Handy formulas for binomial moments

Maciej Skorski

University of Warsaw, Poland

maciej.skorski@uw.edu.pl (M. Skorski)

Abstract   Despite the relevance of the binomial distribution for probability theory and applied statistical inference, its higher-order moments are poorly understood. The existing formulas are either not general enough, or not structured and simplified enough for intended applications.

This paper introduces novel formulas for binomial moments in the form of *polynomials in the variance* rather than in the success probability. The obtained formulas are arguably better structured, simpler and superior in their numerical properties compared to prior works. In addition, the paper presents algorithms to derive these formulas along with working implementation in Python's symbolic algebra package.

The novel approach is a combinatorial argument coupled with clever algebraic simplifications which rely on symmetrization theory. As an interesting byproduct *asymptotically sharp estimates for central binomial moments* are established, improving upon previously known partial results.

Keywords   Binomial distribution, high-order moments, moment asymptotics, symbolic algebra

2010 MSC   60E05

1   Introduction

1.1   Background and related work

The binomial distribution $\text{Binom}(n, p)$, which counts the total number of successes within n independent trials each succeeding with probability p, is of historical and fundamental importance for probability theory and applied statistical inference. In particular, it appears in quantitative variants of the central limit theorem [16, 8, 25, 29], and is broadly used in statistical modelling [17, 1, 9, 7, 26] also as a building block of more advanced models [27]; this includes performing A/B tests on conversion rates popular in e-commerce [11].

VTeX

Despite this large body of work on approximate inference, little is known about the *exact higher moments* of the binomial distribution. Besides being of natural interest, such formulas are needed to develop various concentration bounds; for example, binomial moment bounds were proven useful for analyzing random projections [23].

While the textbooks usually cover only the variance, sometimes also the skewness and kurtosis), there have been only few research papers discussing formulas for binomial moments of order $d > 4$. The first recursion formula for binomial moments appeared in [3] for the special case of $p = \frac{1}{2}$. The case of general p was handled in [13] by means of recursions utilizing Stirling numbers of the first kind. This was subsequently simplified by means of moment generating functions in [15], and resulted in a more compact formula involving Stirling numbers of the second kind. Very recently, a recursion-free derivation of raw moments has been presented in [19]. Overall, the common idea is to see the moments as (more or less explicit) polynomials in n and p and group terms, in order to simplify the formula or establish other desired properties such as nonnegativity of terms or boundedness of recursion depth (as in the recent result on Beta moments [24]).

The discussed approaches still do not offer a satisfactory answer, as the formulas are not handy enough to be directly applicable. The author of the most general formula in [15] didn't manage to obtain nonnaive bounds on the binomial moments: the bound $O(nq)^d$ with $q = 1 - p$ on the d-th central moment [15] valid for $p < \frac{1}{2}$ is trivial as the centered binomial random variable is bounded between $-pn$ and qn – no extra formulas are needed; this bound is far from the true behaviour $O(nq)^{d/2}$ when $nq \to \infty$ (obtained by the Central Limit Theorem). The main formula in [15] is actually a mixture of positive and negative contributions, which makes its numerical convergence problematic and theoretical analysis very difficult (as seen above). Moreover, all the prior works do not fully exploit the symmetry and produce rather complicated formulas in terms of p; one could expect simpler expressions on the central moments with the variance $n\sigma^2 = np(1-p)$ as a variable. Lastly, the discussed prior works are rather scarce in their presentation of related works and techniques, in particular they seem to have overlooked that the formulas with the appearance of Stirling numbers follow easier by the established approach of factorial moments [14, 2].

1.2   Summary of contributions

Addressing the aforementioned issues with approaches in prior works, this paper offers the following novel contributions to computing binomial moments:

- **link to factorial moment,** which simplifies the approach from prior works,

- **variance-formula** for equivalent yet simpler expressions in $\sigma^2 = p(1-p)$,

- **algorithms with implementation** to compute the variance-formula,

- **stable formula** as an explicit sum with positive terms,

- **asymptotically sharp bounds** on binomial moments as an application.

In summary, when compared to prior works, these results bring a broader scope of the techniques, as well as lead to arguably more handy formula; another added value is the contributed algorithm and its Python implementation.[1]

1.3 Organization

The remainder of the paper is organized as follows: the necessary background is given in Section 2, proofs appear in Section 4, the results are presented in Section 3, the sharp asymptotics are discussed in Section 5, and Section 6 concludes the work. The Python implementation is presented in Appendix A.

2 Preliminaries

2.1 Binomial distribution

A random variable S follows the binomial distribution with parameters n and p, denoted as $S \sim \text{Binom}(n, p)$, when the probability density function is

$$\Pr[S = k] = \binom{n}{k} p^k q^{1-k}, \quad q \triangleq 1 - p, \quad k = 0, \ldots, n. \tag{1}$$

2.2 Moments

Let d be a positive integer. The *raw moment* of order d of a random variable S is defined as $\mathbb{E}[S^d]$, while the *central moment* of order d of S equals $\mathbb{E}[(S - \mathbb{E}[S])^d]$. We also use the *factorial moment*, defined as $\mathbb{E}[S^{\underline{d}}]$ where $x^{\underline{d}} = x(x-1) \cdots (x - (d-1))$ is called the d-th *falling factorial* [12].

