Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2020) 44: 1881 – 1898
© TÜBİTAK
doi:10.3906/mat-1907-121

Hypergeometric distribution of the number of draws from an urn with two types of items before one of the counts reaches a threshold

Juan Luis GONZÁLEZ-SANTANDER*
Department of Mathematics, Universidad de Oviedo, Oviedo, Asturias, Spain

Received: 30.07.2019 • **Accepted/Published Online:** 31.07.2020 • **Final Version:** 21.09.2020

Abstract: We consider an urn with R elements of one type and B elements of other type. We calculate the probability distribution $P_{n_R,n_B}^{R,B}(s)$ wherein the random variable s is the number of draws from the urn until we reach n_R elements of type R or n_B elements of type B. We calculate the mean value $\langle s\rangle$ and the standard deviation σ of $P_{n_R,n_B}^{R,B}(s)$ in terms of hypergeometric functions. For $n_R = n_B$ and $B = R$, we reduce $\langle s\rangle$ and σ in terms of elementary functions. Also, the normalization condition leads to a new hypergeometric summation formula involving ${}_3F_2$ terminating series with unity argument. For $n_R = n_B$, we provide an alternative proof of this summation formula using q-hypergeometric functions. As a consistency test, computer simulations have been performed to confirm the analytical results obtained.

Key words: Generalized hypergeometric function, hypergeometric probability distribution, q-hypergeometric function

1. Introduction

In probability theory, the hypergeometric distribution describes the probability to obtain n successes in q draws without replacement from a finite population of size $B + R$, where, say, B is the number of failures in the population and R the number of successes. If X is a random variable following the hypergeometric distribution, then the probability mass function is given by [1, Sect. 2.6]

$$P_q(X = n) = \frac{\binom{R}{n}\binom{B}{q-n}}{\binom{B+R}{q}}. \tag{1.1}$$

However, we can define other type of hypergeometric probability distribution $P_{n_R,n_B}^{R,B}(s)$ wherein the random variable is the number of draws s without replacement that first obtain n_R items of type R or n_B items of type B, where $n_R \le R$ and/or $n_B \le B$. As an application of this probability distribution, suppose a box that contains a mixture of two types of items, say R and B of each type. The goal is to separate the items into separate bins. An algorithm to do this is as follows: a machine picks items randomly from the bin until the items are fully sorted. The running time of this algorithm has exactly the probability distribution $P_{n_R,n_B}^{R,B}(s)$ if we choose $n_R = R$ and $n_B = B$.

On the one hand, in this article, we derive the expression of the probability distribution $P_{n_R,n_B}^{R,B}(s)$, as well as its first moments. In particular, we calculate the mean value $\langle s\rangle$ and the standard deviation

*Correspondence: gonzalezmarjuan@uniovi.es
2010 *AMS Mathematics Subject Classification:* 33C20, 33D15, 33C90

$\sigma = \sqrt{\langle s^2\rangle - \langle s\rangle^2}$, from which confidence intervals can be computed. On the other hand, the normalization condition of $P_{n_R,n_B}^{R,B}(s)$ provides an hypergeometric summation formula that does not seem to be reported in the literature. However, in the particular case of $n_R = n_B$, this summation formula can be derived in a non-trivial way from the theory of q-hypergeometric functions. In fact, the connection between special functions and probability theory is well-known. For instance, the gamma probability distribution is applied to model the time between occurrence of earthquakes [2]. Also, probability theory can be used to derive some identities of special functions, as in the case of the beta function [3, Sect. 1.11].

This paper is organized as follows. Section 2 derives the expression of the modified hypergeometric probability distribution $P_{n_R,n_B}^{R,B}(s)$ stated above. Section 3 introduces the notation and the common features of the calculation of the m-th moment $\langle s^m\rangle$. Next, in Sections 4–6 we specifically calculate $\langle s^0\rangle$, $\langle s\rangle$ and σ in terms of hypergeometric functions. As aforementioned, we derive a new summation formula involving ${}_3F_2$ terminating hypergeometric series from the normalization condition $\langle s^0\rangle = 1$. In the particular case $n_R = n_B$ and $R = B$, we calculate $\langle s\rangle$ and σ in terms of elementary functions. Also, we calculate $\langle s\rangle$ and σ in terms of elementary functions for the case $\max(n_R, n_B) \leq \max(B, R)$ and $\min(n_R, n_B) > \min(B, R)$. Section 7 presents some computer simulations to confirm the analytical results given in Sections 4–6. Finally, the conclusions are summarized in Section 8.

2. Modified hypergeometric probability distribution

First, consider the probability distribution $P_{n_R,n_B}^{R,B}(s)$ for which $n = n_R = n_B$ (denoted by $P_n^{R,B}(s)$). If we take $n = 2$, then the minimum number of extractions will be $s_{\min} = 2$ and the probability $P_n^{R,B}(s)$ for this case is

$$P_2^{R,B}(s=2) = \frac{R}{B+R}\frac{R-1}{B+R-1} + \frac{B}{B+R}\frac{B-1}{B+R-1},$$

i.e. the probability to obtain two consecutive draws from R or two consecutive draws from B. Notice that the maximum number of extractions for $n = 2$ is $s_{\max} = 3$. In this latter case,

$$\begin{aligned} P_2^{R,B}(s=3) &= PR_2^{1,1}\left(\frac{R}{B+R}\frac{R-1}{B+R-1}\frac{B}{B+R-2}\right. \\ &\quad \left. + \frac{B}{B+R}\frac{B-1}{B+R-1}\frac{R}{B+R-2}\right), \end{aligned}$$

where $PR_2^{1,1} = \frac{2!}{1!1!}$ indicates the number of permutations with repetition. In general, $s_{\min} = n$ and $s_{\max} = 2n-1$, and the probability function is given by

$$\begin{aligned} P_n^{R,B}(s) &= \frac{(s-1)!}{(n-1)!\,(s-n)!} \\ &\quad \frac{\prod_{i=0}^{n-1}(R-i)\prod_{j=0}^{s-n-1}(B-j) + \prod_{i=0}^{n-1}(B-i)\prod_{j=0}^{s-n-1}(R-j)}{\prod_{k=0}^{s-1}(B+R-k)}, \end{aligned} \tag{2.1}$$

where we have assumed that $n \leq \min(B, R)$, i.e. the threshold n is not greater than the population of B or R. In the general case, we have:

Theorem 2.1

$$P_{n_R,n_B}^{R,B}(s) = \frac{(s-1)!\,(B+R-s)!R!B!}{(B+R)!} \left(\frac{1}{(n_R-1)!\,(s-n_R)!\,(R-n_R)!\,(B-s+n_R)!} + \frac{1}{(n_B-1)!\,(s-n_B)!\,(B-n_B)!\,(R-s+n_B)!}\right), \quad n_R \leq R,\ n_B \leq B, \tag{2.2}$$

where $s_{\min} = \min(n_R, n_B)$ *and* $s_{\max} = n_R + n_B - 1$.

