

THE MULTINOMIAL GEOMETRIC DISTRIBUTION

LI Guang-hui¹, LI Jun-peng², ZHANG Chong-qi²

(1. School of Science, Kaili University, Kaili 556011, China)

(2. School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China)

Abstract: In this paper, we studied the collection problem under Bernoulli's trials and derived a probability function of the multinomial geometric distribution by using the theory of mixture lattice point sets. Further, we proposed a uniform multinomial geometric distribution when the probabilities of each of the trial outcomes are assumed to be equal in Bernoulli's trial. Moreover, we obtained the probability functions, the expectation, and the variance of the two types of distributions, and verified the differences between these two types of distributions and the normal distribution by means of simulations. Finally, we developed a polynomial regression model on the probability and the number of trials from the simulation results, which can be used to simplify the calculation effectively.

Keywords: multinomial distribution; geometric distribution; probability function

2010 MR Subject Classification: 62E20; 62G30

Document code: A **Article ID:** 0255-7797(2022)02-0129-17

1 Introduction

Geometric distribution is an important kind of probability distribution, which has been widely used in practice. According to the actual situation, geometric distribution has many important generalizations. For example, Muwafi(1980) studied the k order geometric distribution by Fibonacci sequence, Philippou, et al(1980) further deduced the properties and characteristic functions of this distribution. Miller(2008) introduced the properties and applications of geometric distribution.

In the past decade, geometric distribution has been widely used. Jayakumar(2018), Ahmed, et al (2014) introduced the Weibull geometric distribution . The beta geometric distribution studied by Kemp and Adrienne (2001), and some other applications can be seen in Porwal (2018), Pedro, et al (2014)among others.

In many cases, we need to consider an important kind of problem, which we call "collection problem". The problem is described as: a Bernoulli experiment, one outcome with m different results A_1, A_2, \dots, A_m may observed in a trial. The probability of these m results is $p_i = P(A_i), i = 1, 2, \dots, m$, respectively. There are two questions:

* **Received date:** 2021-03-15

Accepted date: 2021-06-01

Foundation item: Supported by Science and Technology Fund for Basic Research of Guizhou Province ([2020]1Y010); National Nature Sciences Foundation of China (NSFC 11901260).

Biography: Li Guanghui(1985-), male, born at Guiyang, Guizhou, associate professor, major in statistics, E-mail: liguanghui1985@126.com

- How many trials are needed until all the m results appear?
- The total number of the trials Y is subject to what distribution?

In the above problems, the experiment will continue until all of A_1, A_2, \dots, A_m or the r ($r < m$) results appeared, we are concerned about the probability distribution of Y (the total number of the trials). Xiao, et al(2015) studied several special cases of this problem. We further study the characteristics and properties of this kind of distribution basis on these results. For the one hand, the distribution of Y is constructed by using the theory of mixture lattice point set. For the other hand, using the mixexp package in R language to carry out complex calculation in the case of high dimension. The probability distribution, expectation and variance of Y are calculated and proved strictly.

In this paper, we first give the basic notations and their properties. In Section 3, we give the probability distribution and the numerical characteristics of the random variable Y which subject to Multinomial Geometric distribution (MGe). In Section 4, we mainly introduce a special case, the Uniform Multinomial Geometric distribution(UMGe). We also discuss the properties and asymptotic distribution of this kind of distribution. In Section 5, the asymptotic properties and estimation methods of MGe and UMGe are discussed by data simulation. Finally, we put forward some problems that can be further studied.

2 Notations and Preliminaries

We use the notation ,defined in Li and Zhang (2017). For the positive integer $n \in \mathbb{Z}^+$, denote

$$\mathcal{L}\{m, n\} = \left\{ \mathbf{n}_m = (n_1, n_2, \dots, n_m) : \sum_{i=1}^m n_i = n, n_i \geq 0, n_i \in \mathbb{Z}^+, i = 1, 2, \dots, m \right\}$$

as the m components, n order lattice points set on a simplex. It can be decomposed into the combination of interior point set and boundary point set $\mathcal{L}\{m, n\} = \mathcal{N}_m(n) \cup \mathcal{N}_m^0(n)$, where

$$\mathcal{N}_m(n) = \{ \mathbf{n}_m = (n_1, n_2, \dots, n_m) : \mathbf{n}_m \in \mathcal{L}\{m, n\}, n_i > 0, i = 1, 2, \dots, m \} \quad (2.1)$$

is the interior point set of $\mathcal{L}\{m, n\}$, $\mathcal{N}_m^0(n) = \mathcal{L}\{m, n\} \setminus \mathcal{N}_m(n)$ is the boundary point set of $\mathcal{L}\{m, n\}$.

As (2.1) shows, $\mathcal{N}_m(n) \neq \emptyset$ if and only if $n \geq m$, $\mathcal{L}\{m, n\}$ has $\binom{n+m-1}{m-1}$ elements, $\mathcal{N}_m(n)$ has $\binom{n-1}{m-1}$ elements.

We define the index set as

$$\mathcal{I}_m(j) = \{ \mathbf{i}_j = (i_1, i_2, \dots, i_j) : 1 \leq i_1 < i_2 < \dots < i_j \leq m, i_k \in \mathbb{Z}^+, k = 1, 2, \dots, j \}. \quad (2.2)$$

If $\mathcal{I}_m(j)$ does not include an index $i \in \{1, 2, \dots, m\}$, we denote it by

$$\mathcal{I}_m(j \setminus i) = \{ \mathbf{i}_j = (i_1, i_2, \dots, i_j) : \mathbf{i}_j \in \mathcal{I}_m(j), i_k \neq i, k = 1, 2, \dots, j \}.$$

In this paper, we always use the form notation $\mathbf{i}_j = (i_1, i_2, \dots, i_j)$ to indicate the j dimensional vector, $\mathbf{n}_{m \setminus i} = (n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_m)$ to indicate the $m - 1$ dimensional vector obtained by removing the i -th element from \mathbf{n}_m .

