Binomial Sums and Mellin Asymptotics with Explicit Error Bounds: A Case Study

Benjamin Hackl
Department of Mathematics and Scientific Computing, University of Graz, Austria

Stephan Wagner
Institute of Discrete Mathematics, TU Graz, Austria
Department of Mathematics, Uppsala University, Sweden

Abstract

Making use of a newly developed package in the computer algebra system SageMath, we show how to perform a full asymptotic analysis by means of the Mellin transform with explicit error bounds. As an application of the method, we answer a question of Bóna and DeJonge on 132-avoiding permutations with a unique longest increasing subsequence that can be translated into an inequality for a certain binomial sum.

2012 ACM Subject Classification Mathematics of computing → Generating functions; Mathematics of computing → Enumeration; Mathematics of computing → Mathematical software

Keywords and phrases binomial sum, Mellin transform, asymptotics, explicit error bounds, B-terms

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.19

Related Version *Full Version*: `https://arxiv.org/abs/2403.09408`

Supplementary Material *Software (Source Code)*: `https://github.com/behackl/dependent_bterms` archived at `swh:1:dir:da1b484042253b145f0ca5b3d5607c96bcf74771`
Software (Computations): `https://arxiv.org/src/2403.09408/anc/2024-bona-dejonge.ipynb`

Funding *Stephan Wagner*: Supported by the Swedish research council (VR), grant 2022-04030.

1 Introduction

Sums involving binomial coefficients occur frequently in enumerative and analytic combinatorics. For example,

$$\sum_{k=0}^{n} \frac{1}{k+1}\binom{n+k}{n}\binom{n}{k}$$

yields the large Schröder numbers, which count (among other things) many different types of lattice paths and permutations. The sum

$$\sum_{k=0}^{\lfloor n/2\rfloor} \binom{n}{2k}\frac{(2k)!}{2^k k!}$$

counts involutions, or matchings in complete graphs. There is a well-established toolkit for dealing with such sums, based on techniques such as the *(discrete) Laplace method*, the *Stirling approximation* of factorials and binomial coefficients, and the *Mellin transform*. See [5] for a comprehensive account of these and many other tools. While these methods are well known and in some sense mechanical, it is still not straightforward to implement them in a computer as they often involve ad-hoc estimates and careful splitting into different cases/regions of summation that are analyzed separately. Moreover, while a lot of the complications can often be hidden in O-terms, things become more involved when explicit error bounds are desired.

35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 19; pp. 19:1–19:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This paper aims to make a contribution towards building a toolkit for asymptotic analysis in the context of computer algebra, including guaranteed error bounds with explicit constants. The example we use to illustrate the methods is based on a question from a recent paper by Bóna and DeJonge [1]: let a_n be the number of 132-avoiding permutations of length n that have a unique longest increasing subsequence, which is also the number of plane trees with $n+1$ vertices with a single leaf at maximum distance from the root, or the number of Dyck paths of length $2n$ with a unique peak of maximum height. Moreover, let $p_n = \frac{a_n}{C_n}$, where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number. This can be interpreted as the probability that a 132-avoiding permutation of length n chosen uniformly at random has a unique longest increasing subsequence – equivalently, that a plane tree with $n+1$ vertices has a single leaf at maximum distance from the root, or that a Dyck path of length $2n$ has a unique peak of maximum height.

▶ **Problem 1** (Bóna and DeJonge [1]). *Is it true that the sequence p_n is decreasing for $n \geq 3$?*

While it would obviously be interesting to have a combinatorial proof, it turns out (as we will explain in the following section) that the problem can be translated in a fairly mechanical fashion (using generating functions) into a purely analytic problem: specifically, the inequality

$$F(n) = \sum_{k=1}^{n} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k} < 0 \tag{1}$$

for all $n \geq 5$, where $\sigma(k)$ is the sum of divisors of k. The standard approach to deriving an asymptotic formula for such a sum (cf. [4, Section 5]) involves the following steps:

- Split the sum into "small" and "large" values of k.
- Show that the contribution of large values is negligible.
- Approximate the binomial coefficient $\binom{2n}{n-k}$, e.g. by means of Stirling's formula, for small values of k.
- Turn the sum into an infinite sum, again at the expense of a negligible error term.
- Apply the Mellin transform to obtain an integral representation for the resulting infinite sum.
- Use residue calculus to derive the final asymptotic formula.