2.3 Special numbers

To state the results, we need *Stirling numbers of the second kind*. The symbol $\left\{ {n \atop k} \right\}$ stands for the number of ways of partitioning an n element set into k nonempty subsets. We also need *multinomial coefficients* defined as $\binom{d}{d_1 \ldots d_k} = \frac{d!}{d_1! \cdots d_k!}$ when $\sum_{i=1}^{k} d_i = d$ and $\min_i d_i \geqslant 0$ and 0 otherwise, which extend the binomial coefficients. By the multinomial theorem, we have that $(x_1 + \cdots + x_n)^d = \sum_{d_1, \ldots, d_n} \binom{d}{d_1 \ldots d_n} x_1^{d_1} \cdots x_n^{d_n}$, extending the binomial formula.

2.4 Polynomials

To work out the desired polynomial formulas, we need some standard algebraic notation. By $\mathbb{Z}[x_1, \ldots, x_k]$ we denote polynomials with integer coefficients in variables $x_1, \ldots, x_k$. A polynomial is symmetric if it is invariant under interchanging variables, and antisymmetric when it gets negated under exchange of any two variables. The fundamental theorem of symmetric polynomials states that any symmetric polynomial from $\mathbb{Z}[x_1, \ldots, x_k]$ can be written as a polynomial in the elementary symmetric functions $s_j(x_1, \ldots, x_n) = \sum_{1 \leqslant i_1 \ldots \leqslant i_j} x_{i_1} \cdots x_{i_j}$ for $j = 1, \ldots, k$, with integer coefficients. Moreover, antisymmetric polynomials can be written as a product of a symmetric polynomial and Vandermonde's determinant $\prod_{1 \leqslant i < j \leqslant k} (x_i - x_j)$ (see, for example, [22, 28]).

---

[1] For code and examples visit the OSF repository https://osf.io/hjscb/

3 Results

Below we discuss the contributions in more detail, deferring proofs to the end part of the paper. We denote $S \sim \mathrm{Binom}(n, p)$, and fix a positive integer d.

3.1 Raw binomial moments and factorial moments

The first result is the derivation of the closed-form formula for raw binomial moments. While this formula appears in prior works [3, 15, 13], the arguments developed here are simpler, as they avoid the machinery of generating functions. The two alternative proof techniques are, respectively: a) linking central and factorial moments b) a direct counting argument.

Theorem 1 (Raw Binomial Moments). *The raw moments are given by*

$$\mathbb{E}[S^d] = \sum_{k=0}^{d} n^{\underline{k}} \left\{ {d \atop k} \right\} p^k. \tag{2}$$

The two proofs appear respectively in Section 4.1 and Section 4.2. Below, in Table 1, we list the explicit expressions for the first 10 moments.

Table 1. Formulas for Raw Binomial Moments

d	$\mathbb{E}[S]^d, \quad S \sim \mathrm{Binom}(n, p)$
1	pn
2	$p^2 n^{\underline{2}} + pn$
3	$p^3 n^{\underline{3}} + 3p^2 n^{\underline{2}} + pn$
4	$p^4 n^{\underline{4}} + 6p^3 n^{\underline{3}} + 7p^2 n^{\underline{2}} + pn$
5	$p^5 n^{\underline{5}} + 10p^4 n^{\underline{4}} + 25p^3 n^{\underline{3}} + 15p^2 n^{\underline{2}} + pn$
6	$p^6 n^{\underline{6}} + 15p^5 n^{\underline{5}} + 65p^4 n^{\underline{4}} + 90p^3 n^{\underline{3}} + 31p^2 n^{\underline{2}} + pn$
7	$p^7 n^{\underline{7}} + 21p^6 n^{\underline{6}} + 140p^5 n^{\underline{5}} + 350p^4 n^{\underline{4}} + 301p^3 n^{\underline{3}} + 63p^2 n^{\underline{2}} + pn$
8	$p^8 n^{\underline{8}} + 28p^7 n^{\underline{7}} + 266p^6 n^{\underline{6}} + 1050p^5 n^{\underline{5}} + 1701p^4 n^{\underline{4}} + 966p^3 n^{\underline{3}} + 127p^2 n^{\underline{2}} + pn$
9	$p^9 n^{\underline{9}} + 36p^8 n^{\underline{8}} + 462p^7 n^{\underline{7}} + 2646p^6 n^{\underline{6}} + 6951p^5 n^{\underline{5}} + 7770p^4 n^{\underline{4}} + 3025p^3 n^{\underline{3}} + 255p^2 n^{\underline{2}} + pn$
10	$p^{10} n^{\underline{10}} + 45p^9 n^{\underline{9}} + 750p^8 n^{\underline{8}} + 5880p^7 n^{\underline{7}} + 22827p^6 n^{\underline{6}} + 42525p^5 n^{\underline{5}} + 34105p^4 n^{\underline{4}} + 9330p^3 n^{\underline{3}} + 511p^2 n^{\underline{2}} + pn$

3.2 Central binomial moments

Symmetric structure. While in prior works the formulas are derived in terms of p, we go beyond that exploiting the symmetry and showing that the formulas can be written in terms of the variance $\sigma^2 = p(1 - p)$, which makes them much simpler. The following theorem proves what can be conjectured by inspection of known formulas for small-order moments.