Let $\mathbf{p}_m = (p_1, p_2, \dots, p_m)$ be a nonnegative vector, denote

$$S^n(\mathbf{p}_m, j) = \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{i=1}^m p_i - \sum_{k=1}^j p_{i_k} \right)^n, j = 0, 1, \dots, m - 1. \tag{2.3}$$

For example, $S^n(\mathbf{p}_m, 0) = \left(\sum_{i=1}^m p_i \right)^n, S^n(\mathbf{p}_m, 1) = \sum_{k=1}^m \left(\sum_{i=1}^m p_i - p_k \right)^n, \dots, S^n(\mathbf{p}_m, m - 1) = \sum_{k=1}^m p_k^n$. The form of expansion of $S^n(\mathbf{p}_m, j)$ is as follows

$$S^n(\mathbf{p}_m, j) = n! \sum_{k=1}^{m-j} \binom{m-k}{j} \left\{ \sum_{\mathbf{i}_k \in \mathcal{I}_m(k)} \sum_{\mathbf{n}_k \in \mathcal{N}_k(n)} \prod_{l=1}^k \frac{p_{i_l}^{n_l}}{n_l!} \right\}.$$

When $1 \leq n \leq m - 1$, we have

$$\sum_{j=0}^{m-1} (-1)^j S^n(\mathbf{p}_m, j) = n! \sum_{j=0}^{m-n} \left\{ \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \sum_{\mathbf{n}_j \in \mathcal{N}_j(m)} \prod_{k=1}^j \frac{p_{i_k}^{n_k}}{n_k!} \left[\sum_{l=0}^{m-j} \binom{m-j}{l} (-1)^l \right] \right\} = 0.$$

When $n \geq m$, we have

$$\sum_{j=0}^{m-1} (-1)^j S^n(\mathbf{p}_m, j) = n! \sum_{\mathbf{n}_m \in \mathcal{N}_m(n)} \prod_{k=1}^m \frac{p_k^{n_k}}{n_k!}.$$

Property 2.1 For any nonnegative vector $\mathbf{p}_m = (p_1, p_2, \dots, p_m)$ and integer $n \geq 1$,

$$b^n(\mathbf{p}_m) = \sum_{j=0}^{m-1} (-1)^j S^n(\mathbf{p}_m, j) = I(n \geq m) \left(n! \sum_{\mathbf{n}_m \in \mathcal{N}_m(n)} \prod_{i=1}^m \frac{p_i^{n_i}}{n_i!} \right),$$

where $I(\cdot)$ is indicator function.

To prove our conclusions, there three properties are proposed as follows.

Property 2.2 The sequence i_1, i_2, \dots, i_{j+1} is an arbitrary permutation of $1, 2, \dots, j+1$, if the function $g(\cdot)$ satisfies $g(p_{i_1}, p_{i_2}, \dots, p_{i_{j+1}}) = g(p_1, p_2, \dots, p_{j+1})$, then we called the $g(p_1, p_2, \dots, p_j)$ is symmetrical, and we have

$$\sum_{i=1}^m p_i \sum_{\mathbf{i}_j \in \mathcal{I}_m(j \setminus i)} \frac{g(p_i, p_{i_1}, \dots, p_{i_j})}{(p_i + p_{i_1} + \dots + p_{i_j})^{r+1}} = \sum_{\mathbf{i}_{j+1} \in \mathcal{I}_m(j+1)} \frac{g(p_{i_1}, \dots, p_{i_j}, p_{i_{j+1}})}{(p_{i_1} + \dots + p_{i_j} + p_{i_{j+1}})^r}.$$

Property 2.3 When $n \geq m, (p_1 + p_2 + \dots + p_m)^n = b^n(\mathbf{p}_m) + n! \sum_{\mathbf{n}_m \in \mathcal{N}_m^0(n)} \prod_{k=1}^m \frac{p_k^{n_k}}{n_k!} > 0$,

so we have

$$b^n(\mathbf{p}_m) < (p_1 + p_2 + \dots + p_m)^n. \tag{2.4}$$

Remark In the above discussion, only required \mathbf{p}_m to be a nonnegative vector. In the following, we use \mathbf{p}_m to represent a probability vector, that is, to satisfy $\sum_{i=1}^m p_i = 1, p_i > 0, i = 1, 2, \dots, m$.

Now, we discuss the case that \mathbf{p}_m is a parameter in a multinomial distribution.

Definition 2.1 Random variable X subject to geometric distribution, denote as $X \sim Ge(p), 0 < p < 1$, the probability function of X is $P(X = n) = (1 - p)^{n-1}p, n = 1, 2, \dots$. Random variable vector (X_1, X_2, \dots, X_m) subject to multiple distribution, denote as $(X_1, X_2, \dots, X_m) \sim M(n, \mathbf{p}_m)$, the joint probability function is

$$P(X_1 = n_1, X_2 = n_2, \dots, X_m = n_m) = n! \prod_{k=1}^m \frac{p_k^{n_k}}{n_k!},$$

where $\sum_{i=1}^m p_i = 1, \sum_{k=1}^m n_k = n$.

Following the notations and assumption of Definition 2.1 and Property 2.1, for $1 \leq n \leq m - 1$, we know that the following equation holds.

$$S^n(\mathbf{p}_m, 0) - \sum_{j=1}^{m-1} (-1)^{j-1} S^n(\mathbf{p}_m, j) = 0,$$

and $S^n(\mathbf{p}_m, 0) = \left(\sum_{i=1}^m p_i \right)^n = 1$, then we get the following property.

Property 2.4 When $1 \leq n \leq m - 1$, we have

$$\begin{aligned} \sum_{j=1}^{m-1} (-1)^{j-1} S^n(\mathbf{p}_m, j) &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} (-1)^{j-1} \left(1 - \sum_{k=1}^j p_{i_k} \right)^n = 1, \\ \sum_{j=1}^{m-1} (-1)^{j-1} S^n(\mathbf{p}_m, m-j) &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} (-1)^{j-1} \left(\sum_{k=1}^j p_{i_k} \right)^n = (-1)^m. \end{aligned}$$

3 Multinomial Geometric Distribution and Its Properties

Based on the Definition 2.1 and the notations given in Section 2, we discuss the probability distribution function of multiple geometric distribution (MGe) and its properties.

Theorem 3.1 Suppose that A_1, A_2, \dots, A_m are the m different results in each Bernoulli experiment, $p_i = P(A_i) > 0, i = 1, 2, \dots, m$. When all m results appear, the total number of the trials Y is subject to MGe distribution, denote it as $Y \sim MGe(\mathbf{p}_m)$, the probability distribution function is

$$P(Y = n) = \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right) \left(1 - \sum_{k=1}^j p_{i_k} \right)^{n-1}, n = m, m+1, \dots \quad (3.1)$$