As we will see, the problem is complicated in this particular instance by the occurrence of nontrivial cancellations, making precise estimates challenging. The asymptotic formula (that will be proven in this paper)

$$F(n) = \binom{2n}{n}\Big(-\frac{n^2}{8} + \frac{n}{24} + o(n)\Big) \tag{2}$$

shows that the answer to the question of Bóna and DeJonge is affirmative for *sufficiently large n*. However, the o-notation hides an error term that is potentially huge for small values of n, so it is not clear what "sufficiently large" means. In order to show that p_n is increasing for *all* $n \geq 3$, we will have to prove a version of (2) with *explicit error bounds*. To this end, we present a new package for the computer mathematics system SageMath [11] that enhances the core implementation of asymptotic expansions, and in particular the arithmetic with SageMath's analogue of O-terms with explicit error bounds, called B-terms. See Section 3 for a guided tour through the features of our package. We then demonstrate the practical usage of our package in Section 4 in which we derive the desired explicit bounds for $F(n)$.

2 Reducing the problem

One of the possible combinatorial interpretations of the sequence a_n is in terms of lattice paths. Specifically, as it was mentioned before, a_n is the number of Dyck paths of length $2n$ (i.e., lattice paths starting at $(0,0)$ and ending at $(2n,0)$ whose steps are either "up" $(1,1)$ or "down" $(1,-1)$) with a unique peak of maximum height. Such a path can be decomposed into two pieces: before and after the peak. The part before the peak needs to be a path that finishes at its maximum height h (but does not reach it earlier, since the peak is unique), and the path after the peak needs to be a path that starts at its maximum height h and never returns to it (which can also be seen as the reflection of a path that finishes at its maximum height but does not reach it earlier). Such paths were analyzed in [2] and [8]. Specifically, [8, Proposition 2.1] states that the probability that a simple symmetric random walk of length n never drops below 0 and finishes at its maximum height h (which can also be reached earlier) is precisely

$$2[z^{n+1}]\frac{1}{U_{h+1}(1/z)},$$

where U_{h+1} is the Chebyshev polynomial of the second kind of degree $h+1$. A path that finishes at its maximum height h without reaching that height before is obtained from a path that finishes at its maximum height $h-1$ by adding one more step up. Since every path of length n has probability 2^{-n} to occur under the model of a simple symmetric random walk, it follows that the (ordinary) generating function for paths of maximum height h that finish at the maximum and do not reach it earlier is

$$\sum_{n\geq 0} 2^n x^{n+1}\cdot 2[z^{n+1}]\frac{1}{U_h(1/z)} = \frac{1}{U_h(1/(2x))}.$$

For example,

$$\frac{1}{U_3(1/(2x))} = \frac{x^3}{1-2x^2} = x^3 + 2x^5 + 4x^7 + \cdots$$

is the generating function for paths that finish at their maximum height 3 and do not reach this height before the final step. The formula is even true for $h=0$: $\frac{1}{U_0(1/(2x))} = 1$ is indeed the correct generating function in this case.

Since the paths we are interested in can be seen as pairs of paths that finish at their maximum height and do not reach this height before, we find that the generating function of a_n is

$$A(x) = \sum_{n=0}^{\infty} a_n x^{2n} = \sum_{h=0}^{\infty} \frac{1}{U_h(1/(2x))^2}.$$

We can simplify the expression by means of the substitution $x = \frac{\sqrt{t}}{1+t}$. Note that this yields $\frac{1}{2x} = \frac{1+t}{2\sqrt{t}} = \cosh(\frac{1}{2}\log t)$. Since $U_h(\cosh w) = \frac{\sinh((h+1)w)}{\sinh w}$, this implies that

$$U_h(1/(2x)) = \frac{\sinh(\frac{h+1}{2}\log t)}{\sinh(\frac{1}{2}\log t)} = t^{-h/2}\cdot\frac{1-t^{h+1}}{1-t}.$$