Theorem 2 (Existence of Variance-Based Formula). *For $S \sim \mathrm{Binom}(n, p)$ the moment $\mathbb{E}[(S - \mathbb{E}[S])^d]$ is a symmetric polynomial in p and q when d is even, and antisymmetric when d is odd. In particular, denoting $\sigma^2 \triangleq pq$ we have*

$$\mathbb{E}[(S - \mathbb{E}[S])^d] \in \begin{cases} \mathbb{Z}[n, \sigma^2], & d \text{ even,} \\ (1 - 2p)\mathbb{Z}[n, \sigma^2], & d \text{ odd.} \end{cases} \tag{3}$$
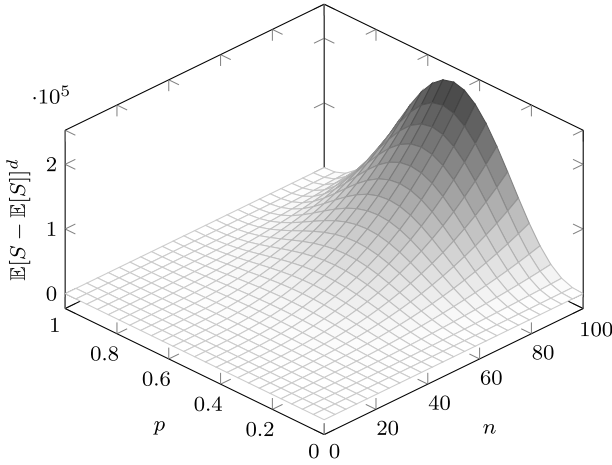
Fig. 1. The d-th central moment of $S(n, p)$, where $d = 6$

Table 2 illustrates this result, providing explicit moments of order $d = 2, \ldots, 10$. The usefulness of the formula in Theorem 2 is its simplicity when compared to the representation in terms of p alone. The result is intuitive, but not straightforward to prove; two arguments are given, based on a) the theory of symmetric functions, see Section 4.3 and b) the novel combinatorial formula, see Section 4.4. The algorithm deriving the exact formulas is discussed later.

Table 2. Central Moments of Binomial Distribution. As above we denote $\sigma^2 = p(1 - p)$

d	$\mathbb{E}[(S - \mathbb{E}[S])^d]$, $S \sim \mathrm{Binom}(n, p)$
2	$n\sigma^2$
3	$n\sigma^2(-2p + 1)$
4	$3n^2\sigma^4 + n\left(-6\sigma^4 + \sigma^2\right)$
5	$(-2p + 1)\left(10n^2\sigma^4 + n\left(-12\sigma^4 + \sigma^2\right)\right)$
6	$15n^3\sigma^6 + n^2\left(-130\sigma^6 + 25\sigma^4\right) + n\left(120\sigma^6 - 30\sigma^4 + \sigma^2\right)$
7	$(-2p + 1)\left(105n^3\sigma^6 + n^2\left(-462\sigma^6 + 56\sigma^4\right) + n\left(360\sigma^6 - 60\sigma^4 + \sigma^2\right)\right)$
8	$105n^4\sigma^8 + n^3\left(-2380\sigma^8 + 490\sigma^6\right) + n^2\left(7308\sigma^8 - 2156\sigma^6 + 119\sigma^4\right) + n\left(-5040\sigma^8 + 1680\sigma^6 - 126\sigma^4 + \sigma^2\right)$
9	$(-2p + 1)\left(1260n^4\sigma^8 + n^3\left(-13216\sigma^8 + 1918\sigma^6\right) + n^2\left(32112\sigma^8 - 6948\sigma^6 + 246\sigma^4\right) + n\left(-20160\sigma^8 + 5040\sigma^6 - 252\sigma^4 + \sigma^2\right)\right)$
10	$945n^5\sigma^{10} + n^4\left(-44100\sigma^{10} + 9450\sigma^8\right) + n^3\left(303660\sigma^{10} - 99120\sigma^8 + 6825\sigma^6\right) + n^2\left(-623376\sigma^{10} + 240840\sigma^8 - 24438\sigma^6 + 501\sigma^4\right) + n\left(362880\sigma^{10} - 151200\sigma^8 + 17640\sigma^6 - 510\sigma^4 + \sigma^2\right)$

Corollary 1 (Skewness and Kurtosis of Binomial Distribution). *The skewness and excess kurtosis equal, respectively,* $\frac{q-p}{\sqrt{npq}}$ *and* $\frac{1-6pq}{npq}$.

Positive polynomial representation. As mentioned in the introduction, the only closed-form formula due to [15] is an alternating sum with no readable leading term, which makes it hard to use; in particular the discussion in [15] fails to give nontrivial bounds on binomial moments. The novelty of this work is a formula consisting of *positive terms*. This makes it more stable for numerical computations and more handy in theoretical analysis.

