

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*

(ii) The joint frequency function of $(X_{(1)}, Y_1), \dots, (X_{(r)}, Y_r)$ is

$$\begin{aligned}
 & P(X_{(1)} = k_1, Y_1 = m_1, \dots, X_{(r)} = k_r, Y_r = m_r, N \geq r) \\
 &= \frac{n!}{m_1! \dots m_{r+1}!} [P(X = k_1)]^{m_1} \dots [P(X = k_r)]^{m_r} [P(X > k_r)]^{m_{r+1}} \\
 &= \frac{n!}{m_1! \dots m_{r+1}!} (pq^{k_1-1})^{m_1} \dots (pq^{k_r-1})^{m_r} \left(\sum_{k=k_r+1}^{\infty} pq^{k-1} \right)^{m_{r+1}} \\
 &= \frac{n!}{m_1! \dots m_{r+1}!} p^{\sum_{i=1}^r m_i} q^{\sum_{i=1}^r m_i k_i + m_{r+1} k_r - \sum_{i=1}^r m_i}, \tag{6}
 \end{aligned}$$

where $m_1 + \dots + m_{r+1} = n$ and $1 \leq k_1 < \dots < k_r$.

From Equation (6), one has

$$\begin{aligned}
 & P(X_{(r)} = k_r, Y_r = m_r, N \geq r | X_{(1)} = k_1, Y_1 = m_1, \dots, X_{(r-1)} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r-1) \\
 &= \frac{P(X_{(1)} = k_1, Y_1 = m_1, \dots, X_{(r)} = k_r, Y_r = m_r, N \geq r)}{P(X_{(1)} = k_1, Y_1 = m_1, \dots, X_{(r-1)} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r-1)} \\
 &= \binom{n - m_1 - \dots - m_{r-1}}{m_r} \left(\frac{p}{q} \right)^{m_r} (q^{n - m_1 - \dots - m_{r-1}})^{k_r - k_{r-1}},
 \end{aligned}$$

therefore,

$$\begin{aligned}
 & P(X_{(r)} = k_r, N \geq r | Y_1 = m_1, \dots, Y_{r-2} = m_{r-2}, X_{(r-1)} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r-1) \\
 &= \sum_{m_r=1}^{n - m_1 - \dots - m_{r-1}} \binom{n - m_1 - \dots - m_{r-1}}{m_r} (q^{n - m_1 - \dots - m_{r-1}})^{k_r - k_{r-1}} \\
 &= (q^{n - m_1 - \dots - m_{r-1}})^{k_r - k_{r-1} - 1} \left[\left(\frac{p}{q} + 1 \right)^{n - m_1 - \dots - m_{r-1}} - 1 \right] \\
 &= (1 - q^{n - m_1 - \dots - m_{r-1}}) (q^{n - m_1 - \dots - m_{r-1}})^{k_r - k_{r-1} - 1},
 \end{aligned}$$

hence,

$$\begin{aligned}
 & P(X_{(r)} - X_{(r-1)} = k, N \geq r | Y_1 = m_1, \dots, Y_{r-2} = m_{r-2}, X_{(r-1)} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r-1) \\
 &= (1 - q^{n - m_1 - \dots - m_{r-1}}) (q^{n - m_1 - \dots - m_{r-1}})^{k-1}. \tag{7}
 \end{aligned}$$

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

$$\begin{aligned}
 & P(X_{(r)} - X_{(r-1)} = k, N \geq r | Y_1 = m_1, \dots, Y_{r-1} = m_{r-1}, N \geq r-1) \\
 &= (1 - q^{n - m_1 - \dots - m_{r-1}}) (q^{n - m_1 - \dots - m_{r-1}})^{k-1}.
 \end{aligned}$$