Thus

$$A(x) = \sum_{h=0}^{\infty} \frac{t^h(1-t)^2}{(1-t^{h+1})^2}.$$

Now we can obtain an alternative expression for a_n by applying Cauchy's integral formula to the generating function $A(x)$. For suitable contours $\mathcal{C}$ and $\mathcal{C}'$ around 0, we have

$$\begin{aligned} a_n &= \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{A(\sqrt{z})}{z^{n+1}}\,dz = \frac{1}{2\pi i}\oint_{\mathcal{C}'} \sum_{h=0}^{\infty} \frac{t^h(1-t)^2}{(1-t^{h+1})^2} \cdot \frac{(1+t)^{2n+2}}{t^{n+1}} \cdot \frac{(1-t)\,dt}{(1+t)^3} \\ &= \sum_{h=0}^{\infty} \frac{1}{2\pi i}\oint_{\mathcal{C}'} \frac{(1-t)^3(1+t)^{2n-1}}{t^{n+1-h}(1-t^{h+1})^2}\,dt, \end{aligned}$$

using the substitution $\sqrt{z} = \frac{\sqrt{t}}{1+t}$ (or equivalently $z = \frac{t}{(1+t)^2}$). It follows that

$$\begin{aligned} a_n &= \sum_{h=0}^{\infty} [t^{n-h}] \frac{(1-t)^3(1+t)^{2n-1}}{(1-t^{h+1})^2} = [t^{n+1}] \sum_{h=0}^{\infty} \frac{t^{h+1}}{(1-t^{h+1})^2}(1-t)^3(1+t)^{2n-1} \\ &= [t^{n+1}] \sum_{k=1}^{\infty} \sigma(k) t^k (1-t)^3(1+t)^{2n-1} = \sum_{k=1}^{\infty} \sigma(k) [t^{n+1-k}](1-t)^3(1+t)^{2n-1} \\ &= \sum_{k=1}^{n+1} \sigma(k) \left(\binom{2n-1}{n+1-k} - 3\binom{2n-1}{n-k} + 3\binom{2n-1}{n-1-k} - \binom{2n-1}{n-2-k} \right) \\ &= \sum_{k=1}^{n+1} \frac{4k\sigma(k)(2k^2-3n-2)(2n-1)!}{(n+1-k)!(n+1+k)!}. \end{aligned}$$

We remark here that the identity $[t^a](1+t)^b = \binom{b}{a}$ that we are using even remains true for negative a or for $a > b$ if the binomial coefficient is considered to be 0 then. The manipulation in the final step is consistent with this.

Problem 1 is equivalent to the inequality $C_{n+1}a_n > C_n a_{n+1}$ for $n \geq 3$, and since $C_{n+1} = \frac{4n+2}{n+2} C_n$, it can also be expressed as $(4n+2)a_n > (n+2)a_{n+1}$. Hence we are left to consider the inequality

$$\sum_{k=1}^{n+1} \frac{4k\sigma(k)(2k^2-3n-2)(4n+2)(2n-1)!}{(n+1-k)!(n+1+k)!} > \sum_{k=1}^{n+2} \frac{4k\sigma(k)(2k^2-3n-5)(n+2)(2n+1)!}{(n+2-k)!(n+2+k)!},$$

which reduces to

$$\sum_{k=1}^{n+2} \frac{8k\sigma(k)(k^2-3n-4)(2k^2-n-2)(2n+1)(2n-1)!}{(n+2-k)!(n+2+k)!} < 0$$

after some manipulations. After multiplication by

$$n(n+1)(n+2)(2n+3) = \frac{(2n)(2n+2)(2n+3)(2n+4)}{8},$$

this can be expressed as

$$\sum_{k=1}^{n+2} k\sigma(k)(k^2-3n-4)(2k^2-n-2)\binom{2n+4}{n+2-k} < 0.$$

Finally, replacing $n+2$ by n, what we have to prove in order to settle Problem 1 is that

$$\sum_{k=1}^{n} k\sigma(k)(k^2-3n+2)(2k^2-n)\binom{2n}{n-k} < 0$$

for all $n \geq 5$, which is precisely (1).

