ORDER STATISTICS OF GEOMETRIC DISTRIBUTION

Chaobing He*

*Corresponding Author:

Abstract:
This paper mainly studies the order statistics of geometric distribution. The paper deduces the joint frequency function and conditional joint frequency function of the order statistics, and, obtain and prove some important propositions of order statistics of geometric distribution. Certain propositions are different from and also similar to corresponding propositions of exponential distribution.

Index Terms: Geometric distribution, order statistics, exponential distribution, joint frequency function, identical distribution
I. INTRODUCTION

Geometric distribution has already been applied to more fields, and it has an extremely important position especially in some fields such as information engineering, electronic engineering, control theory and economics. It is well known that exponential distribution plays quite an important role in the statistical analysis of reliability. However in discrete life case, geometric distribution play the role of exponential distribution in continuous life case, so the study on geometric distribution becomes more and more important. [1] first proposed that the characteristics of geometric distribution might be described by order statistics. [2] made the further study on order statistics of geometric distribution. [3] obtained certain characterizations of exponential and geometric distributions. [4] studied a characterization of the geometric distribution. [5] proved a characterization of the geometric distribution. [6] gave a note on characterizations of the geometric distribution. [7] obtained some results for type I censored sampling from geometric distributions. [8] gave and proved two characterizations of geometric distributions. [9] compared some characterizations of the geometric with exponential random variables. [10] made statistical analysis for geometric distribution based on records. [11] got a generalization of the geometric distribution. [12] gives a generalization of geometric distribution. [13] proved characterizations of the geometric distribution via residual lifetime. Although both geometric distribution and exponential distribution have no memory, properties of their order statistics make an obvious difference because of their individual differences. This paper obtains and proves some propositions of order statistics of geometric distribution, and certain propositions are different from and also similar to corresponding propositions of exponential distribution.

II. THE RESULTS AND PROOFS

random variable X is said to have a geometric distribution with parameter p if its probability function is \( P(X = k) = p(1-p)^{k-1} \) for \( k = 1, 2, \ldots \). This work was supported in part by the Key Scientific Research Program of Colleges and Universities of Henan Province of China under Grant 16A110001 and Grant 18A110009. C. He is with the School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China (Email: chaobing5@163.com.) where \( 0 < p < 1 \) and \( q = 1 - p \). We will sometimes write \( X \sim \text{Geo}(p) \). Suppose that \( X_1, \ldots, X_n \) are i.i.d. geometric random variables. We arrange \( X_i \)'s in ascending order, and some of \( X_i \)'s are taken as the same group whose values are equal. Therefore \( X_i \)'s are divided into finite group. Then we define \( Y_i \) to be the number of variables included by the \( i \)-th group and \( X(i) \) the common value of the \( i \)-th group random variables with \( 1 \leq i \leq n \). Let \( D_i = X(i) - (i-1) \) with \( X(0) = 0 \).

**Proposition 1**: Based on above symbols defined, the following are consequences of \( X_i \)'s:

(i) The joint probability function of \( X(1), X(2), \ldots, X(r) \) is

\[
P(X(1) = k_1, X(2) = k_2, \ldots, X(r) = k_r) = P(K_{r1}) = P(X = k_1)P(X = k_2) \cdots P(X = k_r)
\]

(ii) Conditional \( Y_1 = m_1, Y_2 = m_2, \ldots, Y_{r-1} = m_{r-1}, \) \( Y_r \sim \text{Geo}(1 - q^{m_r} - \cdots - m_{r-1}) \), where \( m_1 + m_2 + \cdots + m_{r-1} \leq r \).

(iii) Conditional \( Y_1 = m_1, Y_2 = m_2, \ldots, Y_{r-1} = m_{r-1}, \) \( Y_r \sim \text{Geo}(1 - q^{m_r} - \cdots - m_{r-1}) \), where \( m_1 + m_2 + \cdots + m_{r-1} \leq r \).