Proof Let X_1, X_2, \dots, X_m are the appeared times of A_1, A_2, \dots, A_m in first $n - 1$ trials, so $(X_1, X_2, \dots, X_m) \sim M(n - 1, \mathbf{p}_m)$. According to the total probability formula, there are

$$\begin{aligned}
P(Y = n) &= \sum_{i=1}^m P(A_i)P(X_1 \geq 1, \dots, X_{i-1} \geq 1, X_i = 0, X_{i+1} \geq 1, \dots, X_m \geq 1) \\
&= \sum_{i=1}^m p_i \left\{ (n-1)! \sum_{\mathbf{n}_{m-1} \in \mathcal{N}_{m-1}(n-1)} \prod_{k \neq i} \binom{p_k^{n_k}}{n_k!} \right\} \\
&= \sum_{i=1}^m p_i b^{n-1}(\mathbf{p}_{m \setminus i}) \\
&= \sum_{i=1}^m p_i (1 - p_i)^{n-1} - \sum_{i=1}^m p_i \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} (1 - p_i - p_{i_1})^{n-1} \\
&\quad + \sum_{i=1}^m p_i \sum_{\mathbf{i}_2 \in \mathcal{I}_m(2 \setminus i)} (1 - p_i - p_{i_1} - p_{i_2})^{n-1} + \dots + (-1)^{m-2} \sum_{i=1}^m p_i \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} p_{i_1}^{n-1} \\
&= \sum_{i=1}^m p_i (1 - p_i)^{n-1} - \sum_{\mathbf{i}_2 \in \mathcal{I}_m(2)} (p_{i_1} + p_{i_2}) (1 - p_{i_1} - p_{i_2})^{n-1} \\
&\quad + \sum_{\mathbf{i}_3 \in \mathcal{I}_m(3)} (p_{i_1} + p_{i_2} + p_{i_3}) (1 - p_{i_1} - p_{i_2} - p_{i_3})^{n-1} \\
&\quad + \dots + (-1)^{m-2} \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1)} (1 - p_{i_1}) p_{i_1}^{n-1} \quad (\text{Property 2.2}) \\
&= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right) \left(1 - \sum_{k=1}^j p_{i_k} \right)^{n-1}.
\end{aligned}$$

Next, we verify its probability regularity.

$$\begin{aligned}
\sum_{n=m}^{\infty} P(Y = n) &= \sum_{n=m}^{\infty} \sum_{i=1}^m p_i b^{n-1}(\mathbf{p}_{m \setminus i}) \\
&= \sum_{i=1}^m p_i \sum_{n=m}^{\infty} \left\{ (1 - p_i)^{n-1} - \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} (1 - p_i - p_{i_1})^{n-1} + \sum_{\mathbf{i}_2 \in \mathcal{I}_m(2 \setminus i)} (1 - p_i - p_{i_1} - p_{i_2})^{n-1} \right. \\
&\quad \left. + \dots + (-1)^{m-2} \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} p_{i_1}^{n-1} \right\} \\
&= \sum_{i=1}^m p_i \left\{ \frac{(1 - p_i)^{m-1}}{p_i} - \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} \frac{(1 - p_i - p_{i_1})^{m-1}}{(p_i + p_{i_1})} + \sum_{\mathbf{i}_2 \in \mathcal{I}_m(2 \setminus i)} \frac{(1 - p_i - p_{i_1} - p_{i_2})^{m-1}}{(p_i + p_{i_1} + p_{i_2})} \right. \\
&\quad \left. + \dots + (-1)^{m-2} \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} \frac{p_{i_1}^{m-1}}{(1 - p_{i_1})} \right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^m (1-p_i)^{m-1} - \sum_{\mathbf{i}_2 \in \mathcal{I}_m(2)} (1-p_{i_1}-p_{i_2})^{m-1} + \sum_{\mathbf{i}_3 \in \mathcal{I}_m(3)} (1-p_{i_1}-p_{i_2}-p_{i_3})^{m-1} \\
&\quad + \cdots + (-1)^{m-2} \sum_{i=1}^m p_i^{m-1} \\
&= \sum_{j=1}^{m-1} (-1)^{j-1} S^{m-1}(\mathbf{p}_m, j) \quad (\text{Property 2.4}) \\
&= 1.
\end{aligned}$$

This completes the proof.

If the stopping condition of the experiment is changed, we give the following two generalizations.

Corollary 3.1 Stopping the experiment until the specified r results A_1, A_2, \dots, A_r appear, the total number of trails is V_r , denote it as $V_r \sim MGe(p_1, p_2, \dots, p_r)$. The probability distribution function of V_r is

$$P(V_r = n) = \sum_{j=1}^r (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_r(j)} \left(\sum_{k=1}^j p_{i_k} \right) \left(1 - \sum_{k=1}^j p_{i_k} \right)^{n-1}, \quad n = r, r+1, \dots.$$

Corollary 3.2 Stopping the experiment until the arbitrary r results appear, the total number of trails is W_r , denote it as $W_r \sim MGe(\mathbf{p}_m, r)$. The probability distribution function of W_r is

$$P(W_r = n) = \sum_{\mathbf{k}_r \in \mathcal{I}_m(r)} \sum_{j=1}^{r-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_r(j)} \left(\sum_{l=1}^j p_{k_{i_l}} \right) \left(p_{0k_i} - \sum_{l=1}^j p_{k_{i_l}} \right)^{n-1}, \quad n = r, r+1, \dots,$$

where $p_{0k_i} = \sum_{l=1}^r p_{k_{i_l}}$.

According to the Property 2.2, the probability distribution function of $Y \sim MGe(\mathbf{p}_m)$ has upper bound. Let

$$u_m(n) = \begin{cases} \sum_{i=1}^m p_i (1-p_i)^{n-1} & , m \text{ is even,} \\ \sum_{i=1}^m (p_i (1-p_i)^{n-1} - (1-p_i) p_i^{n-1}) & , m \text{ is odd,} \end{cases} \quad (3.2)$$

then, we have $P(Y = n) < u_n$ hold, and $P(Y = n) = u_n + o(p^{-n}), 0 < p < 1$.

Theorem 3.2 Suppose that $Y \sim MGe(\mathbf{p}_m)$, the expectation and second order moment of Y are as follow:

$$E(Y) = \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right)^{-1} + (-1)^{m-1}, \quad (3.3)$$

$$E(Y^2) = \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right)^{-2} \left(2 - \sum_{k=1}^j p_{i_k} \right) + (-1)^{m-1}. \quad (3.4)$$