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

(iii) Firstly, it can be obtained that

$$\begin{aligned}
 & P(Y_r = m_r, N \geq r | Y_1 = m_1, \dots, Y_{r-2} = m_{r-2}, X_{(r-1)} = k_{r-1}, Y_{r-1} = m_{r-1}, N \geq r-1) \\
 &= \sum_{k_r=k_{r-1}+1}^{\infty} \binom{n - m_1 - \dots - m_{r-1}}{m_r} (q^{n - m_1 - \dots - m_{r-1}})^{k_r - k_{r-1}} \\
 &= \binom{n - m_1 - \dots - m_{r-1}}{m_r} \left(\frac{p}{q} \right)^{m_r} \frac{q^{n - m_1 - \dots - m_{r-1}}}{1 - q^{n - m_1 - \dots - m_{r-1}}} \\
 &= P(Y_r = m_r, N \geq r | Y_1 = m_1, \dots, Y_{r-1} = m_{r-1}, N \geq r-1),
 \end{aligned}$$

therefore, conditional $Y_1 = m_1, \dots, Y_{r-1} = m_{r-1}$, the frequency function of Y_r is

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

(iv) Since

$$\begin{aligned} & P(X_{(r)} = k_r, Y_r = m_r, N \geq r | Y_1 = m_1, \dots, Y_{r-2} = m_{r-2}, X_{(r-1)} = k_{r-1}) \\ &= P(X_{(r)} = k_r, N \geq r | Y_1 = m_1, \dots, Y_{r-2} = m_{r-2}, X_{(r-1)} = k_{r-1}) \\ &\quad \times P(Y_r = m_r, N \geq r | Y_1 = m_1, \dots, Y_{r-2} = m_{r-2}, X_{(r-1)} = k_{r-1}), \end{aligned}$$

hence, conditional $Y_1 = m_1, \dots, X_{r-1} = k_{r-1}, Y_{r-1} = m_{r-1}$, variables $X_{(r)}$ and Y_r are independent.

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

$$\begin{aligned} P(Y_1 = m_1, \dots, Y_r = m_r, N \geq r) &= \sum_{k_1=1}^{\infty} \sum_{k_2=k_1+1}^{\infty} \dots \sum_{k_r=k_{r-1}+1}^{\infty} A q^{\sum_{i=1}^r m_i k_i + m_{r+1} k_r} \\ &= A \sum_{k_1=1}^{\infty} \sum_{k_2=k_1+1}^{\infty} \dots \sum_{k_r=k_{r-1}+1}^{\infty} q^{\sum_{i=1}^{r-1} m_i k_i + (n - m_1 - \dots - m_{r-1}) k_r} \\ &= A \sum_{k_1=1}^{\infty} \sum_{k_2=k_1+1}^{\infty} \dots \sum_{k_{r-1}=k_{r-2}+1}^{\infty} q^{\sum_{i=1}^{r-2} m_i k_i} \frac{q^{(n - m_1 - \dots - m_{r-2}) k_{r-1}}}{1 - q^{n - m_1 - \dots - m_{r-1}}} \\ &= \frac{A q^{n + (n - m_1) + \dots + (n - m_1 - \dots - m_{r-1})}}{(1 - q^n)(1 - q^{n - m_1}) \dots (1 - q^{n - m_1 - \dots - m_{r-1}})} \sum_{k_1=1}^{\infty} q^{n k_1} \\ &= \frac{n!}{m_1! \dots m_{r+1}!} \frac{p^{\sum_{i=1}^r m_i} q^{n + (n - m_1) + \dots + (n - m_1 - \dots - m_{r-1}) - \sum_{i=1}^r m_i}}{(1 - q^n)(1 - q^{n - m_1}) \dots (1 - q^{n - m_1 - \dots - m_{r-1}})}, \end{aligned}$$

hence,

$$P(Y_1 = m_1, \dots, Y_r = m_r, N \geq r) = \frac{n!}{m_1! \dots m_{r+1}!} \frac{p^{\sum_{i=1}^r m_i} q^{n + (n - m_1) + \dots + (n - m_1 - \dots - m_{r-1}) - \sum_{i=1}^r m_i}}{(1 - q^n)(1 - q^{n - m_1}) \dots (1 - q^{n - m_1 - \dots - m_{r-1}})},$$

where $m_1 + \dots + m_{r+1} = n$.

The other solution is as follows.