(iv) Conditional \( Y_1 = m_1, \ldots, Y_{r-2} = m_{r-2}, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1} \), \( k_1, \ldots, k_{r-1} \) are i.i.d. geometric random variables with \( 1 \leq i \leq r \), \( Y_r \sim \text{Geo}(1 - q^{m_r} - \cdots - m_{r-1}) \), where \( m_1 + m_2 + \cdots + m_{r-1} \leq r \).

(v) The joint probability function of \( Y_1, Y_2, \ldots, Y_{r-1} \) is

\[
Y_1 = m_1, Y_2 = m_2, \ldots, Y_{r-1} = m_{r-1}, Y_r = m_r \sim \text{Geo}(1 - q^{m_1} - \cdots - m_{r-1}) \), where \( m_1 + m_2 + \cdots + m_{r-1} \leq r \).

(vi) Conditional \( Y_1 = m_1, \ldots, Y_{r-1} = m_{r-1}, \) \( X_{r-1} = k_{r-1} \), \( D_r \sim \text{Geo}(1 - q^{m_r} - \cdots - m_{r-1}) \), where \( m_1 + m_2 + \cdots + m_{r-1} \leq r \).

Set \( A_i = B \) with \( i = 1, \ldots, N \). It is easy to check that \( A_i \) is the values of \( X_i \)'s are only \( k_1, \ldots, k_i, k_{i+1}, \ldots, k_{r-1} \), k_{r+1} more than \( k_r \).

By applying properties of probability and combinatorial arguments, it follows that

\[
P(X(1) = k_1, \ldots, X(r) = k_r, N \geq r) = \]

\[
P(B \cap C_1) = P(B \cap B \cap \bigcup_{i=1}^{r} C_i) = P(B) - P\left( \bigcup_{i=1}^{r} B \cap C_i \right) = P(B) - P\left( \bigcup_{i=1}^{r} B \cap C_i \right) = P(B) - P\left( \bigcup_{i=1}^{r} A_i \right)
\]

\[
P(B) - \sum_{i=1}^{r} P(A_i) - \sum_{1 \leq i < j \leq r} P(A_i A_j) - \sum_{1 \leq i < j < k \leq r} P(A_i A_j A_k) + \cdots + (-1)^{r} P(A_1 \cdots A_r)
\]

\[
= \left( \sum_{i=1}^{r} p_{k_i} + \hat{p}_{k_i} \right)^n - \sum_{1 \leq i < j < k < \cdots < r} \left( \sum_{j=1}^{r-1} p_{k_{ij}} + \hat{p}_{k_{ij}} \right)^n - \sum_{1 \leq i < j < k < \cdots < r} \left( \sum_{j=1}^{r-2} p_{k_{ij}} + \hat{p}_{k_{ij}} \right)^n.
\]
(ii) The joint frequency function of \((X_1, Y_1), \ldots, (X_r, Y_r)\) is

\[
P(X_1 = k_1, Y_1 = m_1, \ldots, X_r = k_r, Y_r = m_r, N \geq r) = \frac{n!}{m_1! \cdots m_r!} [P(X = k_1)]^{m_1} \cdots [P(X = k_r)]^{m_r} [P(X > k_r)]^{m_{r+1}}
\]

\[
= \frac{n!}{m_1! \cdots m_r!} (pq^{k_1-1})^{m_1} \cdots (pq^{k_r-1})^{m_r} \left( \sum_{k=k_r+1}^{\infty} pq^{k-1} \right)^{m_{r+1}}
\]

\[
= \frac{n!}{m_1! \cdots m_r!} \left[ \sum_{m} m q^{\sum_{r=1}^{r} m_i} m_i \cdots m_{r+1} \right] (q^{\sum_{r=1}^{r} m_i} - 1)^{m_{r+1}}
\]

where \(m_1 + \cdots + m_{r+1} = n\) and \(1 \leq k_1 < \cdots < k_r\).