Proof Let us first calculate the expectation

$$\begin{aligned} E(Y) &= \sum_{n=m}^{\infty} nP(Y=n) = \sum_{n=m}^{\infty} \sum_{i=1}^m np_i b^{n-1} (\mathbf{p}_{m \setminus i}) \\ &= \sum_{i=1}^m p_i \sum_{n=m}^{\infty} \left\{ n(1-p_i)^{n-1} - \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} n(1-p_i-p_{i_1})^{n-1} + \sum_{\mathbf{i}_2 \in \mathcal{I}_m(2 \setminus i)} n(1-p_i-p_{i_1}-p_{i_2})^{n-1} \right. \\ &\quad \left. + \cdots + (-1)^{m-2} \sum_{\mathbf{i}_1 \in \mathcal{I}_m(1 \setminus i)} np_{i_1}^{n-1} \right\} \\ &= \sum_{i=1}^m p_i \left\{ \sum_{j=0}^{m-2} (-1)^j \sum_{\mathbf{i}_j \in \mathcal{I}_m(j \setminus i)} \left(p_i + \sum_{k=1}^j p_{i_k} \right)^{-2} \left(1 - p_i - \sum_{k=1}^j p_{i_k} \right)^{m-1} \left[(m-1) \sum_{k=1}^j p_{i_k} + 1 \right] \right\} \\ &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(1 - \sum_{k=1}^j p_{i_k} \right)^{m-1} \left[(m-1) + \left(\sum_{k=1}^j p_{i_k} \right)^{-1} \right] \\ &= (m-1) \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(1 - \sum_{k=1}^j p_{i_k} \right)^{m-1} + \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(1 - \sum_{k=1}^j p_{i_k} \right)^{m-1} \left(\sum_{k=1}^j p_{i_k} \right)^{-1} \\ &= (m-1) + \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right)^{-1} + \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \sum_{l=1}^{m-1} (-1)^l \binom{m-1}{l} \left(\sum_{k=1}^j p_{i_k} \right)^{l-1}. \end{aligned}$$

Let the third item in the above formula be

$$\begin{aligned} Q &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \sum_{l=1}^{m-1} (-1)^l \binom{m-1}{l} \left(\sum_{k=1}^j p_{i_k} \right)^{l-1} \\ &= \binom{m-1}{1} \sum_{j=1}^{m-1} (-1)^j \binom{m}{j} + \sum_{l=2}^{m-1} (-1)^l \binom{m-1}{l} \sum_{j=1}^{m-1} (-1)^{j-1} S^{l-1}(\mathbf{p}_m, m-j) \\ &= (m-1)(-1 - (-1)^m) + (-1)^m \sum_{l=2}^{m-1} (-1)^l \binom{m-1}{l} \\ &= -(m-1) + (-1)^{m-1}. \end{aligned}$$

We can get the conclusion (3.3) by substituting Q into

$$E(Y) = (m-1) + \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right)^{-1} + Q.$$

Next, we calculate the second moment. Let

$$c^{-h}(\mathbf{p}_m) = \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right)^{-h}, \quad h = 1, 2.$$

It can be seen from the above expected calculation process, there is conclusion as

$$\sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(1 - \sum_{k=1}^j p_{i_k}\right)^{m-1} \left(\sum_{k=1}^j p_{i_k}\right)^{-1} = c^{-1}(\mathbf{p}_m) - (m-1) + (-1)^{m-1}.$$

The second moment can be obtained by the following calculation.

$$\begin{aligned} E(Y^2) &= \sum_{n=m}^{\infty} n(n+1)P(Y=n) - E(Y) \\ &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left\{ m(m+1) \left(1 - \sum_{k=1}^j p_{i_k}\right)^{m-1} \right. \\ &\quad \left. + 2 \left(1 - \sum_{k=1}^j p_{i_k}\right)^m \left(\sum_{k=1}^j p_{i_k}\right)^{-2} \left(1 + m \sum_{k=1}^j p_{i_k}\right) \right\} \\ &\quad - \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(1 - \sum_{k=1}^j p_{i_k}\right)^{m-1} \left[(m-1) + \left(\sum_{k=1}^j p_{i_k}\right)^{-1} \right] \\ &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left\{ \left(1 - \sum_{k=1}^j p_{i_k}\right)^{m-1} (m-1)^2 + (2m-3) \left(\sum_{k=1}^j p_{i_k}\right)^{-1} + 2 \left(\sum_{k=1}^j p_{i_k}\right)^{-2} \right\} \\ &= (m-1)^2 \sum_{j=1}^{m-1} (-1)^{j-1} S^{m-1}(\mathbf{p}_m, j) \\ &\quad + (2m-3) \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(1 - \sum_{k=1}^j p_{i_k}\right)^{m-1} \left(\sum_{k=1}^j p_{i_k}\right)^{-1} \\ &\quad + 2 \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \sum_{l=0}^{m-1} (-1)^l \binom{m-1}{l} \left(\sum_{k=1}^j p_{i_k}\right)^{l-2} \\ &= (m-1)^2 + (2m-3) [c^{-1}(\mathbf{p}_m) - (m-1) + (-1)^m] \\ &\quad + 2 \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\left(\sum_{k=1}^j p_{i_k}\right)^{-2} - \binom{m-1}{1} \left(\sum_{k=1}^j p_{i_k}\right)^{-1} + \binom{m-1}{2} \right) \\ &\quad + 2 \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \sum_{l=3}^{m-1} (-1)^l \binom{m-1}{l} \left(\sum_{k=1}^j p_{i_k}\right)^{l-2} \\ &= -(m-1)(m-2) + (2m-3)c^{-1}(\mathbf{p}_m) + (2m-3)(-1)^m \\ &\quad + 2c^{-2}(\mathbf{p}_m) - 2(m-1)c^{-1}(\mathbf{p}_m) + \binom{m-1}{2} \sum_{j=1}^{m-1} (-1)^{j-1} \binom{m}{j} \\ &\quad + 2 \sum_{l=3}^{m-1} (-1)^l \binom{m-1}{l} \sum_{j=1}^{m-1} (-1)^{j-1} S^{l-2}(\mathbf{p}_m, m-j) \end{aligned}$$

$$\begin{aligned}
&= 2c^{-2}(\mathbf{p}_m) - c^{-1}(\mathbf{p}_m) + 2(-1)^m \left\{ \sum_{l=2}^{m-1} (-1)^l \binom{m-1}{l} \right\} - (2m-3)(-1)^m \\
&= 2c^{-2}(\mathbf{p}_m) - c^{-1}(\mathbf{p}_m) + (-1)^{m-1}.
\end{aligned}$$