Proof From (2.1), the general case reads as:

$$P_{n_R,n_B}^{R,B}(s) = \frac{(s-1)!}{(n_R-1)!\,(s-n_R)!} \frac{\prod_{i=0}^{n_R-1}(R-i) \prod_{j=0}^{s-n_R-1}(B-j)}{\prod_{k=0}^{s-1}(B+R-k)} + \frac{(s-1)!}{(n_B-1)!\,(s-n_B)!} \frac{\prod_{i=0}^{n_B-1}(B-i) \prod_{j=0}^{s-n_B-1}(R-j)}{\prod_{k=0}^{s-1}(B+R-k)} \tag{2.3}$$

which can be rewritten as (2.2). □

Notice that if the thresholds are greater than the respective populations, according to the statement of the experiment, the probability must be zero. This is so due to the factors $(R-n_R)!$ and $(B-n_B)!$ in both denominators of (2.2). Therefore, we have the following corollary:

Corollary 2.2 *If* $n_R > R$ *and* $n_B > B$,

$$P_{n_R,n_B}^{R,B}(s) = 0. \tag{2.4}$$

However, if we denote $n_{\max} = \max(n_R, n_B)$ and $n_{\min} = \min(n_R, n_B)$, there is an intermediate case, for which we have:

Corollary 2.3

$$P_{n_R,n_B}^{R,B}(s) = \frac{(s-1)!}{(n_{\max}-1)!\,(s-n_{\max})!} \frac{(B+R-s)!}{(B+R)!} \frac{R!B!}{(\max(B,R)-n_{\max})!\,(\min(B,R)-s+n_{\max})!}, \quad \min(B,R) < n_{\min} \leq n_{\max} \leq \max(B,R). \tag{2.5}$$

3. Preliminaries for the calculation of the moments

The definition of the following functions will be useful in the calculation of the moments.

Definition 3.1

$$S_m(X,Y,\nu) = \frac{1}{(X-\nu)!\,(\nu-1)!}\sum_{s=\nu}^{\nu+\bar{\nu}-1}\frac{(X+Y-s)!\,(s-1)!s^m}{(Y-s+\nu)!\,(s-\nu)!}, \tag{3.1}$$

where $\nu = n_R, n_B \to \bar{\nu} = n_B, n_R$, *thus* $\nu + \bar{\nu} = n_R + n_B$.

Definition 3.2

$$S_m^{(a)}(X,Y,\nu) = \sum_{t=0}^{\infty}\frac{(X+Y-t-\nu)!\,(t+\nu)!\,(t+\nu)^{m-1}}{t!\,(Y-t)!}, \tag{3.2}$$

$$S_m^{(b)}(X,Y,\nu) = \sum_{t=0}^{\infty}\frac{(X+Y-t-\nu-\bar{\nu})!\,(t+\nu+\bar{\nu})!\,(t+\nu+\bar{\nu})^{m-1}}{(t+\bar{\nu})!\,(Y-t-\bar{\nu})!}. \tag{3.3}$$

3.1. Case $n_R \le R,\ n_B \le B$

According to (2.2), the m-th moment is given by

$$\begin{aligned}\langle s^m\rangle &= \sum_{s=s_{\min}}^{s_{\max}} s^m P_{n_R,n_B}^{R,B}(s) \\ &= \frac{R!B!}{(B+R)!}\left[\frac{1}{(R-n_R)!\,(n_R-1)!}\sum_{s=s_{\min}}^{s_{\max}}\frac{(B+R-s)!\,(s-1)!s^m}{(B-s+n_R)!\,(s-n_R)!}\right. \\ &\quad \left.+\frac{1}{(B-n_B)!\,(n_B-1)!}\sum_{s=s_{\min}}^{s_{\max}}\frac{(B+R-s)!\,(s-1)!s^m}{(R-s+n_B)!\,(s-n_B)!}\right],\end{aligned} \tag{3.4}$$

where, as aforementioned, $s_{\min} = \min(n_R, n_B)$ and $s_{\max} = n_R + n_B - 1$. Recall (3.1) to rewrite (3.4) as,

$$\langle s^m\rangle = \frac{R!B!}{(B+R)!}\left[S_m(R,B,n_R) + S_m(B,R,n_B)\right], \tag{3.5}$$

$$n_R \le R,\ n_B \le B.$$

Perform now the change of variables $t = s - \nu$ in (3.1), and split the result in two sums, according to the definitions given in (3.2) and (3.3), to obtain,

$$S_m(X,Y,\nu) = \frac{1}{(X-\nu)!\,(\nu-1)!}\left[S_m^{(a)}(X,Y,\nu) - S_m^{(b)}(X,Y,\nu)\right], \tag{3.6}$$

3.2. Case $\min(B,R) < n_{\min} \le n_{\max} \le \max(B,R)$

If we take the probability distribution (2.5), the m-th moment is given by

$$\langle s^m\rangle = \frac{R!B!}{(B+R)!}S_m(\max(R,B), \min(R,B), n_{\max}).$$

Notice that according to (3.3), $S_m^{(b)}(\max(R,B), \min(R,B), n_{\max}) = 0$ in this case, thus

$$\langle s^m\rangle = \frac{R!B!\ S_m^{(a)}(\max(R,B), \min(R,B), n_{\max})}{(B+R)!\,(\max(R,B)-n_{\max})!\,(n_{\max}-1)!}. \tag{3.7}$$

4. Zeroth moment

4.1. Case $n_R \leq R,\ n_B \leq B$

For $m=0$, (3.2) reduces to

$$\begin{aligned} S_0^{(a)}(X,Y,\nu) &= \sum_{t=0}^{\infty} \frac{(X+Y-t-\nu)!\,(t+\nu-1)!}{t!\,(Y-t)!} \\ &= \frac{(X+Y-\nu)!\,(\nu-1)!}{Y!}\,{}_2F_1\left(\left.\begin{array}{c}-Y,\nu\\ \nu-X-Y\end{array}\right|1\right), \end{aligned} \tag{4.1}$$

where we have rewritten the sum as an hypergeometric function according to [3, Sect. 2.1]. Applying Chu-Vandermonde summation formula [3, Corollary 2.2.3]

$${}_2F_1\left(\left.\begin{array}{c}-k,a\\ c\end{array}\right|1\right) = \frac{(c-a)_k}{(c)_k}, \quad k=0,1,2,\ldots \tag{4.2}$$

and the following properties of the Pochhammer symbol [4, Eqs. 18:5:1 and 18:3:3]:

$$(-x)_k = (-1)^k (x-k+1)_k, \tag{4.3}$$

and

$$(x)_k = k!\binom{x+k-1}{k}, \tag{4.4}$$

(4.1) is reduced to

$$S_0^{(a)}(X,Y,\nu) = \frac{(\nu-1)!\,(X-\nu)!\,(X+Y)!}{X!Y!}. \tag{4.5}$$

On the other hand, $S_0^{(b)}(X,Y,\nu)$ can be expressed also as a hypergeometric function:

$$\begin{aligned} &S_0^{(b)}(X,Y,\nu) \\ &= \frac{(X+Y-\nu-\bar{\nu})!\,(\nu+\bar{\nu}-1)!}{\bar{\nu}!\,(Y-\bar{\nu})!}\,{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu},1\\ \bar{\nu}+1,\nu+\bar{\nu}-X-Y\end{array}\right|1\right). \end{aligned} \tag{4.6}$$

Inserting (4.5) and (4.6) in (3.6) for $m=0$, we arrive at

$$\begin{aligned} &S_0(X,Y,\nu) \\ &= \frac{1}{(X-\nu)!\,(\nu-1)!}\left[S_0^{(a)}(X,Y,\nu) - S_0^{(b)}(X,Y,\nu)\right] \\ &= \frac{(X+Y)!}{X!Y!}\left[1 - \frac{\nu\binom{X}{\nu}\binom{Y}{\bar{\nu}}}{(\nu+\bar{\nu})\binom{X+Y}{\nu+\bar{\nu}}}\,{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu},1\\ \bar{\nu}+1,\nu+\bar{\nu}-X-Y\end{array}\right|1\right)\right]. \end{aligned} \tag{4.7}$$

Finally, take $m=0$ in (3.5), hence the normalization condition becomes

$$1 = \left\langle s^0 \right\rangle = \frac{R!B!}{(B+R)!}\left[S_0(R,B,n_R) + S_0(B,R,n_B)\right]. \tag{4.8}$$

Inserting in (4.8) the result given in (4.7), we obtain the following summation formula:

Theorem 4.1 *For* $n_R \le R,\ n_B \le B$

$$\begin{aligned}
&(n_R+n_B)\frac{\binom{B+R}{n_R+n_B}}{\binom{R}{n_R}\binom{B}{n_B}} \qquad (4.9)\\
=\quad & n_R\, {}_3F_2\left(\begin{array}{c} n_B-B, n_R+n_B, 1 \\ n_B+1, n_R+n_B-R-B \end{array}\middle| 1\right) \\
&+n_B\, {}_3F_2\left(\begin{array}{c} n_R-R, n_R+n_B, 1 \\ n_R+1, n_R+n_B-R-B \end{array}\middle| 1\right).
\end{aligned}$$

The summation formula (4.9) is a kind of reflection formula since it is invariant against the exchange of n_R for n_B and R for B. Also, it does not seem to be reported in the most common literature. However, for $n = n_R = n_B$, (4.9) reduces to a summation formula that can be proved by using the theory of q-hypergeometric functions.