3 B-terms and asymptotics with explicit error bounds

In this section, we provide the necessary background on B-terms and their software implementation. We base our work on the module for computing with asymptotic expansions [7] in SageMath [11]. While this module presently also offers some basic support for B-terms, we have extended its capabilities to add support for computations involving an additional monomially bounded variable (e.g., k with $n^\alpha \leq k \leq n^\beta$ for some $0 \leq \alpha < \beta$ where $n \to \infty$), as well as Taylor expansions with explicit error bounds. These improvements are not yet included in the module directly, but can be made available to your local installation of SageMath simply by running

```
sage -pip install dependent_bterms
```

from your terminal. Alternatively, the module can be installed by executing a cell containing

```
!pip install dependent_bterms
```

from within a SageMath Jupyter notebook.

We will now briefly walk through the core functionalities of our toolbox. The central interface is the function

`AsymptoticRingWithDependentVariable`,

which generates a suitable parent structure for our desired asymptotic expansions. Listing 1 demonstrates how it is used to instantiate the structure that will be used throughout the following examples. We consider $1 = n^0 \leq k \leq n^{4/7}$, i.e., $\alpha = 0$ and $\beta = 4/7$.

Listing 1 Setup of the modified `AsymptoticRing`.

```
sage: import dependent_bterms as dbt
sage: AR, n, k = dbt.AsymptoticRingWithDependentVariable(
....:     'n^QQ', 'k', 0, 4/7, bterm_round_to=3, default_prec=5
....: )
sage: AR
Asymptotic Ring <n^QQ> over Symbolic Ring
```

The arguments passed to the interface are, in order,

- `growth_group` – the (univariate) growth group[1] modeling the desired structure of the asymptotic expansions. For example, `'n^QQ'` represents terms like $42n^{9/13}$ or $O(n^{7/42})$.
- `dependent_variable` – a string representation of the symbolic variable being endowed with asymptotic growth bounds, e.g., `'k'`.
- `lower_bound_power` – a real number $\alpha \geq 0$ representing the power to which the ring's independent variable is raised to obtain the lower monomial power.
- `upper_bound_power` – a real number $\beta > \alpha \geq 0$, analogous to `lower_bound_power`, just for the upper bound,
- `bterm_round_to` – a non-negative integer or `None` (the default), determining the number of floating point digits to which the coefficients of B-terms are automatically rounded. If `None`, no rounding is performed.

[1] See SageMath's documentation on asymptotic expansions and the `AsymptoticRing` for an introduction to the algebraic terminology used here.

Any other keyword arguments (like `default_prec` in Listing 1 above) are passed to the constructor of `AsymptoticRing`.

In this structure, arithmetic with asymptotic expansions in n can be carried out as usual, see Listing 2. The `default_prec` parameter specified above controls the order of the automatic expansions.

Listing 2 Arithmetic and automatic expansions in `AsymptoticRing`.

```
sage: (1 + 3*n) * (4*n^(-7/3) + 42/n + 1)
3*n + 127 + 42*n^(-1) + 12*n^(-4/3) + 4*n^(-7/3)
sage: prod((1 + n^(-j)) for j in srange(1, 10)) * (1 + O(n^(-10)))
1 + n^(-1) + n^(-2) + 2*n^(-3) + 2*n^(-4) + 3*n^(-5) + 4*n^(-6)
+ 5*n^(-7) + 6*n^(-8) + 8*n^(-9) + O(n^(-10))
sage: n / (n - 1)
1 + n^(-1) + n^(-2) + n^(-3) + n^(-4) + O(n^(-5))
```

In the implementation of the `AsymptoticRing` shipped with SageMath, asymptotic expansions internally rely on ordering their summands with respect to the growth of the independent variable(s), regardless of attached coefficients.

In the extension of our `dependent_bterms` module, expansions are aware of the growth range contributed by the dependent variable appearing in coefficients. In fact, in our modified ring, expansions are ordered with respect to the upper bound of the coefficient growth combined with the growth of the independent variable. This explains the – at first glance counterintuitive – ordering of the summands in Listing 3. The individual growth ranges of the summands are printed at the end of the listing.