For example, suppose $Y^{(m)} \sim MGe(\mathbf{p}_m)$, $m = 2, 3, 4$, there are some results:

$$P(Y^{(2)} = n) = p^{n-1}(1-p) + (1-p)^{n-1}p, n = 2, 3, \dots;$$

$$P(Y^{(3)} = n) = \sum_{i=1}^3 (p_i(1-p_i)^{n-1} - (1-p_i)p_i^{n-1}), n = 3, 4, \dots;$$

$$P(Y^{(4)} = n) = \sum_{i=1}^4 [p_i(1-p_i)^{n-1} + p_i^{n-1}(1-p_i)] - \sum_{\mathbf{i}_2 \in \mathcal{I}_2(4)} (p_{i_1} + p_{i_2})(1-p_{i_1} - p_{i_2})^{n-1}, n = 4, 5, \dots;$$

$$E(Y^{(2)}) = \frac{1}{p(1-p)} - 1.$$

$$E(Y^{(3)}) = 1 + \sum_{i=1}^3 \left(\frac{1}{p_i} - \frac{1}{1-p_i} \right).$$

$$E(Y^{(4)}) = \sum_{i=1}^4 \left(\frac{1}{p_i} + \frac{1}{1-p_i} \right) - \sum_{\mathbf{i}_2 \in \mathcal{I}_2(4)} \frac{1}{p_{i_1} + p_{i_2}} - 1.$$

4 Uniform Multiple Geometric Distribution

In this section, we mainly discuss a special case of MGe. When the parameter vector $\mathbf{p}_m = (p_1, p_2, \dots, p_m)$, $p_i = \frac{1}{m}$, $i = 1, 2, \dots, m$, we denote $MGe(\mathbf{p}_m)$ as $MGe(m)$, named it as Uniform Multiple Geometric distribution(UMGe).

Theorem 4.1 Suppose Y subject to UMGe distribution, we denote it as $Y \sim MGe(m)$, then the probability distribution function of Y is

$$P(Y = n) = \sum_{j=1}^{m-1} (-1)^{j-1} \binom{m-1}{j-1} \left(1 - \frac{j}{m}\right)^{n-1}, n = m, m+1, \dots \quad (4.1)$$

Proof From the condition $p_i = \frac{1}{m}$, $i = 1, 2, \dots, m$, we can get

$$\sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} (-1)^{j-1} \left(\sum_{k=1}^j p_{i_k} \right) \left(1 - \sum_{k=1}^j p_{i_k} \right)^{n-1} = (-1)^{j-1} \binom{m}{j} \left(\frac{j}{m} \right) \left(1 - \frac{j}{m} \right)^{n-1}.$$

According to equation (3.1), we have

$$P(Y = n) = \sum_{j=1}^{m-1} (-1)^{j-1} \binom{m-1}{j-1} \left(1 - \frac{j}{m}\right)^{n-1}, n = m, m+1, \dots.$$

To obtain the expectation and variance of the UMGe distribution, we first give the following three combined formulas.

Property 4.1 The following three combination formulas are always held for arbitrary positive integers $m \in \mathbb{Z}^+$.

$$r_1(m) = \sum_{j=1}^m (-1)^{j-1} \frac{1}{j} \binom{m}{j} = \sum_{j=1}^m \frac{1}{j}. \quad (4.2)$$

$$r_2(m) = \sum_{j=0}^m (-1)^j \frac{1}{(j+1)^2} \binom{m}{j} = \frac{1}{m+1} \sum_{j=1}^{m+1} \frac{1}{j}. \quad (4.3)$$

$$\sum_{j=1}^m (-1)^{j-1} \binom{m}{j} \left(\frac{m}{j}\right)^2 = \frac{m^2}{2} \left\{ \left(\sum_{j=1}^m \frac{1}{j}\right)^2 + \sum_{j=1}^m \frac{1}{j^2} \right\}. \quad (4.4)$$

Proof For the formula (4.2), because $r_1(m-1) = \sum_{j=1}^{m-1} (-1)^{j-1} \frac{1}{j} \binom{m-1}{j}$, where

$$\begin{aligned} r_1(m) &= \sum_{j=1}^{m-1} (-1)^{j-1} \frac{1}{j} \binom{m}{j} + (-1)^{m-1} \frac{1}{m} \\ &= \sum_{j=1}^{m-1} (-1)^{j-1} \frac{1}{j} \binom{m-1}{j} + \sum_{j=1}^{m-1} (-1)^{j-1} \frac{1}{j} \binom{m-1}{j-1} + (-1)^{m-1} \frac{1}{m} \\ &= r_1(m-1) + \sum_{j=1}^{m-1} (-1)^{j-1} \frac{1}{j} \frac{j}{m} \binom{m}{j} + (-1)^{m-1} \frac{1}{m} \\ &= r_1(m-1) + \frac{1}{m} \left\{ \left[-1 + \sum_{j=1}^m (-1)^{j-1} \binom{m}{j} \right] + 1 \right\} \\ &= r_1(m-1) + \frac{1}{m}. \end{aligned}$$

So we have $r_1(m) = r_1(1) + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m} = \sum_{j=1}^m \frac{1}{j}$.

For the formula (4.3), let $R_2(x) = \sum_{j=0}^m (-1)^j \frac{1}{(j+1)^2} \binom{m}{j} e^{(j+1)x}$, so the second derivative of $R_2(x)$ is

$$R_2''(x) = \sum_{j=0}^m (-1)^j \binom{m}{j} e^{(j+1)x} = e^x (1 - e^x)^m.$$

$$R_2(x) = \int \left(\int e^x (1 - e^x)^m dx \right) dx = -\frac{1}{m+1} \left\{ x + \sum_{j=1}^{m+1} (-1)^j \frac{1}{j} \binom{m+1}{j} e^x \right\} + C_1 x + C_2$$

where C_1, C_2 is the undetermined constant, then from $R_2(0) = r_2(m), r_2(1) = 1$, and equation (4.3), we have

$$r_2(m) = \frac{1}{m+1} \sum_{j=1}^{m+1} \frac{1}{j}.$$

Let us prove the equation (4.4) by mathematical induction. Let

$$f_1(m) = \sum_{j=1}^m (-1)^{j-1} \binom{m}{j} \left(\frac{m}{j}\right)^2, \quad f_2(m) = \frac{m^2}{2} \left\{ \left(\sum_{j=1}^m \frac{1}{j}\right)^2 + \sum_{j=1}^m \frac{1}{j^2} \right\}.$$

When $m = 1$, $f_1(1) = f_2(1) = 1$, the equation (4.4) held.

