Expectation identity of the hypergeometric distribution and its application in the calculations of high-order origin moments

Yuan-Quan Wang\textsuperscript{a†}, Ying-Ying Zhang\textsuperscript{a,b†}, and Jia-Lei Liu\textsuperscript{a}

\textsuperscript{a}Department of Statistics and Actuarial Science, College of Mathematics and Statistics, Chongqing University, Chongqing, China; \textsuperscript{b}Chongqing Key Laboratory of Analytic Mathematics and Applications, Chongqing University, Chongqing, China

ABSTRACT
We provide a novel method to analytically calculate the high-order origin moments of a hypergeometric distribution, that is, the expectation identity method. First, the expectation identity of the hypergeometric distribution is discovered and summarized in a theorem. After that, we analytically calculate the first four origin moments of the hypergeometric distribution by using the expectation identity. Furthermore, we analytically calculate the general $k$th ($k = 1, 2, \ldots$) origin moment of the hypergeometric distribution by using the expectation identity, and the results are summarized in a theorem. Moreover, we use the general $k$th origin moment to validate the first four origin moments of the hypergeometric distribution. Next, the coefficients of the first ten origin moments of the hypergeometric distribution are summarized in a table containing Stirling numbers of the second kind. Moreover, the general $k$th origin moment of the hypergeometric distribution by using the expectation identity is restated by another theorem involving Stirling numbers of the second kind. Finally, we provide some numerical and theoretical results.

1. Introduction
The hypergeometric distribution has attracted continuous interest in the literature since 2000. Childs and Balakrishnan (2000) examined some approximations to the multivariate hypergeometric distribution with applications to hypothesis testing. Kumar (2002) introduced extended generalized hypergeometric probability distributions. Hida and Akahira (2003) considered an improvement on the approximation to the generalized hypergeometric distribution. Dinwoodie, Matusevich, and Mosteig (2004) presented two new transform methods for computing with hypergeometric distributions on lattice points. Hush and Scovel (2005) provided an improved concentration of measure theorem for the hypergeometric distribution. Kumar (2007) studied some properties of bivariate generalized hypergeometric probability distributions. Lahiri, Chatterjee, and Maiti (2007) investigated normal approximation to the hypergeometric distribution in nonstandard cases and established a
sub-Gaussian Berry-Esseen theorem. Fog (2008) developed sampling methods for Wallenius’ and Fisher’s noncentral hypergeometric distributions. Li and Tian (2010) provided a simple algorithm for the high-order origin moments of the hypergeometric distribution. Eisenga and Pelzer (2011) proposed saddlepoint approximations to the mean and variance of the extended hypergeometric distribution. Lebrun (2013) proposed an efficient time/space algorithm to compute rectangular probabilities of multinomial, multivariate hypergeometric and multivariate Plya distributions. Liu and Wang (2014) obtained the high-order origin moments, the high-order central moments, and the high-order cumulants of the hypergeometric distribution. de Klerk, Laurent, and Sun (2015) considered an error analysis for polynomial optimization over the simplex based on the multivariate hypergeometric distribution. Danielian, Chitchyan, and Farbod (2016) proposed a new regularly varying generalized hypergeometric distribution of the second type. Greene and Wellner (2017) established exponential bounds for the hypergeometric distribution. Mano (2017) discussed partition structure and the A-hypergeometric distribution associated with the rational normal curve. Takayama, Kuriki, and Takemura (2018) considered A-hypergeometric distributions and Newton polytopes. Hafid, Hafid, and Samih (2020) proposed a novel methodology-based joint hypergeometric distribution to analyze the security of sharded blockchains. Krishnamoorthy and Lv (2020) constructed prediction intervals for hypergeometric distributions. Themangani, Porwal, and Magesh (2020) introduced a generalized hypergeometric distribution and discussed its applications on univalent functions. For earlier literature on the hypergeometric distribution, we refer readers to Riordan (1937); Kemp and Kemp (1956); Lieberman and Owen (1962); Harkness (1965); Brunk, Holstein, and Williams (1968); Ord (1968); Skibinsky (1970); Molenaar (1973); Tripathi and Gurland (1977); Guenther (1978); Chvatal (1979); Lai (1979); Sibuya and Shimizu (1981); Panaretos (1983); Ling and Pratt (1984); Kachitvichyanukul and Schmeiser (1988); Kochman, Murray, and West (1989); Tohma et al. (1991); Wu (1993); Lin and Hsiung (1994); Lin, Hsiung, and Hsiao (1994); Ma (1999).

Exponential families include the continuous families (normal, gamma, and beta) and the discrete families (binomial, Poisson, and negative binomial). For the continuous exponential families, there are some expectation identities that rely on integration by parts. Stein’s Lemma (Lemma 3.6.5 in Casella and Berger (2002)) gives an expectation identity for the normal family. Moreover, the expectation identities of the gamma and beta families are given in Exercise 3.49 of Casella and Berger (2002). It is worth noting that these expectation identities are useful in the calculations of high-order origin moments of the corresponding families. The discrete analogs of the expectation identities of the continuous exponential families are given in Theorem 3.6.8 of Casella and Berger (2002), which gives the expectation identities for the Poisson and negative binomial families (see also Hwang (1982)). Note that the Poisson and negative binomial families take countable infinite values. For the binomial family taking finite values, similar to the deviations of the Poisson and negative binomial expectation identities, Zhang, Rong, and Li (2019) discovered an expectation identity for the binomial family and obtained a closed-form formula of the high-order origin moments for the binomial family by using the binomial expectation identity.

The analytical calculations of the high-order origin moments of the hypergeometric distribution are quite challenging. In this paper, we provide a novel method to analytically calculate the high-order origin moments of the hypergeometric distribution, that is, the
expectation identity method. First, the expectation identity of the hypergeometric distribution is discovered and summarized in a theorem. After that, we analytically calculate the first four origin moments and the general \( k \)th \((k = 1, 2, \ldots)\) origin moment of the hypergeometric distribution by the expectation identity method. Moreover, we use the general \( k \)th origin moment to validate the first four origin moments of the hypergeometric distribution. Finally, the coefficients of the first 10 origin moments of the hypergeometric distribution are shown in a table. By comparing the tables in the literature, we find that the coefficients are Stirling numbers of the second kind.