Listing 3 Arithmetic involving the dependent variable.

```
sage: k*n^2 + O(n^(3/2)) + k^3*n
k^3*n + k*n^2 + O(n^(3/2))
sage: for summand in expr.summands.elements_topological():
....:     print(f"{summand}␣->␣{summand.dependent_growth_range()}")
O(n^(3/2)) -> (n^(3/2), n^(3/2))
k*n^2 -> (n^2, n^(18/7))
k^3*n -> (n, n^(19/7))
```

Automatic power series expansion (with an O-term error) also works natively in our modified ring, see Listing 4. Observe that the error term $O(n^{-15/7})$ would actually be able to *partially absorb* some of the terms in the automatic expansion like $(k/2 + 1/6)n^{-3}$. This partial absorption is, however, not carried out automatically due to performance reasons. Using the `simplify_expansion` function included in our module expands the symbolic coefficients and enables the error terms to carry out all allowed (partial) absorptions.

Listing 4 Automatic expansions and manual simplifications.

```
sage: auto_expansion = exp((1 + k)/n)
sage: auto_expansion
1 + (k + 1)*n^(-1) + (1/2*(k + 1)^2)*n^(-2) + (1/6*(k + 1)^3)*n^(-3)
+ (1/24*(k + 1)^4)*n^(-4) + O(n^(-15/7))
sage: dbt.simplify_expansion(auto_expansion)
1 + (k + 1)*n^(-1) + (1/2*k^2 + k + 1/2)*n^(-2)
+ (1/6*k^3 + 1/2*k^2)*n^(-3) + 1/24*k^4*n^(-4) + O(n^(-15/7))
```

Now let us turn to the core feature of our extension: B-terms. In a nutshell, B-terms are O-terms that come with an explicitly specified constant and validity point. For example, the term $B_{n\geq 10}(42n^3)$ represents an error term that is bounded by $42n^3$ for $n \geq 10$.

Listing 5 demonstrates basic arithmetic with B-terms. It is worth spending a moment to understand how the resulting constants are determined. In the first example, the B-term $B_{n\geq 10}(5/n)$ absorbs the exact term $3/n^2$ of weaker growth. It does so by automatically estimating $\frac{3}{n^2} \leq \frac{3}{10n}$ (as the term is valid for $n \geq 10$) and then directly absorbing the upper bound; $\frac{53}{10} = 5 + \frac{3}{10}$.

The same mechanism happens in the second example. In order to avoid the (otherwise rapid) accumulation of complicated symbolic expressions in the automatic estimates, we have specified (via the `bterm_round_to`-parameter that we have set to 3) that B-terms should automatically be rounded to three floating point digits. This is why the constant is given as $\lceil (1 + 10^{-1/3}) \cdot 10^3 \rceil \cdot 10^{-3} = \frac{293}{200}$.

Listing 5 Arithmetic with B-terms and the dependent variable.

```
sage: 7*n + AR.B(5/n, valid_from=10) + 3/n^2
7*n + B(53/10*n^(-1), n >= 10)
sage: AR.B(1/n, valid_from=10) + n^(-4/3)
B(293/200*n^(-1), n >= 10)
sage: AR.B(3*k^2/n^3, valid_from=10) + (1 - 2*k + 3*k^2 - 4*k^3)/n^5
B(3373/1000*abs(k^2)*n^(-3), n >= 10)
```

The third example in Listing 5 illustrates arithmetic involving the dependent variable, which requires additional care. With $1 \leq k \leq n^{4/7}$ in place, the growth of the given B-term ranges from $\Theta(n^{-3})$ to $\Theta(n^{-13/7})$. The growth of the explicit term that is added ranges from $\Theta(n^{-5})$ to $\Theta(n^{-23/7})$. In this setting, we consider the explicit term to be of weaker growth, as the lower bound of the B-term is stronger than the lower bound of the explicit term, and likewise for the upper bound. Thus we may let the B-term absorb it. We do so by first estimating

$$\left| \frac{1 - 2k + 3k - 4k^3}{n^5} \right| \leq \frac{(1 + 2 + 3 + 4)k^3}{n^5} = \frac{10k^3}{n^5}.$$

As the power of k in this bound is larger than the maximal power of k in the B-term, we may not yet proceed as above (otherwise we would increase the upper bound of the B-term, which we must avoid). Instead, we use first use $k \leq n^{4/7}$, followed by $n \geq 10$, to obtain

$$\frac{10k^3}{n^5} \leq \frac{10k^2 n^{4/7}}{n^5} = \frac{10k^2}{n^{31/7}} \leq \frac{10k^2}{10^{10/7} \cdot n^3} = 10^{-3/7} \cdot \frac{k^2}{n^3},$$

which the B-term can now absorb directly. Hence the value of the B-term constant is determined by $\lceil (3 + 10^{-3/7}) \cdot 10^3 \rceil \cdot 10^{-3} = \frac{3373}{1000}$.