If for a given $m \in \mathbb{Z}^+$, $f_1(m) = f_2(m)$ holds, then we prove that the conclusion is also true with $m + 1$.

$$\begin{aligned} f_1(m+1) &= \sum_{j=1}^{m+1} (-1)^{j-1} \binom{m+1}{j} \left(\frac{m+1}{j}\right)^2 \\ &= \sum_{j=1}^m (-1)^{j-1} \binom{m}{j} \left(\frac{m+1}{j}\right)^2 + \sum_{j=1}^m (-1)^{j-1} \binom{m}{j-1} \left(\frac{m+1}{j}\right)^2 + (-1)^m \\ &= \frac{(m+1)^2}{m^2} f_1(m) + (m+1)^2 \left\{ \frac{1}{m+1} \sum_{j=1}^{m+1} \frac{1}{j} - \frac{(-1)^m}{(m+1)^2} \right\} + (-1)^m \\ &= \frac{(m+1)^2}{2} \left\{ \left(\sum_{j=1}^m \frac{1}{j}\right)^2 + \sum_{j=1}^m \frac{1}{j^2} \right\} + (m+1)^2 \left\{ \frac{1}{m+1} \sum_{j=1}^m \frac{1}{j} + \frac{1}{(m+1)^2} \right\} \\ &= \frac{(m+1)^2}{2} \left\{ \left(\sum_{j=1}^m \frac{1}{j}\right)^2 + \frac{2}{m+1} \sum_{j=1}^m \frac{1}{j} + \frac{1}{(m+1)^2} + \sum_{j=1}^m \frac{1}{j^2} + \frac{1}{(m+1)^2} \right\} \\ &= \frac{(m+1)^2}{2} \left\{ \left(\sum_{j=1}^{m+1} \frac{1}{j}\right)^2 + \sum_{j=1}^{m+1} \frac{1}{j^2} \right\} \\ &= f_2(m+1). \end{aligned}$$

So the equation hold for any $m \in \mathbb{Z}^+$.

Theorem 4.2 Suppose $Y \sim MGe(m)$, the expectation and variance of Y are as follows

$$E(Y) = \sum_{j=1}^m \frac{m}{j}, \quad (4.5)$$

$$\text{Var}(Y) = \sum_{j=1}^m \binom{m}{j}^2 \left(1 - \frac{j}{m}\right). \quad (4.6)$$

Proof According to Theorem 3.2 and the equation (4.2) and basic condition $\mathbf{p}_m =$

$(p_1, p_2, \dots, p_m), p_i = \frac{1}{m}, i = 1, 2, \dots, m$, we have

$$\begin{aligned}
 E(Y) &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_j(m)} \left(\sum_{k=1}^j p_{i_k} \right)^{-1} + (-1)^{m-1} \\
 &= \sum_{j=1}^{m-1} (-1)^{j-1} \frac{m}{j} \binom{m}{j} + (-1)^{m-1} \\
 &= m \sum_{j=1}^m (-1)^{j-1} \frac{1}{j} \binom{m}{j} \\
 &= m \sum_{j=1}^m \frac{1}{j}. \quad (\text{Property 4.1 (4.2)})
 \end{aligned}$$

The variance of Y ,

$$\begin{aligned}
 \text{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\
 &= \sum_{j=1}^{m-1} (-1)^{j-1} \sum_{\mathbf{i}_j \in \mathcal{I}_m(j)} \left(\sum_{k=1}^j p_{i_k} \right)^{-2} \left(2 - \sum_{k=1}^j p_{i_k} \right) + (-1)^{m-1} - \left(m \sum_{j=1}^m \frac{1}{j} \right)^2 \\
 &= 2 \sum_{j=1}^m (-1)^{j-1} \binom{m}{j} \left(\frac{m}{j} \right)^2 - m \sum_{j=1}^m \frac{1}{j} - \left(m \sum_{j=1}^m \frac{1}{j} \right)^2 \\
 &= m^2 \left\{ \left(\sum_{j=1}^m \frac{1}{j} \right)^2 + \sum_{j=1}^m \frac{1}{j^2} \right\} - m \sum_{j=1}^m \frac{1}{j} - \left(m \sum_{j=1}^m \frac{1}{j} \right)^2 \quad (\text{Property 4.1 (4.3)}) \\
 &= \sum_{j=1}^m \left(\frac{m}{j} \right)^2 \left(1 - \frac{j}{m} \right). \quad (\text{Property 4.1 (4.4)})
 \end{aligned}$$

There is another explanation for the expectation and variance of the UMGe distribution: let the random variable $X_i, i = 1, 2, \dots, m$ represent the number of trails carried out when the " i -th new result" appears. So $X_i \sim Ge \left(1 - \frac{i-1}{m} \right), i = 1, 2, \dots, m$. The total number of trails is $Y = \sum_{i=1}^m X_i$, then it can be calculated as

$$\begin{aligned}
 E(Y) &= \sum_{i=1}^m E(X_i) = 1 + \frac{m}{m-1} + \frac{m}{m-2} + \dots + m, \\
 \text{Var}(Y) &= \sum_{i=1}^m \text{Var}(X_i) = 0 + \frac{1}{m} \left(\frac{m}{m-1} \right)^2 + \frac{2}{m} \left(\frac{m}{m-2} \right)^2 + \dots + \frac{m-1}{m} \left(\frac{m}{1} \right)^2.
 \end{aligned}$$

This is consistent with the formula we have deduced. According to this idea, we can generate random numbers that obey UMGe distribution, i.e., by using the sum of random numbers that subject to different geometric distribution.

5 Simulation Example

In this section, we consider the probability distribution function(pdf) and cumulative distribution function(cdf) of MGe distribution. Note $\Phi(y)$ represents the cumulative distribution function of standard normal distribution.

Example 1 Suppose $Y^{(m)} \sim MGe(\mathbf{p}_m)$, where $\mathbf{p}_m = (p_1, p_2, \dots, p_m)$. The parameters are generated as follows

$$p_i = \frac{2i}{m(m+1)}, i = 1, 2, \dots, m.$$

Let $\mu_m = E(Y^{(m)})$, $\sigma_m^2 = \text{Var}(Y^{(m)})$, discretization of normal distribution, let $\varphi(y) = \Phi[(y - \mu_m)/\sigma_m] - \Phi[(y - \mu_m - 1)/\sigma_m]$, $p(y) = P(Y^{(m)} = y)$, and $F(y) = P(Y^{(m)} \leq y)$ is the cdf of $Y^{(m)}$.

Consider four situations when $m = 5, 10, 15, 20$, respectively, the expectation and variance of the random variable are calculated as

$$\begin{aligned} \mu_5 &= 18.67, & \mu_{10} &= 68.98, & \mu_{15} &= 150.61, & \mu_{20} &= 263.58; \\ \sigma_5^2 &= 169.57, & \sigma_{10}^2 &= 2420.43, & \sigma_{15}^2 &= 11680.48, & \sigma_{20}^2 &= 35959.53. \end{aligned}$$

We draw the image of the pdf of $Y^{(m)}$, the upper bound $u_n = u_m(n)$ of $Y^{(m)}$'s pdf according to formula (3.2), and the scatter plot of discretization of normal distribution $N(\mu_m, \sigma_m^2)$, respectively. The images in four situations are shown in Figure 1.