The rest of the paper is organized as follows. In the next Section 2, we discover an expectation identity of the hypergeometric distribution. In Section 3, we will use the expectation identity of the hypergeometric distribution to calculate its high-order origin moments. First, we will use the expectation identity of the hypergeometric distribution to calculate its first four origin moments. Second, we will use the expectation identity of the hypergeometric distribution to calculate its general \( k \)th \((k = 1, 2, \ldots)\) origin moment. Third, we will use the general \( k \)th origin moment to validate the first four origin moments of the hypergeometric distribution. Finally, the coefficients of the first 10 origin moments of the hypergeometric distribution are summarized in a table. In Section 4, we will provide some numerical and theoretical results. Some conclusions and discussions are provided in Section 5.

2. Expectation identity of the hypergeometric distribution

In this section, we discover an expectation identity of the hypergeometric distribution. The expectation identity is useful in analytical calculations of the high-order origin moments of the hypergeometric distribution.

Let \( Y \sim HG(N, M, n) \) have a hypergeometric distribution with probability mass function (pmf) given by (see Casella and Berger (2002))

\[
P(Y = y|N, M, n) = \binom{M}{y} \binom{N-M}{n-y} \binom{N}{n}, \quad y = 0, 1, \ldots, n,
\]

where \( N \) is the total number of identical balls in a large urn, \( M \) is the number of red balls in the urn, \( N - M \) is the number of green balls in the urn, \( n \) is the total number of selected balls, \( y \) is the number of selected balls that are red, and \( n - y \) is the number of selected balls that are green. Similarly, let \( X \sim HG(N - 1, M - 1, n - 1) \) have a hypergeometric distribution with pmf given by

\[
P(X = x|N - 1, M - 1, n - 1) = \binom{M - 1}{x} \binom{N-M}{n-1-x} \binom{N - 1}{n - 1}, \quad x = 0, 1, \ldots, n - 1.
\]

By definition,

\[
EY^k = \sum_{y=0}^{n} y^k \binom{M}{y} \binom{N-M}{n-y} \binom{N}{n},
\]

for \( k = 0, 1, 2, \ldots \).
The expectation identity of the hypergeometric distribution is discovered and summarized in the following theorem whose proof is simple as with many important results.

**Theorem 1.** Let \( X \sim HG(N - 1, M - 1, n - 1) \) and \( Y \sim HG(N, M, n) \). Moreover, let \( g(x) \) be a function satisfying \(-\infty < E[g(X)] < \infty\) and \(-\infty < g(-1) < \infty\). Then we have the following expectation identity of the hypergeometric distribution:

\[
E[g(X)] = E \left[ \frac{N}{Mn} Yg(Y - 1) \right].
\]  

**Proof.** By the definition of expectation, we have

\[
E[g(X)] = \sum_{x=0}^{n-1} g(x)P(X = x)
\]

\[
= \sum_{x=0}^{n-1} g(x) \frac{\binom{M-1}{x} \binom{N-M}{n-1-x}}{\binom{N-1}{n-1}}
\]

\[
= \sum_{y=1}^{n} g(y-1) \frac{\binom{M-1}{y-1} \binom{N-M}{n-y}}{\binom{N-1}{n-1}} \quad \text{(let } y = x + 1).\]

According to the identities of combinatorial numbers

\[
\binom{M-1}{y-1} = \frac{y}{M} \binom{M}{y} \quad \text{and} \quad \binom{N-1}{n-1} = \frac{n}{N} \binom{N}{n},
\]

we obtain

\[
E[g(X)] = \sum_{y=1}^{n} g(y-1)y \frac{N}{Mn} \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}}
\]

\[
= \sum_{y=0}^{n} g(y-1)y \frac{N}{Mn} \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}} \quad \text{(the added term is 0)}
\]

\[
= \sum_{y=0}^{n} g(y-1)y \frac{N}{Mn} P(Y = y)
\]

\[
= E \left[ \frac{N}{Mn} Yg(Y - 1) \right].
\]

The proof of the theorem is complete. \(\square\)

3. **Using the expectation identity of the hypergeometric distribution to calculate its high-order origin moments**

In this section, we will use the expectation identity of the hypergeometric distribution to calculate its high-order origin moments. First, we will use the expectation identity of
the hypergeometric distribution to calculate its first four origin moments. Second, we will use the expectation identity of the hypergeometric distribution to calculate its general \( k \)th \( (k = 1, 2, \ldots) \) origin moment. Third, we will use the general \( k \)th origin moment to validate the first four origin moments of the hypergeometric distribution. Finally, the coefficients of the first 10 origin moments of the hypergeometric distribution are summarized in a table.

### 3.1. The analytical calculations of the first four origin moments of the hypergeometric distribution by using the expectation identity

In this subsection, we will analytically calculate the first four origin moments of the hypergeometric distribution \( Y \sim HG(N, M, n) \) by the expectation identity (2).

First, let us calculate \( E_Y \). Let \( g(x) = (1 + x)^0 = 1 \). By the expectation identity (2), we have

\[
E_Y = E \left( \frac{N}{Mn} Y \right).
\]

Hence,

\[
E_Y = \frac{Mn}{N}. \tag{3}
\]

Therefore,

\[
E_X = \frac{(M - 1)(n - 1)}{N - 1}, \tag{4}
\]

which can be obtained by replacing \( N, M, \) and \( n \) by \( N - 1, M - 1, \) and \( n - 1 \) in the formula of \( E_Y \) (3).