From Equation (6), one has

\[
P(X_r = k_r, Y_r = m_r, N \geq r | X_1 = k_1, Y_1 = m_1, \ldots, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r - 1)
\]

\[
= \frac{P(X_1 = k_1, Y_1 = m_1, \ldots, X_r = k_r, Y_r = m_r, N \geq r)}{P(X_1 = k_1, Y_1 = m_1, \ldots, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r - 1)}
\]

\[
= \left( \frac{n - m_1 - \cdots - m_{r-1}}{m_r} \right) \left( \frac{p}{q} \right)^{m_r} \left( q^{n-m_1-\cdots-m_{r-1}} - 1 \right)
\]

therefore,

\[
P(X_r = k_r, N \geq r | Y_1 = m_1, \ldots, Y_{r-2} = m_{r-2}, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r - 1)
\]

\[
= \sum_{m_r=1}^{n-m_1-\cdots-m_{r-1}} \left( \frac{n - m_1 - \cdots - m_{r-1}}{m_r} \right) \left( q^{n-m_1-\cdots-m_{r-1}} - 1 \right)
\]

\[
= (q^{n-m_1-\cdots-m_{r-1}})^{k_r} (q^{n-m_1-\cdots-m_{r-1}} - 1)
\]

hence,

\[
P(X_r - X_{r-1} = k, N \geq r | Y_1 = m_1, \ldots, Y_{r-2} = m_{r-2}, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r - 1)
\]

\[
= \left(1 - q^{n-m_1-\cdots-m_{r-1}} \right) (q^{n-m_1-\cdots-m_{r-1}})^{k_r}
\]

(7)

Obviously, the right of Equation (7) has nothing to do with \(k_{r-1}\), hence we have

\[
P(X_r - X_{r-1} = k, N \geq r | Y_1 = m_1, \ldots, Y_{r-1} = m_{r-1}, N \geq r - 1)
\]

\[
= \left(1 - q^{n-m_1-\cdots-m_{r-1}} \right) (q^{n-m_1-\cdots-m_{r-1}})^{k_r}
\]

That is to say: Conditional \(Y_1 = m_1, \ldots, Y_{r-1} = m_{r-1}, D_r \sim Geom(1 - q^{n-m_1-\cdots-m_{r-1}})\).

(iii) Firstly, it can be obtained that

\[
P(Y_r = m_r, N \geq r | Y_1 = m_1, \ldots, Y_{r-2} = m_{r-2}, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r - 1)
\]

\[
= \sum_{k_{r-1}=k_r}^{\infty} \left( \frac{n - m_1 - \cdots - m_{r-1}}{m_r} \right) \left( q^{n-m_1-\cdots-m_{r-1}} - 1 \right)
\]

\[
= \left( \frac{n - m_1 - \cdots - m_{r-1}}{m_r} \right) \left( \frac{p}{q} \right)^{m_r} \left( 1 - q^{n-m_1-\cdots-m_{r-1}} \right)
\]

\[
P(Y_r = m_r, N \geq r | Y_1 = m_1, \ldots, Y_{r-1} = m_{r-1}, N \geq r - 1)
\]
therefore, conditional $Y_1 = m_1$, \ldots, $Y_{r-1} = m_{r-1}$, the frequency function of $Y_r$ is

$$P(Y_r = m_r | Y_1 = m_1, \ldots, Y_{r-1} = m_{r-1}) = \binom{r}{m_r} \frac{(n - m_1 - \cdots - m_{r-1})}{n} \frac{q^{n-m_1-\cdots-m_{r-1}}}{1 - q^{n-m_1-\cdots-m_{r-1}}}.$$

(iv) Since

$$P(X_r = k_r, Y_r = m_r, N \geq r | Y_1 = m_1, \ldots, Y_{r-2} = m_{r-2}, X_{r-1} = k_{r-1})$$

and

$$P(Y_r = m_r, N \geq r | Y_1 = m_1, \ldots, Y_{r-2} = m_{r-2}, X_{r-1} = k_{r-1})$$

hence, conditional $Y_1 = m_1$, \ldots, $Y_{r-1} = k_{r-1}$, $Y_r = m_{r-1}$, variables $X_r$ and $Y_r$ are independent.