Theorem 3 (Explicit Stable Formulas). *For $S \sim \mathrm{Binom}(n, p)$, $q \triangleq 1 - p$ and any*

positive integer d the following holds:

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = \sum_{k=1}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{k}(pq)^k \sum_{d_1 \ldots d_k \geqslant 2} \binom{d}{d_1 \ldots d_k} \prod_{i=1}^{k}(q^{d_i-1} - (-p)^{d_i-1}).$$

Remark 1 (Nonnegativity)**.** *The terms under the sum are all nonnegative when $p \leqslant \frac{1}{2}$. Since $n - S \sim \mathrm{Binom}(n, q)$, it follows that we have $S - \mathbb{E}[S] \sim -(\mathrm{Binom}(n, q) - qn)$; with the help of this identity, working with central binomial moments can be always reduced to the case $p < \frac{1}{2}$.*

Remark 2 (Closed-form Symmetric Formula)**.** *The above result gives an explicit formula for Theorem 2, and provides an alternative proof of that result.*

3.3 Asymptotically sharp moment bounds

To illustrate how useful is our *positive* representation established in Theorem 3, we derive the *sharp bounds* on (normalized) central binomial moments. This problem has remained open so far; some recent works used ad hoc upper bounds to estimate the binomial moments and tails (works on random projections, particularly [23]).

Theorem 4. *Let $S \sim \mathrm{Binom}(n, p)$ and $\sigma^2 = p(1 - p)$. Then for any integer $d > 1$ we have*

$$\left|\mathbb{E}[(S - \mathbb{E}[S])^d]\right|^{1/d} = C(n, p, d) \cdot \max\left\{k^{1-\frac{k}{d}} \cdot (n\sigma^2)^{\frac{k}{d}} : k = 1, \ldots, \lfloor d/2 \rfloor\right\},$$

where $C(n, p, d)$ is uniformly bounded by $(3e)^{-1} \leqslant C(n, p, d) \leqslant (5/2)^{1/5}e^{1/2}$.

The estimate is uniformly sharp in all parameters; for the special case when $n\sigma^2 \to \infty$ and d is fixed, the d-th central moment grows, up to a constant, as $(n\sigma^2)^{d/2}$ which matches the central limit theorem combined with the explicit formulas for moments of the normal distribution [20]. In contrast to Theorem 4, the formula in [15] gives in this setup only much worse $O(nq)^d$, which anyway follows trivially since S is bounded by n. The proof uses Theorem 3.

3.4 Algorithms and implementation

We have seen that the variance-based representation in Theorem 2 is useful, yet it is not immediate *how to efficiently compute* this polynomial. To this end, we develop two different algorithms, both implemented in the popular Python symbolic algebra package `Sympy` [18].

Algorithm 1 takes advantage of the *fundamental theorem on symmetric polynomials* (see, for example, [10]). Specifically, there is an explicit procedure for converting any symmetric polynomial in p, q into a polynomial in variables $p + q$, pq (the basic symmetric polynomials in two variables); substituting $p + q = 1$ we are left with a polynomial in $\sigma^2 = pq$. For even d we start with a symmetric polynomial and obtain a polynomial in σ^2. In turn, for odd d we start with an antisymmetric polynomial and apply this procedure to its symmetric part, which results in a polynomial in σ^2 plus the factor $q - p = 1 - 2p$.

In turn, Algorithm 2 uses the power of *elimination theory*, to recover the representation in σ^2 from any formula written in terms of p. Essentially, it simplifies the input

---

Algorithm 1: Variance Formula for Central Binomial Moments

Result: The d-th central binomial moment in terms of n and $\sigma^2 = p(1-p)$.
The outcome is a polynomial with integer coefficients.
$U \leftarrow \mathbb{E}(S - \mathbb{E}[S])^d$, with $U \in \mathbb{Z}[n][p,q]$ // write as polynomial in p,q, e.g. by Theorem 3

if *d even* **then**

> $V \leftarrow$ symmetrize(U), $V \in \mathbb{Z}[n][pq, p+q]$ // represent by elementary symmetric polynomials
> $V \leftarrow V|_{p+q \leftarrow 1}$ // substitute the relation $p + q = 1$

end
else

> $V \leftarrow V/(q-p)$ // subtract 'unsymmetric' part
> $V \leftarrow$ symmetrize(U), $V \in \mathbb{Z}[n][pq, p+q]$ // represent by elementary symmetric polynomials
> $V \leftarrow V|_{p+q \leftarrow 1}$ // substitute the relation $p + q = 1$
> $V \leftarrow (1 - 2p) \cdot V$ // add unsymmetric part back

end
return V

---

polynomial in p with respect to the polynomial $\sigma^2 - p(1-p)$ leveraging the elimination properties of *Groebner bases* (see, for example, [5]). The output is a polynomial in σ^2 (plus the factor $1 - 2p$ for odd d).

Algorithm 2: Variance Formula for Central Binomial Moments

Result: The d-th central binomial moment in terms of n and $\sigma^2 = p(1-p)$.
The output is a polynomial with integer coefficients.
$U \leftarrow \mathbb{E}(S - \mathbb{E}[S])^d$, $U \in \mathbb{Z}[n,p]$ // represent as polynomial in n,p, for example Theorem 3 and substitution $q \leftarrow 1 - p$

if *d even* **then**

> $F_1, F_2 \leftarrow U, \sigma^2 - p(1-p)$ // moment and variance formulas
> $G \leftarrow$ GroebnerB(poly $= \{F_1, F_2\}$, vars $= (p, n, \sigma^2)$, order $=$ lex) // Groebner basis, lex order
> $V \leftarrow G \cap \mathbb{Z}[n, \sigma^2]$ // extract variance-dependent formula

end
else

> $F_1, F_2 \leftarrow U/(1-2p), \sigma^2 - p(1-p)$ // subtract 'odd' part
> $G \leftarrow$ GroebnerB(poly $= \{F_1, F_2\}$, vars $= (p, n, \sigma^2)$, order $=$ lex) // Groebner basis, lex order
> $V \leftarrow G \cap \mathbb{Z}[n, \sigma^2]$ // extract variance-dependent part
> $V \leftarrow (1 - 2p) \cdot V$ // add 'odd' part back

end
return V

---

4 Proofs

4.1 First proof of Theorem 1

The proof is based on the fact that the *factorial moments* of the binomial distribution are easy to compute. Namely (see [21]), we have the following proposition.