Second, let us calculate \( E_{Y^2} \). Let \( g(x) = (1 + x)^1 \). By the expectation identity (2), we have

\[
E(1 + X) = E \left( \frac{N}{Mn} YY \right).
\]

Hence, by (4), we obtain

\[
E_{Y^2} = \frac{Mn}{N} E(1 + X)
= \frac{Mn}{N} (1 + EX)
= \frac{Mn}{N} + \frac{Mn (M - 1)(n - 1)}{N - 1}
= \frac{Mn}{N} + \frac{M(M - 1)n(n - 1)}{N(N - 1)}. \tag{5}
\]

Therefore,

\[
E_{X^2} = \frac{(M - 1)(n - 1)}{N - 1} + \frac{(M - 1)(M - 2)(n - 1)(n - 2)}{(N - 1)(N - 2)}, \tag{6}
\]

which can be obtained by replacing \( N, M, \) and \( n \) by \( N - 1, M - 1, \) and \( n - 1 \) in the formula of \( E_{Y^2} \) (5).
Third, let us calculate $EY^3$. Let $g(x) = (1 + x)^2$. By the expectation identity (2), we have

$$E(1 + X)^2 = E\left(\frac{N}{Mn} YY^2\right).$$

Hence, by (4) and (6), we obtain

$$EY^3 = \frac{Mn}{N} E(1 + X)^2$$

$$= \frac{Mn}{N} E(1 + 2X + X^2)$$

$$= \frac{Mn}{N} (1 + 2EX + EX^2)$$

$$= \frac{Mn}{N} \left[ 1 + 2\frac{(M - 1)(n - 1)}{N - 1} + \frac{(M - 1)(n - 1)}{N - 1} + \frac{(M - 1)(M - 2)(n - 1)(n - 2)}{(N - 1)(N - 2)} \right]$$

$$= \frac{Mn}{N} + 3 \frac{M(M - 1)n(n - 1)}{N(N - 1)} + \frac{M(M - 1)(M - 2)n(n - 1)(n - 2)}{N(N - 1)(N - 2)}. \quad (7)$$

Therefore,

$$EX^3 = \frac{(M - 1)(n - 1)}{N - 1} + 3 \frac{(M - 1)(M - 2)(n - 1)(n - 2)}{(N - 1)(N - 2)}$$

$$+ \frac{(M - 1)(M - 2)(M - 3)(n - 1)(n - 2)(n - 3)}{(N - 1)(N - 2)(N - 3)}, \quad (8)$$

which can be obtained by replacing $N, M,$ and $n$ by $N - 1, M - 1,$ and $n - 1$ in the formula of $EY^3$ (7).

Fourth, let us calculate $EY^4$. Let $g(x) = (1 + x)^3$. By the expectation identity (2), we have

$$E(1 + X)^3 = E\left(\frac{N}{Mn} YY^3\right).$$

Hence, by (4, 6), and (8), we obtain

$$EY^4 = \frac{Mn}{N} E(1 + X)^3$$

$$= \frac{Mn}{N} E(1 + 3X + 3X^2 + X^3)$$

$$= \frac{Mn}{N} (1 + 3EX + 3EX^2 + EX^3)$$

$$= \frac{Mn}{N} \left[ 1 + 3\frac{(M - 1)(n - 1)}{N - 1} + 3\frac{(M - 1)(n - 1)}{N - 1} + 3\frac{(M - 1)(M - 2)(n - 1)(n - 2)}{(N - 1)(N - 2)} \right]$$

$$= \frac{Mn}{N} + 7 \frac{M(M - 1)n(n - 1)}{N(N - 1)} + 6 \frac{M(M - 1)(M - 2)n(n - 1)(n - 2)}{N(N - 1)(N - 2)}$$

$$+ \frac{M(M - 1)(M - 2)(M - 3)n(n - 1)(n - 2)(n - 3)}{N(N - 1)(N - 2)(N - 3)}. \quad (9)$$
Therefore,

\[
\text{EX}^4 = \frac{(M - 1)(n - 1)}{N - 1} + \frac{7(M - 1)(M - 2)(n - 1)(n - 2)}{(N - 1)(N - 2)} + 6 \frac{(M - 1)(M - 2)(M - 3)(n - 1)(n - 2)(n - 3)}{(N - 1)(N - 2)(N - 3)} + \frac{(M - 1)(M - 2)(M - 3)(M - 4)(n - 1)(n - 2)(n - 3)(n - 4)}{(N - 1)(N - 2)(N - 3)(N - 4)},
\]

which can be obtained by replacing \(N, M, \) and \(n\) by \(N/C_0, M/C_0, \) and \(n/C_0\) in the formula of \(\text{EX}^4\) (9).

It is easy to see that our analytical formulas of the first four origin moments of the hypergeometric distribution by using the expectation identity (2) are the same as those obtained in Li and Tian (2010), in which they use the definition of expectation for the calculations.

### 3.2. The analytical calculations of the kth origin moment of the hypergeometric distribution by using the expectation identity

In this subsection, we will analytically calculate the \(k\)th origin moment of the hypergeometric distribution \(Y \sim \text{HG}(N, M, n)\) by the expectation identity (2).

Define

\[
H_k = \frac{\prod_{i=0}^k (M - i) \prod_{i=0}^k (n - i)}{\prod_{i=0}^k (N - i)}, \quad k = 0, 1, 2, 3, \ldots
\]

In particular,

\[
H_0 = \frac{Mn}{N},
\]

\[
H_1 = \frac{M(M - 1)n(n - 1)}{N(N - 1)},
\]

\[
H_2 = \frac{M(M - 1)(M - 2)n(n - 1)(n - 2)}{N(N - 1)(N - 2)},
\]

\[
H_3 = \frac{M(M - 1)(M - 2)(M - 3)n(n - 1)(n - 2)(n - 3)}{N(N - 1)(N - 2)(N - 3)}.
\]

From the calculation results in the previous section, we find that the first four origin moments of the hypergeometric distribution \(Y \sim \text{HG}(N, M, n)\) can be written as a linear combination of \(H_0, H_1, H_2,\) and \(H_3.\) This fact inspires us to claim that the general \(k\)th origin moment of \(Y \sim \text{HG}(N, M, n),\) \(EY^k (k = 0, 1, 2, \ldots),\) can also be written as a linear combination of \(H_0, H_1, \ldots, H_{k-1}\) given in (11).