Theorem 4.2 *For* $n \le \min(B, R)$

$$\begin{aligned}
&\frac{2\binom{B+R}{2n}}{\binom{B}{n}\binom{R}{n}} \qquad (4.10)\\
=\quad & {}_3F_2\left(\begin{array}{c} n-B, 2n, 1 \\ 2n-B-R, n+1 \end{array}\middle| 1\right) + {}_3F_2\left(\begin{array}{c} n-R, 2n, 1 \\ 2n-B-R, n+1 \end{array}\middle| 1\right).
\end{aligned}$$

Proof Consider the following three-term transformation formula [5, Eq. 3.3.2]

$$\begin{aligned}
&{}_3\varphi_2\left(\begin{array}{c} q^{-k}, b, c \\ d, e \end{array}\middle| q, \frac{deq^k}{bc}\right) \\
=\quad & \frac{(e/b, e/c; q)_\infty}{(e, e/(bc); q)_\infty}\, {}_3\varphi_2\left(\begin{array}{c} dq^k, b, c \\ d, bcq/e \end{array}\middle| q, q\right) \\
&+\frac{(b, c; q)_\infty}{(e, bc/e; q)_\infty}\frac{(de/(bc); q)_k}{(d; q)_k}\, {}_3\varphi_2\left(\begin{array}{c} e/b, e/c, deq^k/(bc) \\ de/(bc), eq/(bc) \end{array}\middle| q, q\right),
\end{aligned}$$

where k is a nonnegative integer. Perform the following substitutions: $k = B - n$ (notice that $n \le \min(B, R)$, thus k is a nonnegative integer), $c = q$, and $b = e^2/q^2$, to obtain

$$\begin{aligned}
&{}_3\varphi_2\left(\begin{array}{c} q^{n-B}, e^2/q^2, q \\ d, e \end{array}\middle| q, \frac{dq^{B-n+1}}{e}\right) \qquad (4.11)\\
=\quad & \frac{(q^2/e, e/q; q)_\infty}{(e, q/e; q)_\infty}\, {}_3\varphi_2\left(\begin{array}{c} dq^{B-n}, e^2/q^2, q \\ d, e \end{array}\middle| q, q\right) \\
&+\frac{(e^2/q^2, q; q)_\infty}{(e, e/q; q)_\infty}\frac{(dq/e; q)_{B-n}}{(d; q)_{B-n}}\, {}_2\varphi_1\left(\begin{array}{c} e/q, dq^{B-n+1}/e \\ dq/e \end{array}\middle| q, q\right),
\end{aligned}$$

where the last ${}_3\varphi_2$ series reduces to a ${}_2\varphi_1$ series due to cancelling parameters. By using the property given in [5, Eqn. 1.2.30], we calculate the prefactor of ${}_3\varphi_2$ series of the RHS of (4.11) as

$$\frac{(q^2/e, e/q; q)_\infty}{(e, q/e; q)_\infty} = -\frac{e}{q},$$

hence

$$
\begin{aligned}
& {}_3\varphi_2\left(\left.\begin{array}{c} q^{n-B}, e^2/q^2, q \\ d, e \end{array}\right| q, \frac{dq^{B-n+1}}{e}\right) + \frac{e}{q}\, {}_3\varphi_2\left(\left.\begin{array}{c} dq^{B-n}, e^2/q^2, q \\ d, e \end{array}\right| q, q\right) \\
= \;& \frac{\left(e^2/q^2, q; q\right)_\infty}{\left(e, e/q; q\right)_\infty} \frac{\left(dq/e; q\right)_{B-n}}{\left(d; q\right)_{B-n}}\, {}_2\varphi_1\left(\left.\begin{array}{c} e/q, dq^{B-n+1}/e \\ dq/e \end{array}\right| q, q\right).
\end{aligned}
$$

Now, performing the substitutions $d = q^{2n-B-R}$ and $e = q^{n+1}$, we obtain

$$
\begin{aligned}
& {}_3\varphi_2\left(\left.\begin{array}{c} q^{n-B}, q^{2n}, q \\ q^{2n-B-R}, q^{n+1} \end{array}\right| q, q^{-R}\right) + q^n\, {}_3\varphi_2\left(\left.\begin{array}{c} q^{n-R}, q^{2n}, q \\ q^{2n-B-R}, q^{n+1} \end{array}\right| q, q\right) \\
= \;& \frac{\left(q^{2n}, q; q\right)_\infty}{\left(q^{n+1}, q^n; q\right)_\infty} \frac{\left(q^{n-B-R}; q\right)_{B-n}}{\left(q^{2n-B-R}; q\right)_{B-n}}\, {}_2\varphi_1\left(\left.\begin{array}{c} q^{-R}, q^n \\ q^{n-B-R} \end{array}\right| q, q\right).
\end{aligned} \tag{4.12}
$$

Applying the properties given in [5, Eq. 1.2.30] and [6, Eq. 17.2.17], we calculate the following prefactor of the RHS of (4.12) as

$$
\frac{\left(q^{2n}, q; q\right)_\infty}{\left(q^{n+1}, q^n; q\right)_\infty} = \frac{(q; q)_n}{(q^n; q)_n} = \frac{(q; q)_n\, (q; q)_{n-1}}{(q; q)_{2n-1}}. \tag{4.13}
$$

Also, the ${}_2\varphi_1$ series can be calculated applying the q-Chu-Vandermonde summation formula [5, Eq. 1.5.3], thus

$$
{}_2\varphi_1\left(\left.\begin{array}{c} q^{-R}, q^n \\ q^{n-B-R} \end{array}\right| q, q\right) = \frac{\left(q^{-B-R}; q\right)_R}{\left(q^{n-B-R}; q\right)_R} q^{nR}. \tag{4.14}
$$

Therefore, inserting (4.13) and (4.14) in the RHS of (4.12), we get

$$
\begin{aligned}
& {}_3\varphi_2\left(\left.\begin{array}{c} q^{n-B}, q^{2n}, q \\ q^{2n-B-R}, q^{n+1} \end{array}\right| q, q^{-R}\right) + q^n\, {}_3\varphi_2\left(\left.\begin{array}{c} q^{n-R}, q^{2n}, q \\ q^{2n-B-R}, q^{n+1} \end{array}\right| q, q\right) \\
= \;& \frac{(q; q)_n\, (q; q)_{n-1}}{(q; q)_{2n-1}} \frac{\left(q^{n-B-R}; q\right)_{B-n}}{\left(q^{2n-B-R}; q\right)_{B-n}} \frac{\left(q^{-B-R}; q\right)_R}{\left(q^{n-B-R}; q\right)_R} q^{nR}.
\end{aligned} \tag{4.15}
$$