(v) Define $A = \frac{n}{m_1! \cdots m_r!} (pq^{-1})^{\sum_{i=1}^r m_i}$, it can be obtained that

$$P(Y_1 = m_1, \ldots, Y_r = m_r, N \geq r) = \sum_{k_1=1}^{\infty} \sum_{k_2=k_1+1}^{\infty} \cdots \sum_{k_r=k_{r-1}+1}^{\infty} \frac{(pq^{-1})^{\sum_{i=1}^r m_i} A^r}{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}})}$$

hence,

$$P(Y_1 = m_1, \ldots, Y_r = m_r, N \geq r) = \frac{n!}{m_1! \cdots m_r!} \frac{(pq^{-1})^{\sum_{i=1}^r m_i} A^r}{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}})},$$

where $m_1 + \cdots + m_r = n$.

The other solution is as follows.

It is easy to check that

$$P(X_1 = \cdots = X_r < \min \{X_{i1}, \ldots, X_{im} \}) = \sum_{k_1=1}^{\infty} \binom{pq^{-1}}{k_1} (pq^{-1})^{\sum_{i=1}^r m_i} = \frac{(pq^{-1})^{\sum_{i=1}^r m_i}}{1-q^{-m_r}}.$$

hence,

$$P(Y_1 = m_1, \ldots, Y_r = m_r, N \geq r)$$

$$= P(Y_1 = m_1) \cdots P(Y_r = m_r, N \geq 2 | Y_1 = m_1) \cdots P(Y_r = m_r, N \geq r | Y_1 = m_1, \ldots, Y_r = m_{r-1}, N \geq r - 1)$$

$$= \frac{(pq^{-1})^m A^r}{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}})} \frac{n!}{m_1! \cdots m_r!} \frac{(pq^{-1})^{\sum_{i=1}^r m_i} A^r}{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}})},$$

$$= \frac{n!}{m_1! \cdots m_r!} \frac{(pq^{-1})^{\sum_{i=1}^r m_i} A^r}{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}})}.$$

Let $m_1 = m_2 = \cdots = m_r = 1$ above, it follows that

$$P(X_i \neq X_j \text{ for } i \neq j, i, j = 1, 2, \ldots, n) = P(Y_1 = 1, \ldots, Y_n = 1) = n! \frac{(pq^{-1})^{\sum_{i=1}^r m_i} A^r}{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}})}.$$

(vi) From Equations (4) and (6), one has

$$P(X(1) = k_1, \ldots, X(r) = k_r | Y_1 = m_1, \ldots, Y_r = m_r, N \geq r)$$

$$= \frac{P(Y_1 = m_1, \ldots, Y_r = m_r, N \geq r)}{P(Y_1 = k_1, \ldots, Y_r = k_r | Y_1 = m_1, \ldots, Y_r = m_r, N \geq r)}$$

$$= \frac{(1-q^n)(1-q^{n-m_1}) \cdots (1-q^{n-m_1-\cdots-m_{r-1}}) q^{\sum_{i=1}^r m_i k_i} (n-m_1-\cdots-m_{r-1}) k_r - [n] (n-m_1) \cdots (n-m_1-\cdots-m_{r-1})]}{1-q^{n-m_1-\cdots-m_{r-1}} \cdots (1-q^{n-m_1-\cdots-m_{r-1}})}.$$
where $m_1 + \cdots + m_r \leq n$ and $1 \leq r_1 < \cdots < r_r$.