To prove Theorem 2, we need the following lemma, which gives a relationship between the expressions of \(EY^k\) and \((Mn/N)\text{EX}^k.\)

**Lemma 1.** Let \(X \sim \text{HG}(N - 1, M - 1, n - 1)\) and \(Y \sim \text{HG}(N, M, n).\) If

\[
EY^k = \sum_{j=1}^{k} a_j H_{j-1}
\]
is a linear combination of \(H_0, H_1, \ldots, H_{k-1}\) given in (11), then
\[
\frac{Mn}{N}EX^k = \sum_{j=1}^{k} a_j H_j
\]
is a linear combination of \(H_1, H_2, \ldots, H_k\).

**Proof.** If
\[
EY^k = \sum_{j=1}^{k} a_j H_{j-1}
\]
(12)
is a linear combination of \(H_0, H_1, \ldots, H_{k-1}\), then
\[
EX^k = \sum_{j=1}^{k} a_j \left[ \prod_{i=0}^{j-1} (M - i) \right] \left[ \prod_{i=0}^{j-1} (n - i) \right] \prod_{i=0}^{j-1} (N - i)
\]
is a linear combination of \(H_0, H_1, \ldots, H_{k-1}\), then
\[
EY^k = \sum_{j=1}^{k} a_j \left[ \prod_{i=0}^{j-1} (M - i) \right] \left[ \prod_{i=0}^{j-1} (n - i) \right] \prod_{i=0}^{j-1} (N - i)
\]
which can be obtained by replacing \(N, M, n\) by \(N - 1, M - 1, n - 1\) in the formula of \(EY^k\) (12). Hence,
\[
\frac{Mn}{N}EX^k = \sum_{j=1}^{k} a_j \left[ \prod_{i=0}^{j-1} (M - i) \right] \left[ \prod_{i=0}^{j-1} (n - i) \right] \prod_{i=0}^{j-1} (N - i)
\]
\[
= \sum_{j=1}^{k} a_j \frac{M(M - 1) \cdots (M - j)n(n - 1) \cdots (n - j)}{N(N - 1) \cdots (N - j)}
\]
\[
= \sum_{j=1}^{k} a_j \left[ \prod_{i=0}^{j-1} (M - i) \right] \left[ \prod_{i=0}^{j-1} (n - i) \right] \prod_{i=0}^{j-1} (N - i)
\]
\[
= \sum_{j=1}^{k} a_j H_j,
\]
is a linear combination of \(H_1, H_2, \ldots, H_k\).

The proof of the lemma is complete.

By Lemma 1 and the mathematical induction method, we have the following theorem, in which \(EY^k\) is expressed as a linear combination of \(H_0, H_1, \ldots, H_{k-1}\) given in (11).

**Theorem 2.** Let \(Y \sim HG(N, M, n)\). Then
\[
EY^k = \sum_{j=1}^{k} A(k, j) H_{j-1},
\]
(13)
is a linear combination of \(H_0, H_1, \ldots, H_{k-1}\) given in (11), where the coefficients \(A(k, j)\) are given by the following recurrence formula:
\[ A(k,j) = \begin{cases} 
1, & k = j \text{ or } j = 1, \\
\sum_{i=j+1}^{k-1} \binom{k-1}{i} A(i,j-1), & k > j > 1.
\end{cases} \quad (14) \]

**Proof.** We will use the mathematical induction method to prove this theorem.

1. When \( k = 1 \),

\[ EY = \frac{Mn}{N} = H_0 = A(1,1)H_0 = \sum_{j=1}^{l} A(1,j)H_{j-1}, \]

and thus (13) is correct.

2. Suppose that (13) is true when \( 1 \leq k \leq l \). According to Lemma 1, we have

\[ \frac{Mn}{N}EX^k = \sum_{j=1}^{k} A(k,j)H_j, \quad (15) \]

for \( 1 \leq k \leq l \).

3. When \( k = l + 1 \), let \( g(x) = (1 + x)^l \). By expectation identity (2), we have

\[ E(1 + X)^l = E \left( \frac{N}{Mn} YY^l \right). \]

Hence, by (15), we obtain

\[ EY^{l+1} = \frac{Mn}{N}E(1 + X)^l \]

\[ = \frac{Mn}{N}E \left[ \left( \frac{l}{0} \right) X^0 + \left( \frac{l}{1} \right) X^1 + \left( \frac{l}{2} \right) X^2 + \left( \frac{l}{3} \right) X^3 + \ldots + \left( \frac{l}{l-1} \right) X^{l-1} + \left( \frac{l}{l} \right) X^l \right] \]

\[ = \frac{Mn}{N} \left[ \left( \frac{l}{0} \right) + \left( \frac{l}{1} \right) EX^1 + \left( \frac{l}{2} \right) EX^2 + \left( \frac{l}{3} \right) EX^3 + \ldots + \left( \frac{l}{l-1} \right) EX^{l-1} + \left( \frac{l}{l} \right) EX^l \right] \]

\[ = \frac{Mn}{N} + \left( \frac{l}{l-1} \right) \frac{Mn}{N}EX^{l-1} + \left( \frac{l}{l} \right) \frac{Mn}{N}EX^l \]

\[ = H_0 + \left( \frac{l}{1} \right) [A(1,1)H_1] + \left( \frac{l}{2} \right) [A(2,1)H_1 + A(2,2)H_2] \]

\[ + \left( \frac{l}{3} \right) [A(3,1)H_1 + A(3,2)H_2 + A(3,3)H_3] + \ldots \]

\[ + \left( \frac{l}{l-1} \right) \left[ \sum_{j=1}^{l-1} A(l-1,j)H_j \right] + \left( \frac{l}{l} \right) \left[ \sum_{j=1}^{l} A(l,j)H_j \right] \]

\[ = H_0 + \sum_{i=1}^{l} \left( \frac{l}{i} \right) A(i,1)H_1 + \sum_{i=2}^{l} \left( \frac{l}{i} \right) A(i,2)H_2 + \ldots + \sum_{i=l}^{l} \left( \frac{l}{i} \right) A(i,l)H_l \]

\[ = A(l+1,1)H_0 + A(l+1,2)H_1 + A(l+1,3)H_2 + \ldots + A(l+1,l+1)H_l \]

\[ = \sum_{j=1}^{l+1} A(l+1,j)H_{j-1}. \]

Therefore, when \( k = l + 1 \), (13) also holds true.
In conclusion, by the mathematical induction method, (13) is established and the proof of the theorem is completed.

