\colon p > p_0$$

we similarly construct the u.m.p. unbiased test by finding the best similar test on the boundary $p = p_0$, $-\infty < v < \infty$. On this boundary, $X_{(1)}$ is sufficient and complete. Reducing by sufficiency and using the independence of U and $X_{(1)}$, we see that the u.m.p. similar test is based upon U alone. Since the distribution of U given by (3.3) is in the exponential family, the u.m.p. unbiased level α test rejects with probability

$$\begin{aligned} \phi(u) &= 1 \quad &&\text{if} \quad u < C \\ &= \gamma \quad &&\text{if} \quad u = C \\ &= 0 \quad &&\text{if} \quad u > C \end{aligned}$$

where C, γ are uniquely determined so that

$$\textstyle\sum_{u=0}^{\infty} \phi(u)\left(\binom{n+u-1}{u} - \binom{u-1}{u-n}\right) q_0^{\,u} p_0^{\,n}/(1-q_0^{\,n}) = \alpha.$$

6. Comments. The relationship with the continuous exponential density with location parameter μ given by $\lambda e^{-\lambda(t-\mu)}$ for $t > \mu$ is seen by letting the random variables X_i be the number of time intervals of length r before a failure. With $\mu = rv$, $p = r\lambda$, and $T_i = rX_i$ (the time to failure) we see that the geometric distribution converges to the exponential as $r \to 0$. The unbiased estimator for μ in the exponential distribution is given by $T_{(1)} - \sum_i (T_i \cdots T_{(1)})/n(n-1)$ which can be obtained as a limit from (4.5) after multiplying by r. Similarly for λ, the unbiased estimator $(n-2)/\sum_i (T_i - T_{(1)})$ can also be obtained from (4.2) by dividing by r and taking the limit.

7. Acknowledgments. Thanks go to V. Erickson for programming and to B. Harris and R. C. Milton for helpful conversations.

REFERENCES

[1] Basu, D., (1958). On statistics independent of a sufficient statistic. *Sankhyā* **20** 223.
[2] Ferguson, T. S. (1967). *Mathematical Statistics*. Academic Press, New York.
[3] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.