Proposition 1 (Factorial Binomial Moments). *Let $S \sim \text{Binom}(n, p)$. Then*

$$\mathbb{E}[S^{\underline{d}}] = \binom{n}{d} p^d. \tag{4}$$

Then it remains to connect factorial moments to standard moments, or in other terms: factorial powers to powers. It is well known (see, for example, the discussion in [4]) that this base change is given in terms of the Stirling numbers of the second kind. We state this fact formally below.

Proposition 2 (Base Change from Falling Factorials to Powers). *For positive integers x and d the following holds:*

$$x^d = \sum_{k=0}^{d} \begin{Bmatrix} d \\ k \end{Bmatrix} x^{\underline{k}}. \tag{5}$$

Now Theorem 1 follows by applying Proposition 2 to $x := S$, and then using Proposition 1 to compute the expectation of $S^{\underline{k}}$ for $k = 0, \ldots, d$.

4.2   Second proof of Theorem 1

Here we take a direct approach, writing $S = \sum_{i=1}^{n} X_i$ where $X_i \sim^{iid}$ Bern(p). Using the multinomial expansion and the independence of X_i we obtain

$$\mathbb{E}[S^d] = \sum_{d_1,\ldots,d_n} \binom{d}{d_1 \ldots d_n} \prod_i \mathbb{E}[X_i^{d_i}].$$

We now group the expressions in the above sum, depending on the number of nonzero elements in $(d_i)_i$. Denoting $\|(d_i)\|_0 = \#\{i : d_i > 0\}$ and using the fact that $\mathbb{E}X_i^{d_i} = p$ when $d_i > 0$, we obtain

$$\mathbb{E}[S^d] = \sum_{k=0}^{d} \sum_{(d_i)_i : \|(d_i)_i\|_0 = k} \binom{d}{d_1 \ldots d_n} p^k.$$

By the symmetry of multinomial coefficients this equals

$$\mathbb{E}[S^d] = \sum_{k=0}^{d} \sum_{d_1 \ldots d_k > 0} \binom{d}{d_1 \ldots d_k} \binom{n}{k} p^k.$$

Finally, we observe that the expression $\sum_{d_1 \ldots d_k > 0} \binom{d}{d_1 \ldots d_k}$ counts the number of ways of partitioning $\{1, \ldots, d\}$ into k nonempty *labeled* subsets; thus, this numbers equals $k! \cdot \begin{Bmatrix} d \\ k \end{Bmatrix}$ which finishes the proof.

4.3   First proof of Theorem 2

From Equation (1) we obtain

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = \sum_{k} \binom{n}{k} p^k q^{n-k} (k - np)^d.$$

Replacing k by $n - k$ and using the symmetry of binomial coefficients we obtain

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = \sum_{k} \binom{n}{k} p^{n-k} q^k (nq - k)^d.$$

When d is even, comparing these two equivalent expressions we see that they are symmetric as polynomials in p and q. By the fundamental theorem of symmetric polynomials, this can be written as a polynomial in pq and $p + q$; in our case $p + q = 1$ and the claim follows. If d is odd then $(nq - k)^d = -(k - nq)^d$ and we get antisymmetric polynomials in p, q which can be written as a product of $p - q$ and a symmetric polynomial. The latter, by the fundamental theorem, is a polynomial in $p + q$ and pq; since $p + q = 1$ the result follows.

4.4 Second proof of Theorem 2

By inspecting the products $\prod_{i=1}^{k}(q^{d_i-1} - (-p)^{d_i-1})$ that appear in Theorem 3, it can be seen that each of them is symmetric in p, q when $\sum_i d_i = d$ is even, and antisymmetric when $\sum_i d_i = d$ is odd. This is because $p, q \to q^{d_i-1} - (-p)^{d_i-1}$ is symmetric when d_i is even, and antisymmetric otherwise. The claim now follows.

4.5 Proof of Theorem 3

As in the proof of Theorem 1 we arrive at

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = \sum_{k=0}^{d} \sum_{d_1 \ldots d_k > 0} \binom{n}{k}\binom{d}{d_1 \ldots d_k} \prod_{i=1}^{k} \mathbb{E}[(X_i - \mathbb{E}[X_i])^{d_i}].$$

Denote $x = 1 - \frac{1}{1-p}$, then $\mathbb{E}[(X_i - \mathbb{E}[X_i])]^{d_i} = p(1 - p)^{d_i}(1 - x^{d_i-1})$ and thus

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = (1 - p)^d \sum_{k=0}^{d} \binom{n}{k} p^k \sum_{d_1 \ldots d_k > 0} \binom{d}{d_1 \ldots d_k} \prod_{i=1}^{k} (1 - x^{d_i-1}).$$