3.3. Verifications of the first four origin moments by the kth origin moment

In this subsection, we will verify the first four origin moments of the hypergeometric distribution \(Y \sim HG(N, M, n)\) by the kth origin moment (13). The verifications ensure that the general formula for the kth origin moment (13) is correct.

(1) When \(k = 1\),
\[
EY = \sum_{j=1}^{1} A(1, j)H_{j-1}
\]
\[
= A(1, 1)H_{0}
\]
\[
= H_{0},
\]
which is the same as in (3).

(2) When \(k = 2\),
\[
EY^2 = \sum_{j=1}^{2} A(2, j)H_{j-1}
\]
\[
= A(2, 1)H_{0} + A(2, 2)H_{1}
\]
\[
= H_{0} + H_{1},
\]
which is the same as in (5).

(3) When \(k = 3\),
\[
EY^3 = \sum_{j=1}^{3} A(3, j)H_{j-1}
\]
\[
= A(3, 1)H_{0} + A(3, 2)H_{1} + A(3, 3)H_{2}.
\]
According to (14),
\[
A(3, 1) = A(3, 3) = 1
\]
and
\[
A(3, 2) = \sum_{i=1}^{2} \binom{2}{i}A(i, 1) = \left(\binom{2}{1}\right)A(1, 1) + \left(\binom{2}{2}\right)A(2, 1) = 3.
\]
Therefore,
\[
EY^3 = H_{0} + 3H_{1} + H_{2},
\]
which is the same as in (7).

(4) When \(k = 4\),
\[
EY^4 = \sum_{j=1}^{4} A(4, j)H_{j-1}
\]
\[
= A(4, 1)H_{0} + A(4, 2)H_{1} + A(4, 3)H_{2} + A(4, 4)H_{3}.
\]
According to (14),
A(4, 1) = A(4, 4) = 1,

\[ A(4, 2) = \sum_{i=1}^{3} \binom{3}{i} A(i, 1) = \binom{3}{1} A(1, 1) + \binom{3}{2} A(2, 1) + \binom{3}{3} A(3, 1) = 7, \]

and

\[ A(4, 3) = \sum_{i=2}^{3} \binom{3}{i} A(i, 2) = \binom{3}{2} A(2, 2) + \binom{3}{3} A(3, 2) = 6. \]

Therefore,

\[ EY^4 = H_0 + 7H_1 + 6H_2 + H_3, \]

which is the same as in (9).

### 3.4. The coefficients table of the kth origin moment of the hypergeometric distribution

The coefficients of the first 10 origin moments of the hypergeometric distribution \( Y \sim HG(N, M, n) \) are shown in Table 1.

Comparing Table 1 with the table of Stirling numbers of the second kind and exponential numbers (P310) in Comtet (1974) and Table 1 in Li and Tian (2010), we find that

\( A(k, j) = S(k, j) \)

are Stirling numbers of the second kind. Moreover, the \( A(k, j) \) given by (14) satisfy the “vertical” recurrence relations given in Theorem B [3c] (P209) in Comtet (1974), and this fact further reassures that \( A(k, j) \) are Stirling numbers of the second kind.

Therefore, Theorem 2 can be restated by the following theorem in the symbol of \( S(k, j) \), which are Stirling numbers of the second kind.

**Theorem 3.** Let \( Y \sim HG(N, M, n) \). Then

\[ EY^k = \sum_{j=1}^{k} S(k, j)H_{j-1}, \tag{16} \]

is a linear combination of \( H_0, H_1, \ldots, H_{k-1} \) given in (11), where the coefficients \( S(k, j) \) are Stirling numbers of the second kind.

### 4. Numerical, theoretical results

In this section, we will provide an assessment of the computation complexity (see Chapter 5 of Progri (2011)) which is given in number of multiplications of the Definition Method (DM) (1) and the Analytical Method (AM) (16). Moreover, we will investigate the memory size of the DM and the AM. Furthermore, we will carry out some numerical experiments to compare the computations of \( EY^k \) and the computing times by the DM and the AM. Finally, we will compare the absolute errors and the relative errors between the two methods. See the Numerical, Theoretical Results section of
Progri (2021a, 2021b, 2021c) for details. The numerical experiments are carried out on a Lenovo computer 11th Gen Intel(R) Core(TM) i5-11300H @ 3.10 GHz with 16.0 GB RAM using R version 4.1.1.

Now let us provide an assessment of the computation complexity which is given in number of multiplications of the DM and the AM. For the DM (1),

\[
\binom{M}{y} = \frac{M!}{y!(M-y)!} = \frac{M(M-1) \cdots (M-y+1)}{y!}
\]

which can be computed in \(O(2y-1)\). Similarly, \(\binom{N-M}{n-y}\) can be computed in \(O(2(n-y)-1)\), and \(\binom{N}{n}\) can be computed in \(O(2n-1)\). Moreover, \(y^k\) can be computed in \(O(k-1)\). Hence,

\[
y^k \binom{M}{y} \binom{N-M}{n-y} \binom{N}{n}
\]

can be computed in

\[
O(k-1) + O(2y-1) + O(2(n-y)-1) + O(2n-1) + O(3)
\]
\[
= O(k + 4n - 1) = O(k + 4n).
\]

Therefore, \(EY^k\) by the DM can be computed in

\[
(n + 1)O(k + 4n) = O\left( (n + 1)(k + 4n) \right) = O(n(k + 4n)).
\]