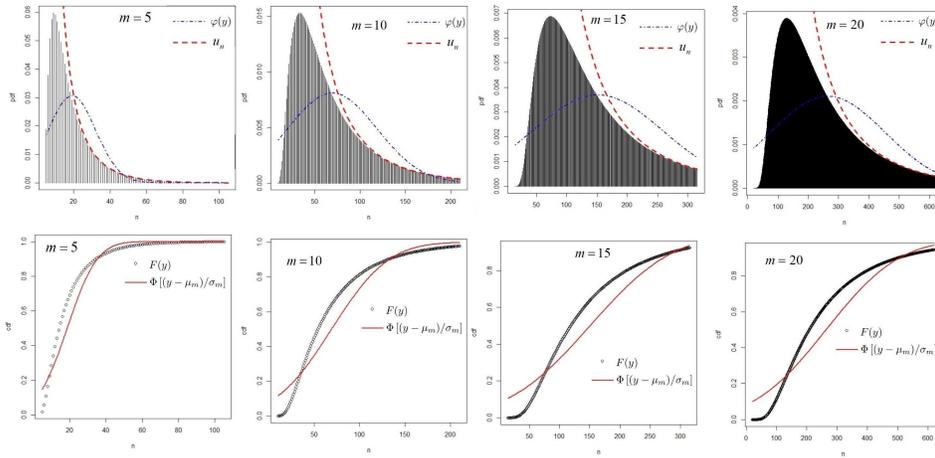


Figure 1 The pdf , cdf , upper bound and normal approximation of MGe distribution in $m = 5, 10, 15, 20$ situations.

From the image point of view, the normal distribution is not a good approximation of the MGe distribution, however, there are two intersections between the density curve of normal distribution and the upper bound u_n of MGe's pdf.

The first intersection is near its expectation, so we can consider to calculate probability by $p(y) = P(Y^{(m)} = y)$ when $y < \mu_m$, using upper bound instead of probability when $y > \mu_m$, that is $p(y) = u_m(n)$.

In addition, the pdf of MGe distribution is always a right-skewed distribution which can not be fitted well by using normal distribution. Because of the complexity of the pdf form of MGe distribution, how to obtain its approximate distribution is an important problem to be studied.

Example 2 Suppose $Y_m \sim MGe(m)$, we still use the notation of Example 1, under conditions $m = 10, 30, 50, 70$, the expectation and variance are

$$\begin{aligned} \mu_{10} &= 29.29, & \mu_{30} &= 119.85, & \mu_{50} &= 224.96, & \mu_{70} &= 338.29; \\ \sigma_{10}^2 &= 125.69, & \sigma_{30}^2 &= 1331.09, & \sigma_{50}^2 &= 3837.87, & \sigma_{70}^2 &= 7652.38. \end{aligned}$$

respectively.

We draw the image of the pdf of Y_m , the upper bound of $Y^{(m)}$'s pdf and the scatter plot of discretization of normal distribution $N(\mu_m, \sigma_m^2)$, respectively. The images in four situations are shown in Figure 2.

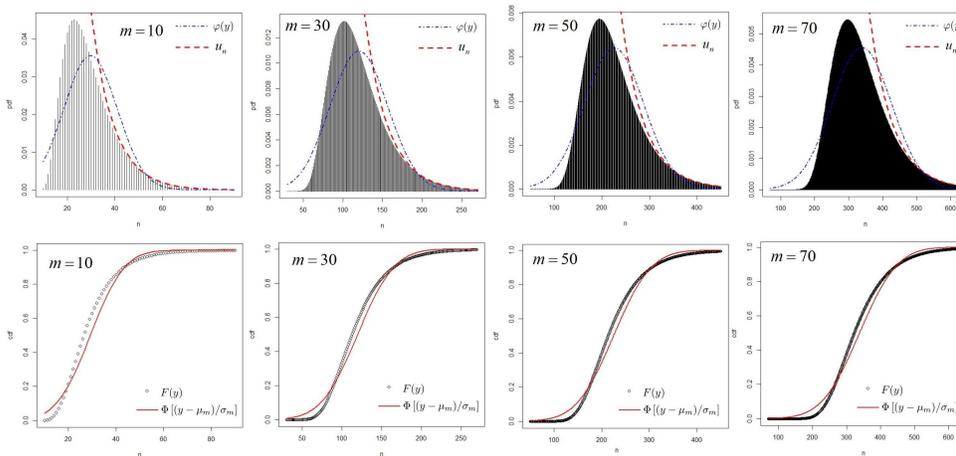


Figure 2 The pdf, cdf, upper bound and normal approximation of UMGe distribution in $m = 10, 30, 50, 70$ situations.

It can be seen that the effect of normal approach to UMGe is better than MGe. We define the quantile of the UMGe distribution as $q_p(m)$ which satisfies

$$q_p(m) = \inf\{y : P(Y_m \leq y) = p\}.$$

We calculate the expectation of each Y_m in the situation of $m = 3, 4, \dots, 40$ as shown in columns 7 and 15 of Table 1, the variance of each Y_m shown in columns 8 and 16 of Table

1, the quantiles of $p = 0.25, 0.5, 0.75, 0.9, 0.95$ are calculated and listed in the corresponding columns in Table 1.

Table 1 The quantile, expectation and variance of UMGe

m	0.25	0.5	0.75	0.9	0.95	μ	σ^2	m	0.25	0.5	0.75	0.9	0.95	μ	σ^2
3	3	4	6	8	10	6	7	22	62	75	93	115	130	81	693
4	5	6	9	12	15	8	14	23	65	80	99	121	137	86	762
5	7	9	13	17	20	11	25	24	69	84	104	127	144	91	833
6	9	12	17	22	26	15	39	25	73	89	110	134	151	95	908
7	12	16	21	27	32	18	56	26	77	93	115	140	158	100	986
8	15	19	25	32	37	22	76	27	81	98	121	147	166	105	1068
9	17	22	29	37	43	25	99	28	85	103	126	153	173	110	1152
10	20	26	34	43	50	29	126	29	89	107	132	160	180	115	1240
11	23	30	38	49	56	33	155	30	93	112	137	167	188	120	1331
12	26	34	43	54	62	37	188	31	97	117	143	173	195	125	1425
13	30	37	48	60	69	41	224	32	101	122	149	180	202	130	1523
14	33	41	53	66	75	46	263	33	105	127	154	187	210	135	1624
15	36	45	57	72	82	50	306	34	110	132	160	193	217	140	1728
16	40	50	62	78	89	54	352	35	114	136	166	200	225	145	1835
17	43	54	67	84	95	58	400	36	118	141	172	207	232	150	1946
18	47	58	73	90	102	63	453	37	122	146	178	214	240	155	2060
19	50	62	78	96	109	67	508	38	127	151	183	221	247	161	2177
20	54	66	83	102	116	72	567	39	131	156	189	228	255	166	2298
21	58	71	88	108	123	77	628	40	135	161	195	234	263	171	2421

Let us calculate the quantile of Y_m at $p = 0.25, 0.5, 0.75, 0.9, 0.95$ for $m = 3, 4, \dots, 40$. The scatter plot of quantile and the fitting curve of the quadratic regression model drawn with the horizontal axis of m and the vertical axis of $q_p(m)$ are shown in Figure 3 (a).