It is easy to check that

$$P(X_1 = \dots = X_l < \min \{X_{l+1}, \dots, X_{l+m}\}) = \sum_{k=1}^{\infty} (pq^{k-1})^l (q^k)^m = \left(\frac{p}{q}\right)^l \frac{q^{l+m}}{1 - q^{l+m}},$$

hence,

$$\begin{aligned} & P(Y_1 = m_1, \dots, Y_r = m_r, n \geq r) \\ &= P(Y_1 = m_1) P(Y_2 = m_2, N \geq 2 | Y_1 = m_1) \dots P(Y_r = m_r, N \geq r | Y_1 = m_1, \dots, Y_{r-1} = m_{r-1}, N \geq r - 1) \\ &= \binom{n}{m_1} \left(\frac{p}{q}\right)^{m_1} \frac{q^n}{1 - q^n} \binom{n - m_1}{m_2} \left(\frac{p}{q}\right)^{m_2} \frac{q^{n - m_1}}{1 - q^{n - m_1}} \dots \binom{n - m_1 - \dots - m_{r-1}}{m_r} \left(\frac{p}{q}\right)^{m_r} \frac{q^{n - m_1 - \dots - m_{r-1}}}{1 - q^{n - m_1 - \dots - m_{r-1}}} \\ &= \frac{n!}{m_1! \dots m_{r+1}!} \frac{p^{\sum_{i=1}^r m_i} q^{n + (n - m_1) + \dots + (n - m_1 - \dots - m_{r-1}) - \sum_{i=1}^r m_i}}{(1 - q^n)(1 - q^{n - m_1}) \dots (1 - q^{n - m_1 - \dots - m_{r-1}})}. \end{aligned}$$

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

$$P(X_i \neq X_j \text{ for } i \neq j, i, j = 1, 2, \dots, n) = P(Y_1 = 1, \dots, Y_n = 1) = n! p^n q^{\frac{n(n-1)}{2}} \prod_{i=1}^n \frac{1}{1 - q^i}.$$

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

$$\begin{aligned} & P(X_{(1)} = k_1, \dots, X_{(r)} = k_r | Y_1 = m_1, \dots, Y_r = m_r, N \geq r) \\ &= \frac{P(X_{(1)} = k_1, \dots, X_{(r)} = k_r, Y_1 = m_1, \dots, Y_r = m_r, N \geq r)}{P(Y_1 = m_1, \dots, Y_r = m_r, N \geq r)} \\ &= (1 - q^n)(1 - q^{n - m_1}) \dots (1 - q^{n - m_1 - \dots - m_{r-1}}) q^{\sum_{i=1}^r m_i k_i + (n - m_1 - \dots - m_r) k_r - [n + (n - m_1) + \dots + (n - m_1 - \dots - m_{r-1})]}, \end{aligned}$$

where $m_1 + \dots + m_r \leq n$ and $1 \leq k_1 < \dots < k_r$.

Therefore,

$$\begin{aligned}
 & P(D_{(1)} = y_1, \dots, D_{(r)} = y_r | Y_1 = m_1, \dots, Y_r = m_r, N \geq r) \\
 = & \frac{P(X_{(1)} = y_1, X_{(2)} = y_1 + y_2, \dots, X_{(r)} = y_1 + \dots + y_r, Y_1 = m_1, \dots, Y_r = m_r, N \geq r)}{P(Y_1 = m_1, \dots, Y_r = m_r, N \geq r)} \\
 = & (1 - q^n)(1 - q^{n-m_1}) \dots (1 - q^{n-m_1-\dots-m_{r-1}}) \\
 & \times q^{[m_1 y_1 + m_2(y_1 + y_2) + \dots + m_r(y_1 + \dots + y_r) + (n-m_1-\dots-m_r)(y_1 + \dots + y_r)] - [n + (n-m_1) + \dots + (n-m_1-\dots-m_{r-1})]} \\
 = & (1 - q^n)(1 - q^{n-m_1}) \dots (1 - q^{n-m_1-\dots-m_{r-1}}) \\
 & \times q^{[n y_1 + (n-m_1) y_2 + \dots + (n-m_1-\dots-m_{r-1}) y_r] - [n + (n-m_1) + \dots + (n-m_1-\dots-m_{r-1})]} \\
 = & [(1 - q^n)(q^n)^{y_1-1}] [(1 - q^{n-m_1})(q^{n-m_1})^{y_2-1}] \dots [(1 - q^{n-m_1-\dots-m_{r-1}})(q^{n-m_1-\dots-m_{r-1}})^{y_r-1}],
 \end{aligned}$$