We can rewrite the RHS of (4.15) in terms of q-binomial coefficients by using the property given in [6, Eq. 17.2.27], thus

$$
\frac{\left(q^{n-B-R}; q\right)_{B-n}}{\left(q^{2n-B-R}; q\right)_{B-n}} = \frac{\binom{B+R-n}{B-n}_q}{\binom{B+R-2n}{B-n}_q} q^{n(n-B)}, \tag{4.16}
$$

and

$$
\frac{\left(q^{-B-R}; q\right)_R}{\left(q^{n-B-R}; q\right)_R} = \frac{\binom{B+R}{R}_q}{\binom{B+R-n}{R}_q} q^{-nR}. \tag{4.17}
$$

Substituting in the RHS of (4.15) the results (4.16) and (4.17), and expanding the q-binomial coefficients, after some simplification, we arrive at

$$\begin{aligned} & \frac{(q;q)_n\,(q;q)_{n-1}}{(q;q)_{2n-1}}\frac{(q;q)_{R-n}\,(q;q)_{B-n}}{(q;q)_{B+R-2n}}\frac{(q;q)_{B+R}}{(q;q)_R\,(q;q)_B}q^{n(n-B)} \\ = & \frac{\binom{B+R}{2n}_q}{\binom{R}{n}_q\binom{B}{n}_q}\frac{(q;q)_{2n}}{(q;q)_{2n-1}}\frac{(q;q)_{n-1}}{(q;q)_n}q^{n(n-B)}. \end{aligned} \tag{4.18}$$

According to the definition of the q-shifted factorial [6, Eq. 17.2.1], we have

$$\frac{(q;q)_k}{(q;q)_{k-1}}=1-q^k,$$

thus (4.18) is reduced to

$$\frac{\binom{B+R}{2n}_q}{\binom{R}{n}_q\binom{B}{n}_q}\left(1+q^n\right)q^{n(n-B)}. \tag{4.19}$$

Substituting now (4.19) in the RHS of (4.15), we obtain the following q-analogue of (4.10),

$$\begin{aligned} & {}_3\varphi_2\left(\begin{array}{c} q^{n-B},q^{2n},q \\ q^{2n-B-R},q^{n+1} \end{array}\middle| q,q^{-R}\right)+q^n\,{}_3\varphi_2\left(\begin{array}{c} q^{n-R},q^{2n},q \\ q^{2n-B-R},q^{n+1} \end{array}\middle| q,q\right) \\ = & \frac{\binom{B+R}{2n}_q}{\binom{R}{n}_q\binom{B}{n}_q}\left(1+q^n\right)q^{n(n-B)}. \end{aligned} \tag{4.20}$$

Therefore, performing in (4.20) the limit $q\to 1^-$, taking into account [6, Eqs. 17.4.2 and 17.2.28], we finally obtain (4.10), as we wanted to prove. □

Corollary 4.3 *Take $B=R=M$ in (4.10), to obtain this nice formula*

$$\begin{aligned} & {}_3F_2\left(\begin{array}{c} n-M,2n,1 \\ 2n-2M,n+1 \end{array}\middle| 1\right)=\frac{\binom{2M}{2n}}{\binom{M}{n}^2}, \\ & n\leq M. \end{aligned} \tag{4.21}$$

Here, we provide a proof of (4.21), considering a result reported in the literature for a terminating ${}_3F_2$ hypergeometric sum.

Proof In [7, Eq. 7.4.4(101)], we find the following equivalent result to (4.21),

$$\begin{aligned} & {}_3F_2\left(\begin{array}{c} -k,a,b \\ -2k,\frac{a+b+1}{2} \end{array}\middle| 1\right)=\frac{\left(\frac{a+1}{2}\right)_k\left(\frac{b+1}{2}\right)_k}{\left(\frac{1}{2}\right)_k\left(\frac{a+b+1}{2}\right)_k}, \\ & k=0,1,2,\ldots \end{aligned} \tag{4.22}$$

taking $-k=n-M\leq 0$, $a=2n$, $b=1$, and applying the formula of the gamma function for half-integral argument [4, Eq. 43:4:3]. □

We will apply the identity (4.22) in the next sections to calculate the mean value and the standard deviation of the modified hypergeometric probability distribution $P_{n_R,n_B}^{R,B}(s)$.

4.2. Case $\min(B,R) < n_{\min} \le n_{\max} \le \max(B,R)$

According to (3.7), we have

$$\left\langle s^0 \right\rangle = \frac{R!B!\ S_0^{(a)}\left(\max(R,B), \min(R,B), n_{\max}\right)}{(B+R)!\left(\max(R,B)-n_{\max}\right)!\left(n_{\max}-1\right)!}. \tag{4.23}$$

Note that applying (4.5) to (4.23), the normalization condition is automatically satisfied:

$$\left\langle s^0 \right\rangle = 1.$$

5. First moment

5.1. Case $n_R \le R,\ n_B \le B$

For the calculation of the mean value $\langle s \rangle$, take $m = 1$ in (3.2) and rewrite the obtained sum as a hypergeometric function:

$$S_1^{(a)}(X,Y,\nu) = \frac{(X+Y-\nu)!\nu!}{Y!}\ {}_2F_1\left(\left. \begin{array}{c} -Y, \nu \\ \nu - X - Y \end{array} \right| 1 \right). \tag{5.1}$$

Applying again (4.2)–(4.4), we calculate the hypergeometric sum given in (5.1). Thereby, we obtain

$$S_1^{(a)}(X,Y,\nu) = \frac{\nu!\,(X-\nu)!\,(X+Y+1)!}{Y!\,(X+1)!}. \tag{5.2}$$

Similarly, we calculate

$$\begin{aligned} & S_1^{(b)}(X,Y,\nu) \\ = & \frac{(X+Y-\nu-\bar{\nu})!\,(\nu+\bar{\nu})!}{\bar{\nu}!\,(Y-\bar{\nu})!}\ {}_3F_2\left(\left. \begin{array}{c} \bar{\nu}-Y, \nu+\bar{\nu}+1, 1 \\ \bar{\nu}+1, \nu+\bar{\nu}-X-Y \end{array} \right| 1 \right). \end{aligned} \tag{5.3}$$

Therefore, inserting (5.2) and (5.3) in (3.6) for $m = 1$, we arrive at

$$\begin{aligned} & S_1(X,Y,\nu) \\ = & \frac{1}{(X-\nu)!\,(\nu-1)!}\left[S_1^{(a)}(X,Y,\nu) - S_1^{(b)}(X,Y,\nu)\right] \\ = & \frac{\nu\,(X+Y)!}{X!Y!}\left[\frac{X+Y+1}{X+1} - \frac{\binom{X}{\nu}\binom{Y}{\bar{\nu}}}{\binom{X+Y}{\nu+\bar{\nu}}}\ {}_3F_2\left(\left. \begin{array}{c} \bar{\nu}-Y, \nu+\bar{\nu}+1, 1 \\ \bar{\nu}+1, \nu+\bar{\nu}-X-Y \end{array} \right| 1 \right)\right]. \end{aligned} \tag{5.4}$$

Finally, taking into account (5.4) in (3.5) for $m = 1$, and simplifying, the mean value results in

$$\begin{aligned} & \langle s \rangle \\ = & \frac{R!B!}{(B+R)!}\left[S_1(R,B,n_R) + S_1(B,R,n_B)\right] \\ = & (R+B+1)\left(\frac{n_R}{R+1} + \frac{n_B}{B+1}\right) - \frac{\binom{R}{n_R}\binom{B}{n_B}}{\binom{R+B}{n_R+n_B}} \\ & \left[n_R\,{}_3F_2\left(\left. \begin{array}{c} n_B - B, n_R+n_B+1, 1 \\ n_B+1, n_R+n_B-R-B \end{array} \right| 1 \right) + n_B\,{}_3F_2\left(\left. \begin{array}{c} n_R - R, n_R+n_B+1, 1 \\ n_R+1, n_R+n_B-R-B \end{array} \right| 1 \right)\right], \end{aligned} \tag{5.5}$$