For the AM (16), \(S(k,j)\) can be precomputed and stored in a lower-triangular matrix \(S\) of size \(k \times k\), which is independent of \(N, M,\) and \(n\), and thus the computation cost of \(S(k,j)\) can be regarded as \(0\). Note that \(H_k\) (11) can be computed in \(O(3k+2)\), and thus \(H_{j-1}\) can be computed in \(O(3j-1)\). Therefore, \(EY^k\) by the AM can be computed in

\[
\sum_{j=1}^{k} O(3j) = O\left( \sum_{j=1}^{k} 3j \right) = O\left( 3 \sum_{j=1}^{k} j \right) = O\left( \frac{3}{2} k(k+1) \right) = O\left( \frac{3}{2} k^2 \right).
\]

Now let us investigate the memory size of the DM and the AM. For the DM, we only need to store a scalar, and thus the memory size of the DM is \(1\) double floating-point number. For the AM, we need to store two vectors of length \(k\), and a precomputed lower-triangular matrix \(S\) of size \(k \times k\). Therefore, the memory size of the AM is

\[
k + k + \frac{1}{2} k(k+1) = \frac{1}{2} k(k + 5)
\]
double floating-point numbers. In most cases, \(k\) would be small, and thus the memory size of the AM would be small.

In summary, compared to the DM, the AM would be computationally more efficient, but suffers from a heavier burden of memory size.

The numerical values of \(EY^k\) \((k = 1, \ldots, 10)\) and the computing times (in seconds, 10, 000 replications) by the two methods (DM and AM) for \(N = 1000, M = 500,\) and \(n = 10\) are summarized in Table 2.
other parameters fixed. Moreover, the two methods produce the same answers accurate to two decimal points. To obtain reliable computing times, we have done 10,000 replications, that is, we have computed each moment for 10,000 times, and the total computing times in seconds are reported in the table. Therefore, for each moment in a single replication, the computing time is \( \frac{1}{10,000} \) of the computing time reported in the table. For example, the computing time for \( EY_1 \) for once by the DM is \( \frac{1.71}{10,000} \) seconds = \( \frac{1.71}{10} \) milliseconds = 0.171 milliseconds.

Hence, the computing for each moment by either method is very fast. Finally, the computation of \( EY^k \) by the AM is consistently faster than that by the DM.

Now let us consider the absolute errors and the relative errors between the two methods. The absolute error between the two methods to compute \( EY^k \) is

\[ AE(EY^k) = |EY^k_{DM} - EY^k_{AM}|, \]

where \( EY^k_{DM} \) and \( EY^k_{AM} \) are the \( EY^k \) values computed by the DM and the AM, respectively. The relative error between the two methods to compute \( EY^k \) is

\[ RE(EY^k) = \frac{|EY^k_{DM} - EY^k_{AM}|}{EY^k_{AM}} = \frac{AE(EY^k)}{EY^k_{AM}}. \]

The absolute errors and the relative errors between the two methods to compute \( EY^k \) for \( k = 1, \ldots, 20 \) with \( N = 1000, M = 500 \), and \( n = 10 \) are reported in Table 3.
the table, we see that the absolute errors between the two methods to compute $EY^k$ are quite small except for $k = 18$ and 20, and the absolute errors are 1 and 32, respectively. However, since $EY^k$ for $k = 18$ and 20 are extremely large,
the relative errors between the two methods to compute \( E^Y_k \) are still acceptably small,

\[
\text{RE}(E^Y_{18}) = 2.98e - 16 \quad \text{and} \quad \text{RE}(E^Y_{20}) = 1.19e - 16.
\]

Now let us carry out some sensitivity analysis to compute \( E^Y_k \) by allowing one of the parameters \((k, N, M, \text{and } n)\) to change, holding other parameters fixed. The parameter values are fixed as

\[
k = 10, \quad N = 1000, \quad M = 500, \quad \text{and} \quad n = 10.
\]

Other parameter values can also be specified.

The sensitivity analysis of \( E^Y_k \) by the DM and the AM against the parameters \( k, N, M, \text{and } n \) are plotted in Figure 1. From the figure, we see that \( E^Y_k \) are increasing functions of \( k, M, \text{and } n \), and it is a decreasing function of \( N \). The two curves in the plots
are indistinguishable, which means that the two methods produce almost the same answers to $EY^k$. It is worth noting that $EY^k$ for $n \geq 385$ by the DM are $\infty$ (see the lower right plot), while the AM produces reasonable $EY^k$ values for $n \geq 385$. Therefore, the AM is more robust to extreme parameter values than the DM.

The sensitivity analysis of the absolute errors and the relative errors against the parameters $k$, $N$, $M$, and $n$ are plotted in Figure 2. The upper left plot is a graphical version of Table 3. From the upper left plot, we see that the absolute errors of $EY^{18}$ and $EY^{20}$ are large, however, their relative errors are acceptably small. Moreover, the absolute errors against $N$ and $M$ are acceptably small, and the relative errors against $N$ and $M$ are much smaller. Furthermore, the absolute errors against $n$ is very large, however, the relative errors against $n$ are acceptably small. All the plots indicate that the relative errors are acceptably small, and in fact, they are almost 0.

5. Conclusions and discussions

We have provided a noval method to calculate the high-order origin moments of the hypergeometric distribution $Y \sim HG(N,M,n)$, that is, the expectation identity method. First, the expectation identity of the hypergeometric distribution is discovered and summarized in Theorem 1. After that, we analytically calculate the first four origin moments of the hypergeometric distribution by using the expectation identity. Furthermore, we analytically calculate the general $k$th ($k = 1, 2, \ldots$) origin moment of the hypergeometric distribution by using the expectation identity, and the results are summarized in Theorem 2. Moreover, we use the general $k$th origin moment to validate the first four origin moments of the hypergeometric distribution. The coefficients of the first 10 origin moments of the hypergeometric distribution are shown in Table 1. Comparing Table 1 with the tables in the literature, we find that $A(k,j)$ in Theorem 2 are Stirling numbers of the second kind. Finally, Theorem 2 is restated by Theorem 3 in the symbol of $S(k,j)$, which are Stirling numbers of the second kind.