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

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

$$D_i \sim Geo(1 - q^{n-m_1-\dots-m_{i-1}}), i = 1, 2, \dots, r.$$

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

$$D_i \sim Geo(1 - q^{n-i+1}), i = 1, 2, \dots, n.$$

The proof is complete.

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

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

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

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

Moreover,

$$\begin{aligned}
 \sum_{i=k}^n D_{(n+1-i)} &= \sum_{i=k}^n (X_{(n+1-i)} - X_{(n-i)}) \\
 &= X_{(n+1-k)}.
 \end{aligned}$$

From the above, it follows that

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

The proof is complete. ■

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

Proof: it is easy to check that

$$X_{(n-l)} = \sum_{i=l+1}^n Y_i \quad \text{and} \quad X_{(n-k)} = \sum_{i=k+1}^n Y_i,$$

hence,

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

Since

$$Z_{(k-l)} \text{ and } \sum_{i=l+1}^k Y_i \text{ have an identical distribution,}$$

therefore,

$$X_{(n-l)} - X_{(n-k)} \text{ and } Z_{(k-l)} \text{ are identically distributed.}$$

The proof is complete. ■

Corollary 2: X_n given $X_i \neq X_j (i \neq j)$ and $\sum_{i=l+1}^k Y_i$ have an identical distribution.

Corollary 3: conditional $X_i \neq X_j (i \neq j)$, variables $X_{(n)} - X_{(1)}$ and $\max (X_1, \dots, X_n)$ have an identical distribution.

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

- [1]. T. S. Ferguson. A characterization of the geometric distribution. *American Mathematical Monthly*, 72(3):256–260, 1965.
- [2]. Thomas S. Ferguson. On characterizing distributions by properties of order statistics. *Sankhy: The Indian Journal of Statistics, Series A* (1961–2002), and 29(3):265–278, 1967.
- [3]. E. B. Fosam and D. N. Shanbhag. Certain characterizations of exponential and geometric distributions. *Journal of the Royal Statistical Society*, 56(1):157–160, 1994.
- [4]. Fernando López Blázquez and Begoña Salamanca Miño. A characterization of the geometric distribution. *Statistical Papers*, 39(2):231–236, 1998.
- [5]. A. C Dallas. A characterization of the geometric distribution. *Journal of Applied Probability*, 11(3):609–611, 1974.
- [6]. Henrik Cobbers and Udo Kamps. A note on characterizations of the geometric distribution. *Journal of Statistical Planning & Inference*, 67(2):187–190, 1998.
- [7]. Gaoxiong Gan and Lee J. Bain. Some results for type I censored sampling from geometric distributions. *Journal of Statistical Planning & Inference*, 67(1):85–97, 1998.
- [8]. A. Y. Yehia and A. N. Ahmed. Two characterizations of geometric distributions. *Microelectronics Reliability*, 33(6):833–843, 1993.
- [9]. R Dodunekova. Characterization of the geometric and exponential random variables. *Communications in Statistics*, 33(8):1755–1765, 2004.
- [10]. M Doostparast and J Ahmadi. Statistical analysis for geometric distribution based on records. *Computers & Mathematics with Applications*, 52(6):905–916, 2006.
- [11]. E. Gmezdniz. Another generalization of the geometric distribution. *Test*, 19(2):399–415, 2010.
- [12]. Subrata Chakraborty and Rameshwar D. Gupta. Exponentiated geometric distribution: Another generalization of geometric distribution. *Communications in Statistics*, 44(6):1143–1157, 2015.
- [13]. Nan Cheng Su and Wan Ping Hung. Characterizations of the geometric distribution via residual lifetime. *Statistical Papers*, 59(1):57C73, 2018.