For $n = n_R = n_B$ and $M = R = B$, (5.5) is reduced to

$$\langle s \rangle = 2n \left[\frac{2M+1}{M+1} - \frac{\binom{M}{n}^2}{\binom{2M}{2n}} \, {}_3F_2 \left(\begin{array}{c} n-M, 2n+1, 1 \\ 2n-2M, n+1 \end{array} \middle| 1 \right) \right]. \tag{5.6}$$

To calculate the hypergeometric function given in (5.6) we need the following result:

Theorem 5.1 *If k is a nonnegative integer, then*

$${}_3F_2 \left(\begin{array}{c} -k, a, b \\ -2k, \frac{a+b}{2} \end{array} \middle| 1 \right) = \frac{a \left(\frac{a}{2}+1\right)_k \left(\frac{b+1}{2}\right)_k + b \left(\frac{a+1}{2}\right)_k \left(\frac{b}{2}+1\right)_k}{(a+b) \left(\frac{1}{2}\right)_k \left(\frac{a+b}{2}+1\right)_k}. \tag{5.7}$$

Proof From the formulas given in [3, Sect. 3.7], we can write

$$\begin{aligned} & {}_3F_2 \left(\begin{array}{c} -k, a+1, b \\ -2k, \frac{a+b+1}{2} \end{array} \middle| 1 \right) \\ = & \; {}_3F_2 \left(\begin{array}{c} -k, a, b \\ -2k, \frac{a+b+1}{2} \end{array} \middle| 1 \right) + \frac{b}{a+b+1} \, {}_3F_2 \left(\begin{array}{c} -k+1, a+1, b+1 \\ -2k+1, \frac{a+b+3}{2} \end{array} \middle| 1 \right), \end{aligned} \tag{5.8}$$

and

$$\begin{aligned} & {}_3F_2 \left(\begin{array}{c} -k+1, a+1, b+1 \\ -2k+1, \frac{a+b+3}{2} \end{array} \middle| 1 \right) \\ = & \; \frac{(a+1)(b+1)}{(1-2k)(a+b+3)} \, {}_3F_2 \left(\begin{array}{c} -(k-1), a+2, b+2 \\ -2(k-1), \frac{a+b+5}{2} \end{array} \middle| 1 \right) \\ & + {}_3F_2 \left(\begin{array}{c} -k, a+1, b+1 \\ -2k, \frac{a+b+3}{2} \end{array} \middle| 1 \right). \end{aligned} \tag{5.9}$$

Therefore, inserting (5.9) in (5.8), and applying the formula given in (4.22), we obtain

$$\begin{aligned} {}_3F_2 \left(\begin{array}{c} -k, a+1, b \\ -2k, \frac{a+b+1}{2} \end{array} \middle| 1 \right) = & \; \frac{\left(\frac{a+1}{2}\right)_k \left(\frac{b+1}{2}\right)_k}{\left(\frac{1}{2}\right)_k \left(\frac{a+b+1}{2}\right)_k} \\ + \frac{b}{a+b+1} & \left\{ \frac{(a+1)(b+1)}{(1-2k)(a+b+3)} \frac{\left(\frac{a+3}{2}\right)_{k-1} \left(\frac{b+3}{2}\right)_{k-1}}{\left(\frac{1}{2}\right)_{k-1} \left(\frac{a+b+5}{2}\right)_{k-1}} + \frac{\left(\frac{a}{2}+1\right)_k \left(\frac{b}{2}+1\right)_k}{\left(\frac{1}{2}\right)_k \left(\frac{a+b+3}{2}\right)_k} \right\}. \end{aligned}$$

Applying the properties $(\alpha+1)_{k-1} = (\alpha)_k / \alpha$ and $\alpha (\alpha+1)_k = (\alpha+k)(\alpha)_k$, thus $(a+b+1)\left(\frac{a+b+3}{2}\right)_k = (a+b+1+2k)\left(\frac{a+b+1}{2}\right)_k$, we have

$$\begin{aligned} & {}_3F_2 \left(\begin{array}{c} -k, a+1, b \\ -2k, \frac{a+b+1}{2} \end{array} \middle| 1 \right) = \frac{1}{\left(\frac{a+b+1}{2}\right)_k} \\ & \left\{ \frac{\left(\frac{a+1}{2}\right)_k \left(\frac{b+1}{2}\right)_k}{\left(\frac{1}{2}\right)_k \left(\frac{a+b+1}{2}\right)_k} + \frac{b}{a+b+2k+1} \left[\frac{\left(\frac{a+1}{2}\right)_k \left(\frac{b+1}{2}\right)_k}{\left(-\frac{1}{2}\right)_{k-1} (2k-1)} + \frac{\left(\frac{a}{2}+1\right)_k \left(\frac{b}{2}+1\right)_k}{\left(\frac{1}{2}\right)_k} \right] \right\}. \end{aligned}$$

Finally, knowing that $(-1/2)_k = \frac{(1/2)_k}{1-2k}$, after some simplification, we arrive at (5.7). □

Corollary 5.2 *Setting $k=M-n$, $a=2n+1$, $b=1$ in (5.7), and taking into account the property given in [4, Eq. 43:4:3], we arrive at*

$$
\begin{aligned}
& {}_3F_2\left(\left.\begin{array}{c} n-M, 2n+1, 1 \\ 2n-2M, n+1 \end{array}\right| 1\right) \\
= \; & \frac{1}{2\,(M+1)}\left[(2M+1)\left(1+\frac{\binom{2M}{2n}}{\binom{M}{n}^2}\right)-2n\right].
\end{aligned}
\tag{5.10}
$$

Taking into account (5.10), (5.6) is reduced to

$$
\begin{aligned}
& \langle s\rangle = \frac{n}{M+1}\left(2M+1-(2M-2n+1)\frac{\binom{M}{n}^2}{\binom{2M}{2n}}\right), \\
& n \leq M.
\end{aligned}
\tag{5.11}
$$

5.2. Case $\min(B,R) < n_{\min} \leq n_{\max} \leq \max(B,R)$

In this case, according to (3.7), we have

$$
\langle s\rangle = \frac{R!B!\; S_1^{(a)}\left(\max(R,B), \min(R,B), n_{\max}\right)}{(B+R)!\,(\max(R,B)-n_{\max})!\,(n_{\max}-1)!}.
$$

Applying the result given in (5.2) and knowing the fact that $B+R=\max(B,R)+\min(B,R)$, we arrive at

$$
\langle s\rangle = \frac{n_{\max}\,(R+B+1)}{\max(R,B)+1}.
\tag{5.12}
$$

6. Second moment

6.1. Case $n_R \leq R$, $n_B \leq B$

Taking $m=2$ in (3.2), we rewrite $S_2^{(a)}(X,Y,\nu)$ as follows:

$$
S_2^{(a)}(X,Y,\nu) = \hat{S}_2^{(a)}(X,Y,\nu) + \nu\, S_1^{(a)}(X,Y,\nu),
\tag{6.1}
$$

where we define and calculate $\hat{S}_2^{(a)}(X,Y,\nu)$ as follows:

$$
\begin{aligned}
& \hat{S}_2^{(a)}(X,Y,\nu) \\
= \; & \sum_{t=0}^{\infty}\frac{(X+Y-t-\nu-1)!\,(t+\nu+1)!}{t!\,(Y-t-1)!} \\
= \; & \frac{(X+Y-\nu-1)!\,(\nu+1)!}{(Y-1)!}\,{}_2F_1\left(\left.\begin{array}{c} 1-Y, \nu+2 \\ \nu+1-X-Y \end{array}\right| 1\right) \\
= \; & \frac{(X+Y+1)!\,(X-\nu)!\,(\nu+1)!}{(Y-1)!\,(X+2)!}.
\end{aligned}
\tag{6.2}
$$