We have provided an assessment of the computation complexity which is given in number of multiplications of the DM and the AM, which are respectively $O(n(k + 4n))$ and $O(3k^2/2)$. Moreover, we have investigated the memory size of the DM and the AM, which are respectively 1 and $k(k + 5)/2$ double floating-point numbers. Furthermore, we have carried out some numerical experiments to compare the computations of $EY^k$ and the computing times by the two methods. Finally, we have compared the absolute errors and the relative errors between the two methods. In summary, compared to the DM, the AM would be computationally more efficient, but suffers from a heavier burden of memory size. Moreover, the AM is more robust to extreme parameter values than the DM, as observed from the lower right plot in Figure 1.

Supporting information

Additional information for this article is available.

The R codes used in the article can be downloaded from the link: https://pan.baidu.com/s/1SzEN1fHwnRvB8HzOdpoxg, with password: 1234. Alternatively, the readers can send emails to the corresponding author.
Acknowledgments

The authors are extremely grateful to the editor, the associate editor, and the reviewer for their insightful comments that led to significant improvement of the article.

Funding

The research was supported by the Ministry of Education (MOE) project of Humanities and Social Sciences on the west and the border area (20XJC910001), the National Social Science Fund of China (21XTJ001), the National Natural Science Foundation of China (72071019), and the Fundamental Research Funds for the Central Universities (2020CDJQY-Z001; 2021CDJQY-047).

ORCID

Ying-Ying Zhang http://orcid.org/0000-0002-6279-3662

References

Brunk, H. D., J. E. Holstein, and F. Williams. 1968. A comparison of binomial approximations to hypergeometric distribution. *The American Statistician* 22 (1):24–6. doi:10.2307/2681878.

Casella, G., and R. L. Berger. 2002. *Statistical inference*. 2nd ed. Duxbury, Pacific Grove.

Childs, A., and N. Balakrishnan. 2000. Some approximations to the multivariate hypergeometric distribution with applications to hypothesis testing. *Computational Statistics & Data Analysis* 35 (2):137–54. doi:10.1016/S0167-9473(00)00007-4.

Chvatal, V. 1979. Tail of the hypergeometric distribution. *Discrete Mathematics* 25 (3):285–7.

Comtet, L. 1974. *Advanced combinatorics*. Revised and enlarged ed. D. Reidel Publishing Company, Boston.

Danielian, E., R. Chitchyan, and D. Farbod. 2016. On a new regularly varying generalized hypergeometric distribution of the second type. *Mathematical Reports* 18 (2):217–32.

de Klerk, E., M. Laurent, and Z. Sun. 2015. An error analysis for polynomial optimization over the simplex based on the multivariate hypergeometric distribution. *SIAM Journal on Optimization* 25 (3):1498–514. doi:10.1137/140976650.

Dinwoodie, I. H., L. F. Matusevich, and E. Mosteig. 2004. Transform methods for the hypergeometric distribution. *Statistics and Computing* 14 (4):287–97. doi:10.1023/B:STCO.0000039478.96565.a5.

Eisinga, R., and B. Pelzer. 2011. Saddlepoint approximations to the mean and variance of the extended hypergeometric distribution. *Statistica Neerlandica* 65 (1):22–31. doi:10.1111/j.1467-9574.2010.00468.x.

Fog, A. 2008. Sampling methods for Wallenius’ and Fisher’s noncentral hypergeometric distributions. *Communications in Statistics – Simulation and Computation* 37 (2):241–57. doi:10.1080/03610910701790236.

Greene, E., and J. A. Wellner. 2017. Exponential bounds for the hypergeometric distribution. *Bernoulli: Official Journal of the Bernoulli Society for Mathematical Statistics and Probability* 23 (3):1911–50.

Guenther, W. C. 1978. Some remarks on runs test and use of hypergeometric distribution. *The American Statistician* 32 (2):71–3.

Hafid, A., A. S. Hafid, and M. Samih. 2020. A novel methodology-based joint hypergeometric distribution to analyze the security of shard ed blockchains. *IEEE Access*. 8:179389–99. doi:10.1109/ACCESS.2020.3027952.

Harkness, W. L. 1965. Properties of the extended hypergeometric distribution. *The Annals of Mathematical Statistics* 36 (3):938–45. doi:10.1214/aoms/1177700066.
Hida, E., and M. Akahira. 2003. An approximation to the generalized hypergeometric distribution. *Statistical Papers* 44 (4):483–97. doi:10.1007/BF02926006.

Hush, D., and C. Scovel. 2005. Concentration of the hypergeometric distribution. *Statistics & Probability Letters* 75 (2):127–32. doi:10.1016/j.spl.2005.05.019.

Hwang, J. T. 1982. Improving on standard estimators in discrete exponential families with applications to poisson and negative binomial cases. *The Annals of Statistics* 10:857–67.

Kachitvichyanukul, V., and B. W. Schmeiser. 1988. Algorithm 668 h2pec – sampling from the hypergeometric distribution. *ACM Transactions on Mathematical Software* 14 (4):397–8. doi:10.1145/50063.214387.

Kemp, C. D., and A. W. Kemp. 1956. Generalized hypergeometric distributions. *Journal of the Royal Statistical Society: Series B (Methodological)* 18 (2):202–11. doi:10.1111/j.2517-6161.1956.tb00224.x.

Kochman, F., A. Murray, and D. B. West. 1989. An entropy inequality for the bi-multivariate hypergeometric distribution. *Proceedings of the American Mathematical Society* 107 (2):479–85. doi:10.1090/S0002-9939-1989-0979050-8.

Krishnamoorthy, K., and S. Lv. 2020. Prediction intervals for hypergeometric distributions. *Communications in Statistics – Theory and Methods* 49 (6):1528–36. doi:10.1080/03610926.2018.1563181.