From the data in Table 1 and the trend of quantile $q_p(m)$ in Figure 3(a), we can suppose the quadratic regression model of quantile $q_p(m)$ and m as

$$\hat{q}_p(m) = \hat{\beta}_p^{(0)} + \hat{\beta}_p^{(1)}m + \hat{\beta}_p^{(2)}m^2 \tag{5.1}$$

The regression model that establishes according to (5.1) is very significant in $p = 0.25, 0.5, 0.75, 0.9, 0.95$ by calculation, and the regression coefficients as shown in Table 2.

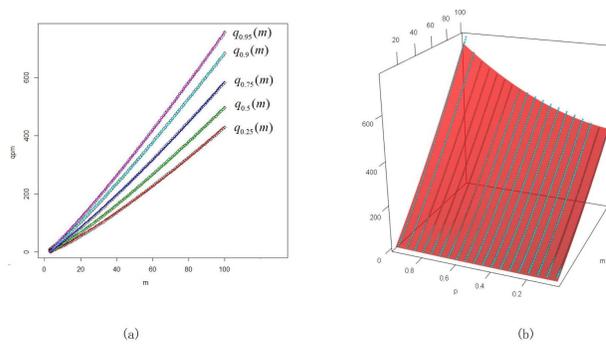
Furthermore, calculate all quantile values for $m = 3, 4, \dots, 100, p = 0.05, 0.06, \dots, 0.97$, after regression analysis, the empirical regression equation is

$$\hat{q}(p, m) = -13.42552 + 2.83025m + 0.01111m^2 + 0.86067mp + 2.82063mp^2. \tag{5.2}$$

Draw the surface graph of (5.2) and the scatter graph of $(m, p, q(p, m))$ shown in Figure 3(b).

Table 2 The quantile fitting curve coefficient of UMGe

p	$\hat{\beta}_p^{(0)}$	$\hat{\beta}_p^{(1)}$	$\hat{\beta}_p^{(2)}$
0.25	-14.43	3.302	0.01163
0.50	-15.55	3.976	0.01175
0.75	-16.23	4.832	0.01192
0.90	-17.11	5.837	0.01189
0.95	-17.52	6.548	0.01199

Figure 3 (a) Quantile $q_p(m)$ for $p = 0.25, 0.5, 0.75, 0.9, 0.95$; (b) Empirical regression surface of $\hat{q}(p, m)$ and scatter of $(m, p, q(p, m))$.

For example, The 0.95 quantile of Y_{10} can be obtained from

$$\hat{q}_{0.95}(10) = \beta_{0.95}^{(0)} + \beta_{0.95}^{(1)} \times 10 + \beta_{0.95}^{(2)} \times 10^2 \approx 49.15.$$

Using (5.2) to estimate as

$$\hat{q}(10, 0.95) \approx 49.62044.$$

The quantile calculated by (5.1) is more accurate than (5.2), however, in many cases (5.2) is more convenient to use. When m is larger, the estimated value is often smaller than the real value, so it can be considered to build a nonparametric regression model based on the calculated data, so the calculation is more accurate.

6 Discussion

In this paper, the multiple geometric distribution is discussed under the condition of putting back, which means that the probability of occurrence of event A_i is constant for each test. If the sampling conditions are changed, and the samples are taken one by one in a finite population, the distribution of the total number of tests will be different if several specified results occur. In many cases, upper bound approximation and normal approximation can be considered.

References

- [1] Ahmed Z Afify, Zohdy M Nofal, Nadeem Shafique Butt. Transmuted complementary weibull geometric distribution[J]. Pakistan Journal of Statistics & Operation Research, 2014, 10(4): 435–454.
- [2] Jayakumar K, Babu M G. Discrete weibull geometric distribution and its properties[J]. Communications in Statistics-Theory and Methods, 2018, 47(7): 1767–1783.
- [3] Kemp, Adrienne. The q-beta-geometric distribution as a model for fecundability[J]. Communications in Statistics: Theory and Methods, 2001, 30(11): 2373–2384.
- [4] Li G H, Zhang C Q. The pseudo component transformation design for experiment with mixture[J]. Statistics & Probability Letters, 2017, 131: 19–24.
- [5] Miller F P, Vandome A F, Mcbrewster J. Geometric distribution[M]. Springer New York, 2008.
- [6] Muwafi A N P A A. Waiting for the kth consecutive success and the fibonacci sequence of order k[J]. Fibonacci Quarterly, 1980, 20(1): 28–32.
- [7] Pedro L Ramos, Fernando A Moala, Jorge A Achcar. Objective priors for estimation of extended exponential geometric distribution[J]. Journal of Modern Applied Statistical Methods Jmasm, 2014, 13(2): 226–243.
- [8] Philippou A N, Georghiou C, Philippou G N. A generalized geometric distribution and some of its properties[J]. Statistics & Probability Letters, 1983, 1(4): 171–175.
- [9] Porwal S. Generalized distribution and its geometric properties associated with univalent functions[J]. Journal of Complex Analysis, 2018, 2018: 1–5.
- [10] Xiao S J, Li G H, Yang W Z. A kind of generalized geometric distribution(in Chinese)[J]. Jilin Normal University Journal, 2015, 36(2): 43–50.

多项几何分布

李光辉¹, 李俊鹏², 张崇岐²

(1.凯里学院理学院, 贵州 凯里 556011)

(2.广州大学经济与统计学院, 广东 广州 510006)

摘要: 本文研究了在伯努利试验下的收集问题, 利用混料格子点集的理论, 推导出了多项几何分布的概率函数. 在伯努利试验中, 如果假设各种试验结果发生的概率都相等, 进一步提出了均匀多项几何分布. 我们得到了两类分布的概率函数以及期望与方差, 通过模拟验证了这两类分布与正态分布的差异, 并由模拟结果建立关于概率与试验次数的多项式回归模型, 使用该模型可以有效的简化计算.

关键词: 多项分布; 几何分布; 概率函数

MR(2010)主题分类号: 62E20; 62G30

中图分类号: O211