Finally, our module also provides support for B-term bounded Taylor expansion (again, also involving the dependent variable) in form of the `taylor_with_explicit_error` function. An example is given in Listing 6: we first obtain a Taylor expansion of $f(t) = (1 - t^2)^{-1}$ around $t = (1 + k)/n + B_{n\geq 10}(k^3/n^3)$. Using the `simplify_expansion` function rearranges the terms and lets the B-term absorb coefficients (partially) as far it is able to. Observe that it may happen that the attempted simplification produces summands with a smaller upper growth bound that the implementation cannot absorb ($B_{n\geq 10}(k^3/n^3)$ vs. n^{-2} in this case). The expansion is still correct; just not as compact as it could be. We can also use the `simplify_expansion` function with the `simplify_bterm_growth` parameter set to `True` to collapse the dependent variables in all B-terms by replacing them with their upper bounds, resulting in a single "absolute" B-term.

Listing 6 B-term bounded Taylor expansions.

```
sage: arg = (1 + k)/n + AR.B(k^3/n^3, valid_from=10)
sage: ex = dbt.taylor_with_explicit_error(
....:     lambda t: 1/(1 - t^2), arg,
....:     order=3, valid_from=10)
sage: ex
1 + ((k + 1)^2)*n^(-2)
+ B((abs(7351/250*k^3 + 30*k^2 + 30*k + 10))*n^(-3), n >= 10)
sage: dbt.simplify_expansion(ex)
1 + k^2*n^(-2)
+ B((abs(7351/250*k^3 + 30*k^2 + 30*k + 10))*n^(-3), n >= 10)
+ (2*k + 1)*n^(-2)
sage: dbt.simplify_expansion(ex, simplify_bterm_growth=True)
1 + k^2*n^(-2) + B(41441/1000*n^(-9/7), n >= 10)
```

4 Asymptotic analysis

In the following, we provide the steps of the analysis of the sum $F(n)$, aided by the software package that was presented in the previous section. We will verify (1) for $n \geq N = 10000$ by means of an asymptotic analysis with explicit error terms. For $n < N$, one can verify the inequality with a computer by determining $F(n)$ explicitly in all cases.

All computations carried out in this section can be found in the SageMath notebook located at

`https://arxiv.org/src/2403.09408/anc/2024-bona-dejonge.ipynb`,

and a corresponding static version (containing computations and results) is available at

`https://arxiv.org/src/2403.09408/anc/2024-bona-dejonge.html`.

4.1 Approximating the binomial coefficients

It is useful to divide the entire sum by $\binom{2n}{n}$ and to approximate the quotient. Note that we have

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} = \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)} = \frac{n}{n-k}\prod_{j=1}^{k}\frac{n-j}{n+j} = \frac{n}{n-k}\prod_{j=1}^{k}\frac{1-j/n}{1+j/n}.$$

This can be rewritten as

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} = \frac{n}{n-k}\exp\Bigl(\sum_{j=1}^{k}\log(1-j/n)-\log(1+j/n)\Bigr) = \frac{n}{n-k}\exp\Bigl(-\sum_{j=1}^{k}\sum_{\substack{r=1\\ r\text{ odd}}}^{\infty}\frac{2j^r}{rn^r}\Bigr), \tag{3}$$

an expression that will also be used later. It follows from it that

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \leq \frac{n}{n-k}\exp\Bigl(-\sum_{j=1}^{k}\frac{2j}{n}\Bigr) \leq \frac{n}{n-k}\exp\Bigl(-\frac{k^2}{n}\Bigr). \tag{4}$$

For small enough k, we can also obtain an asymptotic expansion. This will be discussed later.