Inserting in (6.1) the results (5.2) and (6.2), we arrive at

$$S_2^{(a)}(X,Y,\nu) = \frac{(X+Y+1)!\,(X-\nu)!\nu!}{(Y-1)!\,(X+1)!}\left[\frac{\nu+1}{X+2}+\frac{\nu}{Y}\right]. \tag{6.3}$$

Similarly,

$$S_2^{(b)}(X,Y,\nu) = \hat{S}_2^{(b)}(X,Y,\nu) + \nu\, S_1^{(b)}(X,Y,\nu), \tag{6.4}$$

where

$$\begin{aligned} &\hat{S}_2^{(b)}(X,Y,\nu) \\ = &\sum_{t=0}^{\infty}\frac{(X+Y-t-\nu-\bar{\nu})!\,(t+\nu+\bar{\nu})!}{(t+\bar{\nu}-1)!\,(Y-t-\bar{\nu})!} \\ = &\frac{(X+Y-\nu-\bar{\nu})!\,(\nu+\bar{\nu})!}{(\bar{\nu}-1)!\,(Y-\bar{\nu})!}\,{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu}+1\\ \bar{\nu},\bar{\nu}+\nu-X-Y\end{array}\right|1\right). \end{aligned} \tag{6.5}$$

Inserting in (6.4) the results (5.3) and (6.5), we arrive at

$$\begin{aligned} S_2^{(b)}(X,Y,\nu) = &\frac{(X+Y-\nu-\bar{\nu})!\,(\nu+\bar{\nu})!}{(\bar{\nu}-1)!\,(Y-\bar{\nu})!} \\ &\left[{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu}+1\\ \bar{\nu},\bar{\nu}+\nu-X-Y\end{array}\right|1\right)+\frac{\nu}{\bar{\nu}}\,{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu}+1\\ \bar{\nu}+1,\bar{\nu}+\nu-X-Y\end{array}\right|1\right)\right]. \end{aligned} \tag{6.6}$$

Now, substituting (6.3) and (6.6) in (3.6) for $m=2$, we eventually obtain

$$\begin{aligned} &S_2(X,Y,\nu) \\ = &\frac{1}{(X-\nu)!\,(\nu-1)!}\left[S_2^{(a)}(X,Y,\nu)-S_2^{(b)}(X,Y,\nu)\right] \\ = &\frac{(X+Y)!}{X!Y!}H(X,Y,\nu), \end{aligned} \tag{6.7}$$

where we have defined

$$\begin{aligned} &H(X,Y,\nu) \\ = &\frac{\nu(X+Y+1)\,[\nu(X+Y+2)+Y]}{(X+1)(X+2)} - \nu\bar{\nu}\frac{\binom{X}{\nu}\binom{Y}{\bar{\nu}}}{\binom{X+Y}{\nu+\bar{\nu}}} \\ &\left[{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu}+1\\ \bar{\nu},\bar{\nu}+\nu-X-Y\end{array}\right|1\right)+\frac{\nu}{\bar{\nu}}\,{}_3F_2\left(\left.\begin{array}{c}\bar{\nu}-Y,\nu+\bar{\nu}+1\\ \bar{\nu}+1,\bar{\nu}+\nu-X-Y\end{array}\right|1\right)\right] \end{aligned}$$

Finally, recall (3.5) for $m=2$ and take into account (6.7), to obtain

$$\begin{aligned} \langle s^2\rangle &= \frac{R!B!}{(B+R)!}\left[S_2(R,B,n_R)+S_2(B,R,n_B)\right] \\ &= H(R,B,n_R)+H(B,R,n_B), \end{aligned} \tag{6.8}$$

hence from (5.5) and (6.8), we obtain the standard deviation:

$$\sigma = \sqrt{H(R,B,n_R) + H(B,R,n_B) - \langle s\rangle^2}. \tag{6.9}$$

Notice that in the case $B = R = M$ and $n = n_R = n_B$, the second moment is reduced to

$$\left\langle s^2\right\rangle = \frac{2n(2M+1)}{M+2}\left(2n + \frac{M}{M+1}\right) \tag{6.10}$$
$$-2n^2\frac{\binom{M}{n}^2}{\binom{2M}{2n}}\left[{}_3F_2\left(\begin{array}{c} n-M, 2n+1, 1 \\ 2n-2M, n+1 \end{array}\middle| 1\right) + {}_3F_2\left(\begin{array}{c} n-M, 2n+1, 1 \\ 2n-2M, n \end{array}\middle| 1\right)\right].$$

The first hypergeometric series given in (6.10) has been calculated in (5.10). To calculate the second hypergeometric series, we need the next result.

Theorem 6.1 *If k is a nonnegative integer, then*

$$ {}_3F_2\left(\begin{array}{c} -k, a+1, b \\ -2k, \frac{a+b}{2} \end{array}\middle| 1\right) \tag{6.11}$$
$$= \frac{P_k^{(1)}(a,b)\left(\frac{a}{2}\right)_k\left(\frac{b+1}{2}\right)_k + Q_k^{(1)}(a,b)\left(\frac{a+1}{2}\right)_k\left(\frac{b}{2}+1\right)_k}{\left[a(a+b+2k)(a+b+2k+2)\right]\left(\frac{1}{2}\right)_k\left(\frac{a+b}{2}\right)_k},$$

$$\begin{aligned} P_k^{(1)}(a,b) &= (a+2k)\left[a(a+2)+b^2+2k(a+b)\right], \\ Q_k^{(1)}(a,b) &= 2ab(1+a+2k), \end{aligned}$$

and

$$ {}_3F_2\left(\begin{array}{c} -k, a, b \\ -2k, \frac{a+b}{2}-1 \end{array}\middle| 1\right) \tag{6.12}$$
$$= \frac{P_k^{(2)}(a,b)\left(\frac{a}{2}\right)_k\left(\frac{b+1}{2}\right)_k + Q_k^{(2)}(a,b)\left(\frac{a+1}{2}\right)_k\left(\frac{b}{2}+1\right)_k}{\left[(a+b-2)(a+b+2k)(a+b+2k+2)\right]\left(\frac{1}{2}\right)_k\left(\frac{a+b}{2}\right)_k},$$

$$\begin{aligned} P_k^{(2)}(a,b) &= (a+2k)\left[a^2+3b^2+2k(a+3b-2)-4\right], \\ Q_k^{(2)}(a,b) &= b\left[3a(a+2k)+(b-2)(b+2k+2)\right]. \end{aligned}$$

Proof Taking $z = 1$ and renaming the parameters in the contiguous relation given in [6, Eq. 16.3.7], we have

$$0 = {}_3F_2\left(\begin{array}{c} \alpha, \beta+1, \gamma \\ \delta, \varepsilon \end{array}\middle| 1\right)\beta(\delta+\varepsilon-\alpha-\beta-\gamma-1) \tag{6.13}$$
$$+ {}_3F_2\left(\begin{array}{c} \alpha, \beta, \gamma \\ \delta, \varepsilon \end{array}\middle| 1\right)\left[(2\beta-\delta)(2\beta-\varepsilon)+\beta(1-\beta)-(\beta-\alpha)(\beta-\gamma)\right]$$
$$- {}_3F_2\left(\begin{array}{c} \alpha, \beta-1, \gamma \\ \delta, \varepsilon \end{array}\middle| 1\right)(\beta-\delta)(\beta-\varepsilon).$$

