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

\[
P[X_i = x_i] = q^{x_i-1}p \quad (x_i = v, v+1, \cdots, \infty)
\]

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

\[
P[X_{(1)} = x] = q_n^{x-v}p_n \quad (x = v, v+1, \cdots, \infty)
\]

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 \( \theta \) 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 \( v \).

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

\[
P[X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}, \cdots, X_{(n)} = x_{(n)}] = \frac{n!}{\prod k!} q^{n(x_{(1)}-1)}(1-q^n)I[x_{(1)} \leq x]
\]

\[
q^\sum(x_{(i)}-x_{(1)}) \cdot \frac{p^n}{1-q^n} I[x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}]
\]

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

\[
P[X_{(1)} = x_{(1)}, U = u] = q^{n(x_{(1)}-v)}(1-q^n)I[x_{(1)} \geq v]
\]

\[
q^u \cdot \frac{p^n}{1-q^n} \sum \left( \frac{n!}{\prod k!} I[x_{(1)} \leq \cdots \leq x_{(n)}] \right)
\]

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.

1078
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

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

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

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

Equating coefficients of the power series we have

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

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

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

4. Unbiased estimators of \( \theta \). 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 \( \theta \) and the usual theory of minimum variance unbiased estimation works. For the unbiased estimator of \( p \), we solve for \( h(u) \) in the equation

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

to obtain

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

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

\[
(4.3) \quad EX_{(1)} = v + q^n/(1 - q^n).
\]

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

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

and construct the unbiased estimator of \( v \) to be

\[
(4.5) \quad X_{(1)} - \left( \frac{v}{U - 1} \right)/\left[ \binom{n + U - 1}{U} - \binom{U - 1}{U - 1} \right].
\]

The mean square error for estimator (4.2) is compared with that of the maximum likelihood estimator \( \hat{\theta} = n/(n + U) \) 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 6½ decimal places, and checks from $E_h(U) = p$, $E_f(U) = q^n/(1-q^n)$ 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>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean square error comparison of unbiased and m.1. estimators of $p$</td>
</tr>
<tr>
<td>M.S.E.</td>
</tr>
<tr>
<td>$n = 2$</td>
</tr>
<tr>
<td>m.1.($\hat{p}$)</td>
</tr>
<tr>
<td>$n = 5$</td>
</tr>
<tr>
<td>m.1.($\hat{p}$)</td>
</tr>
<tr>
<td>$n = 10$</td>
</tr>
<tr>
<td>m.1.($\hat{p}$)</td>
</tr>
<tr>
<td>$n = 15$</td>
</tr>
<tr>
<td>m.1.($\hat{p}$)</td>
</tr>
<tr>
<td>$n = 20$</td>
</tr>
<tr>
<td>m.1.($\hat{p}$)</td>
</tr>
</tbody>
</table>

$^1$ The number in parenthesis is the exponent or power of 10 so that 6.632(-2) represents .06632.

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean square error comparison of unbiased and m.1. estimators of $v$</td>
</tr>
<tr>
<td>M.S.E.</td>
</tr>
<tr>
<td>$n = 2$</td>
</tr>
<tr>
<td>m.1.($\hat{f}$)</td>
</tr>
<tr>
<td>$n = 5$</td>
</tr>
<tr>
<td>m.1.($\hat{f}$)</td>
</tr>
<tr>
<td>$n = 10$</td>
</tr>
<tr>
<td>m.1.($\hat{f}$)</td>
</tr>
<tr>
<td>$n = 15$</td>
</tr>
<tr>
<td>m.1.($\hat{f}$)</td>
</tr>
<tr>
<td>$n = 20$</td>
</tr>
<tr>
<td>m.1.($\hat{f}$)</td>
</tr>
</tbody>
</table>
by computing the probabilities by two methods and the m.s.e. of the m.l. estimator \( \hat{\theta} = X_{(1)} \) 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 \( \nu \), 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

\[ H_0: \nu \leq 0 \] against the alternative \( A_\nu: \nu > 0 \),

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

\[
\phi(x_{(1)}) = \begin{cases} 
1 & \text{if } x_{(1)} > C(s) \\
\gamma & \text{if } x_{(1)} = C(s) \\
0 & \text{if } x_{(1)} < C(s)
\end{cases}
\]

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

\[
\sum_{x_{(1)} = 0}^{\infty} \phi(x_{(1)})[(n + s - nx_{(1)} - 1)!(s - nx_{(1)} - 1)!]/(n + s - 1)! = \alpha.
\]

For testing the hypothesis

\[ H_p: p \leq p_0 \] against the alternative \( A_p: p > p_0 \)

we similarly construct the u.m.p. unbiased test by finding the best similar test on the boundary \( p = p_0 \), \( -\infty < \nu < \infty \). 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 \( \alpha \) test rejects with probability

\[
\phi(u) = \begin{cases} 
1 & \text{if } u < C \\
\gamma & \text{if } u = C \\
0 & \text{if } u > C
\end{cases}
\]

where \( C, \gamma \) are uniquely determined so that

\[
\sum_{u=0}^{\infty} \phi(u)[(n + u - 1)!/(u)!]q_0^u p_0^n/(1 - q_0^n) = \alpha.
\]
6. Comments. The relationship with the continuous exponential density with location parameter $\mu$ 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 $\mu$ in the exponential distribution is given by $T_{(1)} - \sum_i(T_i - T_{(1)})/n(n-1)$ which can be obtained as a limit from (4.5) after multiplying by $r$. Similarly for $\lambda$, 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