4.2 The tails

In order to replace the binomial coefficient by a simpler expression that is amenable to a Mellin analysis, we first have to handle the tails of the sum. For this purpose, we require an explicit bound for the divisor function $\sigma(k)$ in form of a constant $A > 0$ such that

$$\sigma(k) \le A \cdot k \cdot \log\log n \tag{5}$$

for $1 \le k \le n$ when $n \ge N$. Assume temporarily that $N \le k \le n$. Then, using an inequality due to Robin [10],

$$\frac{\sigma(k)}{k} \le e^{\gamma} \log\log k + \frac{0.6483}{\log\log k} \le e^{\gamma}\log\log n + \frac{0.6483}{\log\log N} \le \left(e^{\gamma} + \frac{0.6483}{\log\log N}\right)\log\log n.$$

For $N = 10000$, we can choose $A = 52/25 \ge e^{\gamma} + 0.6483/\log\log N$, and we can let the computer verify that (5) also holds for $1 \le k \le n = N$.

For $k > \frac{n}{2}$, (4) combined with the fact that the binomial coefficients $\binom{2n}{n-k}$ are decreasing in k gives us

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \le 2e^{-n/4}.$$

So we have

$$\begin{aligned}
0 &\le \frac{1}{\binom{2n}{n}} \sum_{n/2<k\le n} k\sigma(k)(k^2-3n+2)(2k^2-n)\binom{2n}{n-k} \\
&\le \sum_{n/2<k\le n} 4Ak^6(\log\log n)e^{-n/4} \\
&\le (A\log\log n)n^7 e^{-n/4},
\end{aligned}$$

since it is easily verified that $\sum_{n/2<k\le n} k^6 \le \frac{n^7}{4}$ for $n > 5$. Thus, the contribution of the sum in this range is

$$\frac{1}{\binom{2n}{n}} \sum_{n/2<k\le n} k\sigma(k)(k^2-3n+2)(2k^2-n)\binom{2n}{n-k} = B_{n\ge N}\Big(\frac{52}{25}e^{-n/4}n^7\log\log n\Big). \tag{6}$$

Next, fix a constant $\alpha \in (\frac{1}{2}, \frac{3}{4})$; the precise value is in principle irrelevant if one is only interested in an asymptotic formula. However, for our computations with explicit error bounds it is advantageous to take a value close to $\frac{3}{4}$, so we choose $\alpha = \frac{7}{10}$. We bound the sum over all $k \in [n^{\alpha}, n/2]$. Here, we have

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \le 2e^{-k^2/n}$$

by (4), thus (assuming that N is large enough that $k^2 \ge n^{2\alpha} \ge 3n$ whenever $n \ge N$, which we can easily verify for $N = 10000$ and $\alpha = 7/10$)

$$\begin{aligned}
0 &\le \frac{1}{\binom{2n}{n}} \sum_{n^{\alpha}\le k\le n/2} k\sigma(k)(k^2-3n+2)(2k^2-n)\binom{2n}{n-k} \\
&\le \sum_{n^{\alpha}\le k\le n/2} 4Ak^6(\log\log k)e^{-k^2/n} \\
&\le (4A\log\log n) \sum_{n^{\alpha}\le k\le n/2} k^6 e^{-k^2/n}.
\end{aligned}$$

The function $t \mapsto t^6 e^{-t^2/n}$ is decreasing for $t \geq \sqrt{3n}$, thus in particular for $t \geq n^\alpha$ under our assumptions. This implies that (by a standard estimate for sums in terms of integrals)

$$\sum_{n^\alpha \leq k \leq n/2} k^6 e^{-k^2/n} \leq n^{6\alpha} e^{-n^{2\alpha-1}} + \int_{n^\alpha}^\infty t^6 e^{-t^2/n}\, dt.$$

The integral can be estimated by elementary means: for $T = n^\alpha$,

$$\int_T^\infty t^6 e^{-t^2/n}\, dt \leq \frac{1}{T}\int_T^\infty t^7 e^{-t^2/n}\, dt = \frac{n}{2T}\big(6n^3 + 6n^2T^2 + 3nT^4 + T^6\big)e^{-T^2/n}.$$

For large enough $n \geq N$, this is negligibly small. This can be quantified with the help of some explicit computations with B-terms. We find that

$$\frac{1}{\binom{2n}{n}} \sum_{n^\alpha \leq k \leq n/2} \binom{2n}{n-k} k\sigma(k)(k^2-3n+2)(2k^2-n) = B_{n\geq N}\Big(\frac{25073}{5000} e^{-n^{2/5}} n^{\frac{9}{2}} \log\log n\Big). \tag{7}$$