With some further simplifications and grouping we can write

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = (1 - p)^d \sum_{k=1}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{k} p^k \underbrace{\sum_{d_1 \ldots d_k \geqslant 2} \binom{d}{d_1 \ldots d_k} \prod_{i=1}^{k} (1 - x^{d_i-1})}_{U_k},$$

or equivalently

$$\mathbb{E}[(S - \mathbb{E}[S])^d] = \sum_{k=1}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{k} (pq)^k \sum_{d_1 \ldots d_k > 1} \binom{d}{d_1 \ldots d_k} \prod_{i=1}^{k} (q^{d_i-1} - (-p)^{d_i-1}).$$

This finishes the proof. In addition to that, in what follows, we discuss how to further group terms and speed up computations. We can write

$$U_k = \sum_{j=0}^{k} \binom{k}{j} \sum_{d_1 \ldots d_j > 1} \sum_{d_{j+1} \ldots d_k > 1} \binom{d}{d_1 \ldots d_k} (-1)^j x^{d_1 + \cdots + d_j - j},$$

and since $\binom{d}{d_1 \dots d_k} = \binom{d_1 + \dots + d_j}{d_1 \dots d_j} \cdot \binom{d - (d_1 + \dots + d_j)}{d_{j+1} \dots d_k} \cdot \binom{d}{d_1 + \dots + d_j}$ we obtain

$$U_k = \sum_{j=0}^{k} \binom{k}{j} \sum_{\ell} \binom{d}{\ell} \sum_{d_1 \dots d_k > 1} \binom{\ell}{d_1 \dots d_j} \binom{d - \ell}{d_{j+1} \dots d_k} (-1)^j x^{\ell - j},$$

and thus

$$U_k = \sum_{j=0}^{k} \sum_{\ell=0}^{d} \binom{k}{j} \binom{d}{\ell} j! (k-j)! S_2(\ell, j) S_2(d - \ell, k - j)(-1)^j x^{\ell - j}$$

$$= k! \sum_{\ell=0}^{d} \binom{d}{\ell} \sum_{j=0}^{k} S_2(\ell, j) S_2(d - \ell, k - j)(-1)^j x^{\ell - j},$$

where $S_2(n, k)$ denotes the number of ways of partitioning an n-element set into k subsets of cardinality at least 2 (a variation on Stirling numbers of the second kind). This can be used to develop an equivalent, but faster to compute, formula.

5   Applications

5.1   Proof of Theorem 4

Throughout the proof, we will use the elementary estimates $(m/\mathrm{e})^m \leqslant m! \leqslant m^m$ (the second inequality is obvious, the first one follows by rearranging to $\mathrm{e}^m \geqslant \frac{m^m}{m!}$ and Taylor's expansion) and $(n/k)^k \leqslant \binom{n}{k} \leqslant (n\mathrm{e}/k)^k$ (see, for example, [6]).

By applying Remark 1 we can assume $p \leqslant \frac{1}{2}$. Then $q \geqslant p$, and thus $0 \leqslant q^{d_i - 1} - (-p)^{d_i - 1} \leqslant p + q = 1$ for any $d_i \geqslant 2$. In view of Theorem 3, we obtain the bound

$$\mathbb{E}[(S - \mathbb{E}[S])^d] \leqslant \sum_{k=1}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{k} (pq)^k \sum_{d_1 \dots d_k \geqslant 2} \binom{d}{d_1 \dots d_k}.$$

Since we have

$$\sum_{d_1 \dots d_k \geqslant 2} \binom{d}{d_1 \dots d_k} \leqslant \sum_{d_1 \dots d_k \geqslant 0} \binom{d}{d_1 \dots d_k} = k^d,$$

we further obtain

$$\mathbb{E}[(S - \mathbb{E}[S])^d] \leqslant \sum_{k=1}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{k} (pq)^k k^d.$$

Denoting $\sigma^2 = pq$ and using the bound $\binom{n}{k} \leqslant (n\mathrm{e}/k)^k$, we finally obtain

$$\mathbb{E}[(S - \mathbb{E}[S])^d] \leqslant \lfloor d/2 \rfloor \cdot \max\{\mathrm{e}^k k^{d-k} \cdot (n\sigma^2)^k : 1 \leqslant k \leqslant \lfloor d/2 \rfloor\},$$

so that (noticing that $(d/2)^{1/d} \leqslant (5/2)^{1/5}$ for $d \geqslant 2$)

$$\left(\mathbb{E}[(S - \mathbb{E}[S])^d] \right)^{\frac{1}{d}} \leqslant (5/2)^{1/5} \mathrm{e}^{1/2} \cdot \max\{k^{1 - \frac{k}{d}} \cdot (n\sigma^2)^{\frac{k}{d}} : 1 \leqslant k \leqslant \lfloor d/2 \rfloor\}. \quad (6)$$

We now move on to the lower bound. The idea is to fix k and look at terms such that $d_1, \ldots, d_k$ are nearly equal, to make $\binom{d}{d_1 \ldots d_k}$ possibly large.