Also, taking $p=3$, $q=2$ and $z=1$ in [8, Sect. 48(15)], we obtain

$$0 = {}_3F_2\left(\begin{array}{c}\alpha,\beta,\gamma\\ \delta,\varepsilon\end{array}\middle|1\right)(\beta-\varepsilon+1) - {}_3F_2\left(\begin{array}{c}\alpha,\beta+1,\gamma\\ \delta,\varepsilon\end{array}\middle|1\right)\beta + {}_3F_2\left(\begin{array}{c}\alpha,\beta,\gamma\\ \delta,\varepsilon-1\end{array}\middle|1\right)(\varepsilon-1). \tag{6.14}$$

Substituting $\alpha=-k$, $\beta=a$, $\gamma=b$, $d=-2k$ and $\varepsilon=\frac{a+b}{2}$ in (6.13), and (6.14), we get the following set of equations:

$$(b-a)(a+2k)U + 2a(1+a+2k)V - a(a+b+2k+2)X = 0, \tag{6.15}$$

$$(a+b-2)Y + (a-b+2)V - 2aX = 0, \tag{6.16}$$

where X and Y are given by the hypergeometric functions (6.11) and (6.12) respectively. Also, according to (4.22) and (5.7), we have

$$U = {}_3F_2\left(\begin{array}{c}-k,a-1,b\\ -2k,\frac{a+b}{2}\end{array}\middle|1\right) = \frac{\left(\frac{a}{2}\right)_k\left(\frac{b+1}{2}\right)_k}{\left(\frac{1}{2}\right)_k\left(\frac{a+b}{2}\right)_k}, \tag{6.17}$$

and

$$V = {}_3F_2\left(\begin{array}{c}-k,a,b\\ -2k,\frac{a+b}{2}\end{array}\middle|1\right) = \frac{a\left(\frac{a}{2}+1\right)_k\left(\frac{b+1}{2}\right)_k + b\left(\frac{a+1}{2}\right)_k\left(\frac{b}{2}+1\right)_k}{(a+b)\left(\frac{1}{2}\right)_k\left(\frac{a+b}{2}+1\right)_k}. \tag{6.18}$$

Therefore, solving for X and Y in (6.15) and (6.16), taking into account he property $(x+1)_k = (1+k/x)(x)_k$, after some simplification, we arrive at (6.11) and (6.12), as we wanted to prove. □

Corollary 6.2 *Taking $k=M-n$, $a=2n+1$, $b=1$, (6.12) can be expressed as*

$$ {}_3F_2\left(\begin{array}{c}n-M,2n+1,1\\ 2n-2M,n\end{array}\middle|1\right) = \frac{(2M-2n+1)(M+2n+3Mn) + M(2M+1)(n+1)\frac{\binom{2M}{2n}}{\binom{M}{n}^2}}{2n(M+1)(M+2)}. \tag{6.19}$$

Let us now reduce (6.10) by using (5.10) and (6.19). Thereby, after some simplification, we arrive at

$$\left\langle s^2\right\rangle = \frac{1}{(M+1)(M+2)}\left\{(2M+1)(n-2)\left[M+2n(M+1)\right] - n(2M-2n+1)\left[M+4n(M+1)\right]\frac{\binom{M}{n}^2}{\binom{2M}{2n}}\right\}. \tag{6.20}$$

Finally, according to (5.11) and (6.20), the standard deviation is

$$\begin{aligned}\sigma &= \frac{1}{M+1}\sqrt{\frac{n}{M+2}} \\ &\left\{M(2M+1)(M-n+1)+\frac{(2M-2n+1)\binom{M}{n}^2}{\binom{2M}{2n}}\right. \\ &\left.\left(\frac{n(M+2)(2n-2M-1)\binom{M}{n}^2}{\binom{2M}{2n}}-M(M-2n+1)\right)\right\}^{1/2}, \\ &n\leq M.\end{aligned} \tag{6.21}$$

6.2. Case $\min(B,R) < n_{\min} \leq n_{\max} \leq \max(B,R)$

In this case, according to (3.7), we have

$$\left\langle s^2\right\rangle = \frac{R!B!\,S_2^{(a)}(\max(R,B),\min(R,B),n_{\max})}{(B+R)!\,(\max(R,B)-n_{\max})!\,(n_{\max}-1)!}.$$

Applying the result given in (6.3), we arrive at

$$\left\langle s^2\right\rangle = \frac{n_{\max}(B+R+1)\left[n_{\max}(B+R+2)+\min(B,R)\right]}{(\max(B,R)+1)(\max(B,R)+2)}. \tag{6.22}$$

Taking into account (5.12) and knowing the fact that $B+R=\max(B,R)+\min(B,R)$, after some simplification, the standard deviation $\sigma=\sqrt{\langle s^2\rangle-\langle s\rangle^2}$ is given by

$$\begin{aligned}\sigma &= \frac{1}{\max(B,R)+1} \\ &\sqrt{\frac{n_{\max}(B+R+1)\left[RB-(n_{\max}-1)\min(B,R)\right]}{\max(B,R)+2}}.\end{aligned} \tag{6.23}$$

7. Numerical simulations

Simulations coded in MATHEMATICA confirm the results given for the mean value, i.e. (5.5), (5.11) and (5.12); and for the standard deviation, i.e. (6.9), (6.21), and (6.23). For these computational simulations, we have taken samples sizes from $N=100$ to $N=1000$. Also, for each sample size, the simulation has been repeated $n_r=30$ times, in order to apply the central limit theorem to the sampling distribution. Therefore, for the mean sampling distribution, we have calculated the size of the error bars as $\sigma_m/\sqrt{n_r}$, being σ_m the standard deviation of the statistic $\langle s\rangle$ considering n_r repetitions. For the standard deviation sampling distribution, we have calculated the size of the error bars as $\sigma_s/\sqrt{2(n_r-1)}$, being σ_s the standard deviation of the statistic σ considering n_r repetitions. These error bars are a good approximation when $n_r>10$.

Figures 1 and 2 show the mean value $\langle s\rangle$ and the standard deviation σ respectively, taking $n_R=5$, $n_B=7$ and $B=25$, $R=30$. Therefore, for the analytic calculation of $\langle s\rangle$ and σ, (5.5) and (6.9) have been used respectively.

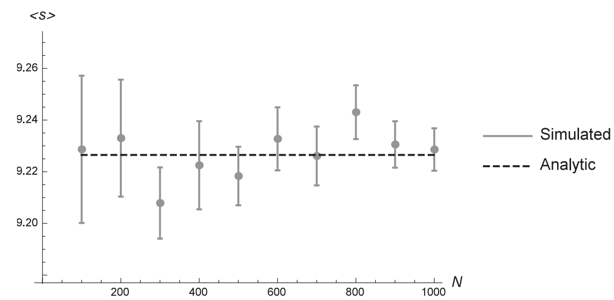

Figure 1. Mean value $\langle s\rangle$ for $n_R=5$, $n_B=7$ and $B=25$, $R=30$.

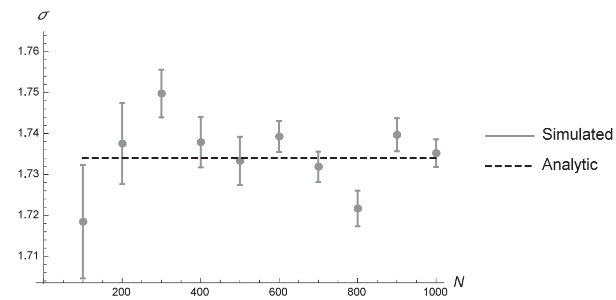

Figure 2. Standard deviation σ for $n_R=5$, $n_B=7$ and $B=25$, $R=30$.