Therefore,

$$
P(D_{(1)} = y_1, \cdots, D_{(r)} = y_r | Y_1 = m_1, \cdots, Y_r = m_r, N \geq r)$$

$$= \frac{P(Y_1 = m_1, \cdots, Y_r = m_r, N \geq r)}{P(Y_1 = m_1, \cdots, Y_r = m_r, N \geq r)}$$

$$= (1 - q^n)(1 - q^{n-m_1}) \cdots (1 - q^{n-m_r})$$

$$\times q^{(m_1 y_1 + m_2 (y_1 + y_2) + \cdots + m_r (y_1 + \cdots + y_r) + (n-m_1-y_1) (y_1 + \cdots + y_r) - [n+(n-m_1)+\cdots+(n-m_r-y_1)+y_1]})$$

$$= (1 - q^n)(1 - q^{n-m_1}) \cdots (1 - q^{n-m_r})$$

$$\times q^{(m_1 + \cdots + m_r) y_1 + \cdots + (n-m_1-y_1) y_1 - [n+(n-m_1)+\cdots+(n-m_r-y_1)+y_1]}$$

$$= (1 - q^n)(q^n)^{n-1} \cdots (1 - q^{n-m_1}) (q^{n-m_1})^{y_1} \cdots (1 - q^{n-m_r}) (q^{n-m_r})^{y_r}.$$ 

where $m_1 + m_2 + \cdots + m_r \leq n$.

Hence, conditional $Y_1 = m_1, \cdots, Y_r = m_r$, variables $D_1, \cdots, D_r$ are independent and

$$D_i \sim \text{Geo}(1 - q^{n-m_1})$$

$i = 1, 2, \cdots, r$.

Let $m_1 = m_2 = \cdots = m_r = 1$ above, it follows that conditional $X_i \neq X_j$ ($i \neq j$), $i, j = 1, 2, \cdots, n$,

$$D_i \sim \text{Geo}(1 - q^{n-1})$$

$i = 1, 2, \cdots, n$.

The proof is complete.

**Proposition 2:** Suppose that $Y_1, Y_2, \cdots, Y_n$ are independent with $Y_i \sim \text{Geo}(1 - q)$, $i = 1, 2, \cdots, n$, then conditional $X_i \neq X_j$ ($i \neq j$),

$$X_{(n-k+1)}$$

and $\sum_{i=k}^{n} Y_i$ have an identical distribution.

**Proof:** Conditional $X_i \neq X_j$ ($i \neq j$), variables $Y_i$ and $D_{(n+1-i)}$ have an identical distribution, hence

$$\sum_{i=k}^{n} Y_i$$

and $\sum_{i=k}^{n} D_{(n+1-i)}$ have an identical distribution.

Moreover,

$$\sum_{i=k}^{n} D_{(n+1-i)} = \sum_{i=k}^{n} (X_{(n+1-i)} - X_{(n-i)})$$

$$= X_{(n+1-k)}.$$

From the above, it follows that $X_{(n-k+1)}$ and $\sum_{i=k}^{n} Y_i$ have an identical distribution.

The proof is complete.

**Corollary 1:** Suppose that $Z_1, Z_2, \cdots, Z_k$ are i.i.d. geometric variables with $k < n$, then $X_{(n-i)} - X_{(n-k)}$ given $X_i \neq X_j$ ($i \neq j$) and $Z_{(k-1)}$ given $Z_i \neq Z_j$ ($i \neq j$) are identically distributed.

**Proof:** It is easy to check that
III. CONCLUSION
The current work concerns the order statistics of geometric distribution. The joint frequency function and conditional joint frequency function of the order statistics has been obtained by applying properties of probability and combinatorial arguments. Several propositions of order statistics are very fresh, interesting and attractive. Results indicate that certain propositions are different from and also similar to corresponding propositions of exponential distribution. According to the theoretical conclusions of this paper, further topics will include the parameter estimation on the basis of observation data.

IV. CONFLICT OF INTEREST
The author declares that there is no conflict of interest regarding the publication of this paper.

REFERENCES