Let us write $d = k \cdot r + \ell$ with an integer $r \geqslant 2$ and nonnegative integer ℓ such that $0 \leqslant \ell < k$; this is possible as $k \leqslant d/2$, by writing $d = k \cdot r + \ell$ where $0 \leqslant \ell < k$. Define $d_i = r + 1$ when $1 \leqslant i \leqslant \ell$ and $d_i = r$ when $\ell < i \leqslant k$, so that $\sum_i d_i = d$. For this tuple (d_i) we have

$$\binom{d}{d_1 \ldots d_k} = \frac{d!}{\prod_{i=1}^{k} d_i!} \geqslant \frac{d!}{(r+1)!^{\sum_i d_i}} = \frac{d!}{(r+1)^d}.$$

Since $d! \geqslant (d/e)^d$, $r + 1 \leqslant \frac{3}{2} r$, and $d/r \geqslant k$, it follows that

$$\binom{d}{d_1 \ldots d_k} \geqslant \left(\frac{2k}{3e}\right)^d.$$

When d_i are even, we have $q^{d_i-1} - (-p)^{d_i-1} = q^{d_i-1} + p^{d_i-1} \geqslant 2 \cdot \frac{1}{2^{d_i-1}}$ by Jensen's inequality applied to the function $u \to u^{d_i-1}$ and $p + q = 1$. Since in the summation we consider $(d_i)_i$ such that $\sum_i d_i = d$ and $d_i \geqslant 2$, we obtain

$$\prod_{i=1}^{k} (q^{d_i-1} - (-p)^{d_i-1}) \geqslant \prod_{k} 2^{2-d_i} = 2^{2k-d}.$$

The above two bounds, in view of Theorem 3, imply that

$$\mathbb{E}[(S - \mathbb{E}[S])^d] \geqslant (3e)^{-d} \sum_{k=1}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{k} (pq)^k k^d.$$

Denoting $\sigma^2 = pq$ and using the bound $\binom{n}{k} \geqslant (n/k)^k$, we finally obtain

$$\left(\mathbb{E}[(S - \mathbb{E}[S])^d]\right)^{1/d} \geqslant (3e)^{-1} \cdot \max\{k^{1-\frac{k}{d}} \cdot (n\sigma^2)^{\frac{k}{d}} : 1 \leqslant k \leqslant \lfloor d/2 \rfloor\}, \quad (7)$$

which finishes the proof.

6   Conclusion

This paper introduces novel and simpler formulas for binomial moments, derived by a combinatorial argument coupled with clever algebraic simplification which relies on symmetrization. An important application leads to sharp asymptotics for the growth of central binomial moments. Moreover, explicit algorithms and the working implementation are provided.

A   Implementation

The implementation with examples is also available at https://osf.io/hjscb/.

```python
import itertools
import sympy as sm
from sympy import Symbol
from sympy import polys
from sympy.functions.combinatorial.numbers import stirling, binomial
from sympy.functions.combinatorial.numbers import factorial
from sympy.functions.combinatorial.factorials import FallingFactorial

def multinomial_coef(n,ks):
if n!=sum(ks):
return 0
elif len(ks)==1:
return 1
else:
return binomial(n,ks[0])*multinomial_coef(n-ks[0],ks[1:])

n=Symbol('n')
p=Symbol('p')
q=Symbol('q')
s=Symbol('sigma')
```

Listing 1. Preliminaries

```python
def central_binom_moment(d=2):
''' output as poly in trials number n, success prob. p and q=1-p '''
out = 0
for k in range(1,int(d/2)+1):
tmp = 0
for dks in itertools.product(range(2,d+1),repeat=k):
polyx = multinomial_coef(d,dks)
for dk in dks:
polyx = polyx * (q**(dk-1)-(-p)**(dk-1))
tmp = tmp + polyx
out = out + binomial(n,k) * (p*q)**k * tmp

return out.simplify()
```

Listing 2. Stable Formula for Central Moments

```python
def central_binom_moment_pretty1(d=2):
''' output as polynomial in number of trials and success variance '''

out = central_binom_moment2(d=d).combsimp()
out = out.subs(q,1-p)
if d % 2 == 0:
out = polys.groebner([out,s**2-p*(1-p)],p,n,s,order='lex')[-1]
elif d % 2 == 1:
out = polys.div(out,1-2*p)[0]
out = polys.groebner([out,s**2-p*(1-p)],p,n,s,order='lex')[-1]
out = (1-2*p)*out
return out
```

Listing 3. Variance Formula for Central Moments by Variable Elimination

```python
def central_binom_moment_pretty2(d=2):
''' output as polynomial in number of trials and success variance '''

out = central_binom_moment2(d=d)
if d % 2 == 0:
out = polys.symmetrize(out,[p,q])[0]
elif d % 2 == 1:
out = polys.div(out,q-p)[0]
out = polys.symmetrize(out,[p,q])[0]
out = (1-2*p)*out
```

```
out = out.subs(p+q,1)
out = out.subs(p*q,s**2)
return out
```

Listing 4. Variance Formula by Symmetrization

Acknowledgement

Funding

References

[1] Agresti, A., Coull, B.A.: Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician **52**(2), 119–126 (1998) MR1628435. https://doi.org/10.2307/2685469

[2] Balakrishnan, N., Johnson, N.L., Kotz, S.: A note on relationships between moments, central moments and cumulants from multivariate distributions. Statistics & probability letters **39**(1), 49–54 (1998) MR1649335. https://doi.org/10.1016/S0167-7152(98)00027-3

[3] Bényi, Á., Manago, S.M.: A recursive formula for moments of a binomial distribution. The College Mathematics Journal **36**(1), 68–72 (2005)

[4] Boyadzhiev, K.N.: Close encounters with the Stirling numbers of the second kind. Mathematics Magazine **85**(4), 252–266 (2012) MR2993818. https://doi.org/10.4169/math.mag.85.4.252