Figures 3 and 4 show the mean value $\langle s\rangle$ and the standard deviation σ respectively, taking $n_R=18$, $n_B=22$ and $B=15$, $R=30$. Thereby, (5.12) and (6.23) have been used respectively for the calculation of $\langle s\rangle$ and σ.

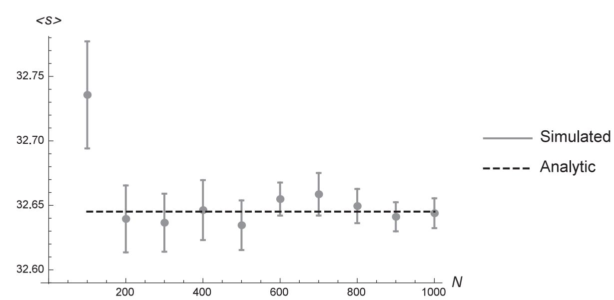

Figure 3. Mean value $\langle s\rangle$ for $n_R=18$, $n_B=22$ and $B=15$, $R=30$.

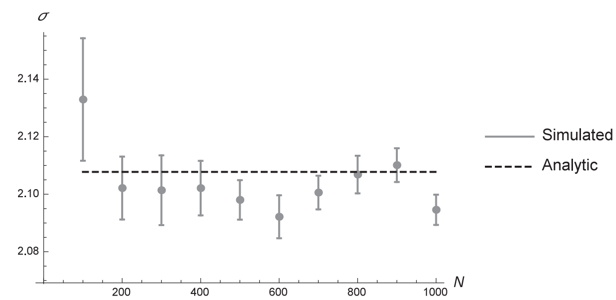

Figure 4. Standard deviation σ for $n_R=18$, $n_B=22$ and $B=15$, $R=30$.

Finally, Figures 5 and 6 show the mean value $\langle s\rangle$ and the standard deviation σ respectively, taking $n_R=n_B=15$ and $B=R=30$. Thereby, for the analytic calculation of $\langle s\rangle$ and σ, (5.11) and (6.21) have been used respectively.

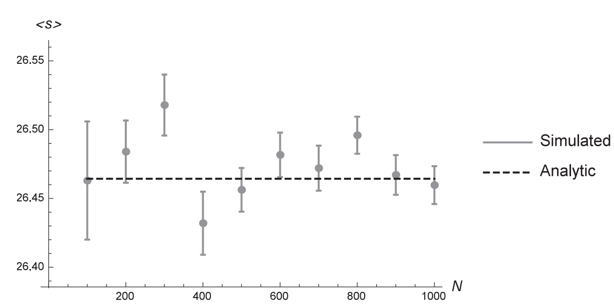

Figure 5. Mean value $\langle s\rangle$ for $n_R=n_B=15$ and $B=R=30$.

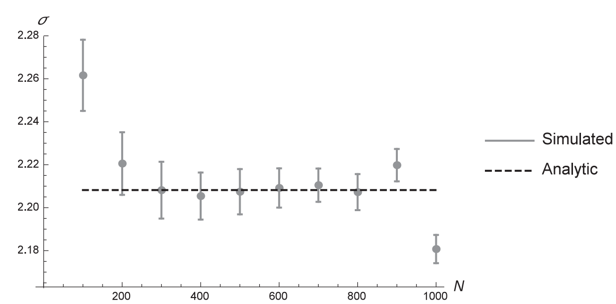

Figure 6. Standard deviation σ for $n_R=n_B=15$ and $B=R=30$.

Notice that the error bars for the graphs of $\langle s\rangle$ contain more or less the analytic value in the erf $\left(1/\sqrt{2}\right)\approx 68\%$ of the times, confirming the theoretical prediction. This is not exactly the case of the error bars for the graphs of σ since the sampling distribution of the standard deviation does not have to follow a normal distribution as the sampling distribution of the mean does.

8. Conclusion

Similarly to the hypergeometric probability distribution, we have considered a probability distribution $P_{n_R,n_B}^{R,B}(s)$ for which the random variable is the number of draws s without replacement that first obtain n_R items of type R or n_B items of type B.

On the one hand, the normalization of $P_{n_R,n_B}^{R,B}(s)$ leads to the hypergeometric formula (4.9). This hypergeometric formula does not seem to be reported in the literature. However, in the particular case of $n_R = n_B$, (4.9) is reduced to (4.10), for which we provide an alternative proof. This derivation is highly nontrivial and involves a three-term ${}_3\varphi_2$ q-hypergeometric series. Also, in the special case $n_R = n_B$ and $B = R$, (4.9) is reduced to (4.21), for which we provide an alternative proof using a known ${}_3F_2$ summation identity.

On the other hand, formulas for the mean value $\langle s \rangle$ and the standard deviation σ have been obtained in (5.5) and (6.9) respectively for the case $n_R \leq R$, $n_B \leq B$, and in (5.12) and (6.23) for the case $\min(B,R) < \min(n_R,n_B)$, $\max(n_R,n_B) \leq \max(B,R)$. Also, with the aid of the new summation formulas given in (5.7) and (6.12), reduced expressions for $\langle s \rangle$ and σ have been derived in (5.11) and (6.21) respectively for the case $n = n_R = n_B$ and $M = B = R$, where $n \leq M$. It is worth noting that as a by-product another new summation formula has been derived in (6.11).

Finally, computer simulations have been performed in order to confirm all the results calculated for the mean value and the standard deviation. These simulations have been coded in MATHEMATICA*.

General speaking, the technique of the normalization condition shown in this paper might be applied to other probability distributions in order to search new identities involving special functions. For instance, the multivariate hypergeometric distribution generalizes the hypergeometric distribution to an arbitrary number of items types, say N_i items of type i. Thereby, it is expected that the modified hypergeometric distribution presented here $P_{n_R,n_B}^{R,B}(s)$ can be generalized to an arbitrary number of items types, i.e. to $P_{n_1,\ldots,n_k}^{N_1,\ldots,N_k}(s)$. We leave this generalization as an open problem to the reader.

Conflict of interest

The author declares that he has not conflict of interest.

References

[1] Feller W. An Introduction to Probability Theory and Its Applications, Vol. 1. 3rd ed. New York, NY, USA: Wiley, 1968.

[2] Udias A, Rice J. Statistical analysis of microearthquake activity near San Andreas geophysical observatory, Hollister, California. Bulletin of the Seismological Society of America 1975; 65: 809-827.

[3] Andrews GE, Askey R, Roy R. Encyclopedia of Mathematics and its Applications, Vol. 71. Special Functions. New York, NY, USA: Cambridge University Press, 2004.

[4] Oldham K, Myland J, Spanier J. An Atlas of functions: with Equator, the Atlas Function Calculator. 2nd ed. New York, NY, USA: Springer, 2008.

*The MATHEMATICA code is available at https://bit.ly/2AseWDv.

[5] Gasper G, Rahman GM. Basic Hypergeometric Series. Cambridge, UK: Cambridge University Press, 1990.

[6] Olver FWJ, Lozier DW, Boisvert RF, Clark CW (editors). NIST Handbook of Mathematical Functions. New York, NY, USA: Cambridge University Press, 2010.

[7] Prudnikov AP, Brychov YA, Marichev OI. Integrals and Series, Vol. 3. More Special Functions. New York, NY, USA: Gordon and Breach Science Publishers, 1986.

[8] Rainville ED. Special Functions. New York, NY, USA: The Macmillan Co., 1960.