Kumar, C. S. 2002. Extended generalized hypergeometric probability distributions. *Statistics & Probability Letters* 59 (1):1–7. doi:10.1016/S0167-7152(02)00102-5.

Kumar, C. S. 2007. Some properties of bivariate generalized hypergeometric probability distributions. *Journal of the Korean Statistical Society* 36 (3):349–55.

Lahiri, S. N., A. Chatterjee, and T. Maiti. 2007. Normal approximation to the hypergeometric distribution in nonstandard cases and a sub-gaussian berry-esseen theorem. *Journal of Statistical Planning and Inference* 137 (11):3570–90. doi:10.1016/j.jspi.2007.03.033.

Lai, T. L. 1979. Sequential tests for hypergeometric distributions and finite populations. *Annals of Statistics* 7 (1):46–59.

Lebrun, R. 2013. Efficient time/space algorithm to compute rectangular probabilities of multinomial, multivariate hypergeometric and multivariate ptya distributions. *Statistics and Computing* 23 (5):615–23. doi:10.1007/s11222-012-9334-8.

Li, J. Q., and Q. J. Tian. 2010. Simple algorithm for the hypergeometric distribution moments about the origin of higher-order. *Science Technology and Engineering* 10:7989–92. In Chinese.

Lieberman, G. J., and D. B. Owen. 1962. Tables of the hypergeometric probability-distribution. *Psychometrika* 27 (3):321–2.

Ling, R. F., and J. W. Pratt. 1984. The accuracy of peizer approximations to the hypergeometric distribution, with comparisons to some other approximations. *Journal of the American Statistical Association* 79 (385):49–60. doi:10.1080/01621459.1984.10477061.

Lin, M. H., and C. A. Hsiung. 1994. Empirical bayes estimates of domain scores under binomial and hypergeometric distributions for test-scores. *Psychometrika* 59 (3):331–59. doi:10.1007/BF02296128.

Lin, M. H., C. A. Hsiung, and C. F. Hsiao. 1994. A program for monotonizing 2 empirical bayes estimators in binomial and hypergeometric data distributions. *Psychometrika* 59 (3):423–4. doi:10.1007/BF02296134.

Liu, C. B., and H. L. Wang. 2014. Researches on the high order moments of hypergeometric distribution. *Journal of Inner Mongolia Normal University (Natural Science Edition)* 43:23–5. (In Chinese).

Ma, C. S. 1999. Representations, bounds and approximations for tail probabilities of multivariate non-central hypergeometric and negative hypergeometric distributions. *Journal of Statistical Computation and Simulation* 62 (3):237–58. doi:10.1080/0094965908811934.

Mano, S. 2017. Partition structure and the a-hypergeometric distribution associated with the rational normal curve. *Electronic Journal of Statistics* 11 (2):4452–87. doi:10.1214/17-EJS1361.

Molenaar, W. 1973. Simple approximations to poisson, binomial, and hypergeometric distributions. *Biometrics* 29 (2):403–7. doi:10.2307/2529405.

Ord, J. K. 1968. Approximations to distribution functions which are hypergeometric series. *Biometrika* 55 (1):243–8. doi:10.1093/biomet/55.1.243.
Panaretos, J. 1983. An elementary characterization of the multinomial and the multivariate hypergeometric distributions. *Lecture Notes in Mathematics* 982:156–64.

Progri, I. F. 2011. *Geolocation of RF signals—Principles and simulations*. Springer, New York.

Progri, I. F. 2021a. Efficient computation of the generalized parabolic cylinder function distribution. *Journal of Geolocation, Geo-Information, and Geo-Intelligence* 2021:13–29. Article ID 2021071602.

Progri, I. F. 2021b. Landmark computation of the generalized parabolic cylinder function distribution. *Journal of Geolocation, Geo-Information, and Geo-Intelligence* 2021:30–44. Article ID 2021071603.

Progri, I. F. 2021c. Special cases of the generalized parabolic cylinder function distribution. *Journal of Geolocation, Geo-Information, and Geo-Intelligence* 2021:45–56. Article ID 2021071604.

Riordan, J. 1937. Moment recurrence relations for binomial, poisson and hypergeometric frequency distributions. *The Annals of Mathematical Statistics* 8 (2):103–11. doi:10.1214/aoms/1177732430.

Sibuya, M., and R. Shimizu. 1981. The generalized hypergeometric family of distributions. *Annals of the Institute of Statistical Mathematics* 33 (2):177–90. doi:10.1007/BF02480931.

Skibinsky, M. 1970. A characterization of hypergeometric distributions. *Journal of the American Statistical Association* 65 (330):926–9. doi:10.1080/01621459.1970.10481135.

Takayama, N., S. Kuriki, and A. Takemura. 2018. A-hypergeometric distributions and newton polytopes. *Advances in Applied Mathematics* 99:109–33. doi:10.1016/j.aam.2018.05.001.

Themangani, R., S. Porwal, and N. Magesh. 2020. Generalized hypergeometric distribution and its applications on univalent functions. *Journal of Inequalities and Applications* 2020 (249):1–10. doi:10.1186/s13660-020-02515-5.

Tohma, Y., H. Yamano, M. Ohba, and R. Jacoby. 1991. The estimation of parameters of the hypergeometric distribution and its application to the software-reliability growth-model. *IEEE Transactions on Software Engineering* 17 (5):483–9. doi:10.1109/32.90450.

Tripathi, R. C., and J. Gurland. 1977. General family of discrete distributions with hypergeometric probabilities. *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (3):349–56. doi:10.1111/j.2517-6161.1977.tb01364.x.

Wu, T. 1993. An accurate computation of the hypergeometric distribution function. *ACM Transactions on Mathematical Software* 19 (1):33–43. doi:10.1145/151271.151274.

Zhang, Y. Y., T. Z. Rong, and M. M. Li. 2019. Expectation identity for the binomial distribution and its application in the calculations of high-order binomial moments. *Communications in Statistics – Theory and Methods* 48 (22):5467–76. doi:10.1080/03610926.2018.1435818.