[5] Buchberger, B.: Gröbner bases: A short introduction for systems theorists. In: International Conference on Computer Aided Systems Theory, pp. 1–19 (2001). Springer MR1868810. https://doi.org/10.1023/A:1011949421611

[6] Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. The MIT Press (2001) MR1848805

[7] Daniel Paulino, C., Soares, P., Neuhaus, J.: Binomial regression with misclassification. Biometrics **59**(3), 670–675 (2003) MR2004272. https://doi.org/10.1111/1541-0420.00077

[8] De Moivre, A.: Approximatio ad summum terminorum binomii seriem expansi. Proceedings of the Royal Society (1733)

[9] Fraas, J.W., Newman, I.: A binomial test of model fit. Structural Equation Modeling: A Multidisciplinary Journal **1**(3), 268–273 (1994)

[10] Gaudry, P., Schost, É., Thiéry, N.M.: Evaluation properties of symmetric polynomials. International Journal of Algebra and Computation **16**(03), 505–523 (2006) MR2241620. https://doi.org/10.1142/S0218196706003128

[11] Goodson, M.: Most winning A/B test results are illusory. Whitepaper, Qubit (2014)

[12] Graham, R.L., Knuth, D.E., Patashnik, O., Liu, S.: Concrete mathematics: a foundation for computer science. Computers in Physics **3**(5), 106–107 (1989) MR1001562

[13] Griffiths, M.: Raw and central moments of binomial random variables via Stirling numbers. International Journal of Mathematical Education in Science and Technology **44**(2), 264–272 (2013) MR3172571. https://doi.org/10.1080/0020739X.2012.678899

[14] Joarder, A.H., Mahmood, M.: Classroom note: An inductive derivation of Stirling numbers of the second kind and their applications in statistics. Advances in Decision Sciences **1**(2), 151–157 (1997) MR1609730. https://doi.org/10.1155/S1173912697000138

[15] Knoblauch, A.: Closed-form expressions for the moments of the binomial probability distribution. SIAM Journal on Applied Mathematics **69**(1), 197–204 (2008) MR2447945. https://doi.org/10.1137/070700024

[16] Laplace, P.-S.: Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur applications aux probabilités. Memoires de l'Academie des Sciences de Paris (1810)

[17] Little, R.J.: Testing the equality of two independent binomial proportions. The American Statistician **43**(4), 283–288 (1989)

[18] Meurer, A., Smith, C.P., Paprocki, M., Čertík, O., Kirpichev, S.B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J.K., Singh, S., Rathnayake, T., Vig, S., Granger, B.E., Muller, R.P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., Curry, M.J., Terrel, A.R., Roučka, v., Saboo, A., Fernando, I., Kulal, S., Cimrman, R., Scopatz, A.: Sympy: symbolic computing in python. PeerJ Computer Science **3**, 103 (2017). https://doi.org/10.7717/peerj-cs.103

[19] Nguyen, D.: A probabilistic approach to the moments of binomial random variables and application. The American Statistician, **75**(1), 101–103 (2019). MR4203486. https://doi.org/10.1080/00031305.2019.1679257

[20] Patel, J.K., Read, C.B.: Handbook of the Normal Distribution, Second Edition. Statistics: A Series of Textbooks and Monographs. Taylor & Francis (1996). https://books.google.fr/books?id=zoVLF0VF9UYC MR0664762

[21] Potts, R.: Note on the factorial moments of standard distributions. Australian Journal of Physics **6**(4), 498–499 (1953)

[22] Prasolov, V.V.: Polynomials, volume **11** of Algorithms and Computation in Mathematics. Springer (2004) MR2082772. https://doi.org/10.1007/978-3-642-03980-5

[23] Skorski, M.: Johnson-Lindenstrauss transforms with best confidence. In: Belkin, M., Kpotufe, S. (eds.) Proceedings of Thirty Fourth Conference on Learning Theory. Proceedings of Machine Learning Research, vol. 134, pp. 3989–4007. PMLR (2021)

[24] Skorski, M.: Bernstein-type bounds for Beta distribution. Modern Stochastics: Theory and Applications **10**(2), 211–228 (2023) MR4573679

[25] Uspensky, J.V.: Introduction to Mathematical Probability. McGraw-Hill Book Company (1937). https://books.google.at/books?id=aeRQAAAAMAAJ

[26] Yang, J., Liu, Y., Liu, Z., Zhu, X., Zhang, X.: A new feature selection algorithm based on binomial hypothesis testing for spam filtering. Knowledge-Based Systems **24**(6), 904–914 (2011)

[27] Young-Xu, Y., Chan, K.A.: Pooling overdispersed binomial data to estimate event rate. BMC medical research methodology **8**(1), 58 (2008)

[28] Zhou, J.: Introduction to symmetric polynomials and symmetric functions. Lecture Notes for Course at Tsinghua University, available at http://cms.zju.edu.cn/course/cn/Symmetric.pdf (2003)

[29] Zolotukhin, A., Nagaev, S., Chebotarev, V.: On a bound of the absolute constant in the Berry–Esseen inequality for i.i.d. Bernoulli random variables. Modern Stochastics: The-

ory and Applications **5**(3), 385–410 (2018). doi:https://doi.org/10.15559/18-VMSTA113 MR3868547. https://doi.org/10.15559/18-vmsta113