4.3 Approximating the summands

So we are left with the sum over $k < n^\alpha$. Here, we can expand the exact expression in (3): this can be done by cutting the sum over r at some point (we choose the cutoff at $R = 9$) and estimating

$$0 \leq \sum_{j=1}^{k} \sum_{\substack{r=R\\ r \text{ odd}}}^{\infty} \frac{2j^r}{rn^r} \leq \sum_{j=1}^{k} \frac{2j^R}{Rn^R(1-j^2/n^2)}$$

by means of a geometric series (observe that the factor $1 - j^2/n^2$ in the denominator stems from the fact that we are only summing over odd r), and then further

$$\begin{aligned}\sum_{j=1}^{k} \frac{2j^R}{Rn^R(1-j^2/n^2)} &\leq \frac{2}{Rn^R(1-k^2/n^2)}\Big(k^R + \int_0^k t^R\, dt\Big)\\ &= \frac{2}{Rn^R(1-k^2/n^2)}\Big(k^R + \frac{k^{R+1}}{R+1}\Big)\\ &\overset{R=9}{=} B_{n\geq N}\Big(\Big(\frac{239}{10000}k^{10} + \frac{2223}{10000}k^9\Big)n^{-9}\Big),\end{aligned}$$

followed by a Taylor expansion of the exponential multiplied with the expansion of $\frac{n}{n-k}$, cf. (3). The full and sufficiently precise asymptotic expansion can be found in our auxiliary SageMath notebook. It reads

$$\begin{aligned}\frac{\binom{2n}{n-k}}{\binom{2n}{n}} = e^{-k^2/n}\Big(1 - \frac{k^4+k^2}{6n^3} + \frac{k^8}{72n^6} + \frac{k^2}{n^2} - \frac{k^{12}}{1296n^9} - \frac{3k^6}{20n^5} + \frac{k^{16}}{31104n^{12}}\\ + \cdots + B_{n\geq N}\Big(\frac{k^{24}}{10000n^{18}}\Big) + B_{n\geq N}\Big(\frac{9k^{21}}{10000n^{16}}\Big) + \cdots\Big),\end{aligned}$$

where the summands are ordered based on their individual upper growth bound (found from substituting $k = n^\alpha$). The ellipses $\cdots$ indicate terms that are left out as the expression would otherwise be very long. If it were required, this expansion could also be made more precise. Let us now split the expression inside the brackets: let $S(n,k)$ denote the sum of all "exact" terms, and $S_B(n,k)$ the sum of all B-terms. We want to evaluate

$$\sum_{1\leq k<n^\alpha} k\sigma(k)\big(S(n,k) + S_B(n,k)\big)(k^2-3n+2)(2k^2-n)e^{-k^2/n}.$$

Let us first deal with the error estimate: since $|(k^2 - 3n + 2)(2k^2 - n)| \leq 2k^4 + 3n^2$ for all k and n, it suffices to bound

$$\begin{aligned}\sum_{1 \leq k < n^\alpha} & kS_B(n,k)\sigma(k)(2k^4 + 3n^2)e^{-k^2/n} \\ &\leq (A \log\log n) \sum_{1 \leq k < n^\alpha} S_B(n,k)(2k^6 + 3n^2k^2)e^{-k^2/n} \\ &\leq (A \log\log n) \sum_{k \geq 1} S_B(n,k)(2k^6 + 3n^2k^2)e^{-k^2/n}.\end{aligned}$$

For a positive function $f(t)$ that is increasing up to some maximum t_0 and decreasing thereafter, it is well known that $\sum_{k \geq 1} f(k) \leq f(t_0) + \int_0^\infty f(t)\,dt$. This can now be applied to $t \mapsto t^j e^{-t^2/2}$ to find, with the help of computer algebra,

$$\sum_{1 \leq k < n^\alpha} kS_B(n,k)\sigma(k)(2k^4 + 3n^2)e^{-k^2/n} = B_{n \geq N}\Big(\frac{146718899}{10000}\sqrt{n} \log\log n\Big). \tag{8}$$

While this error is not quite as small as those collected so far, for $n = 10000$ it is still only about 26.1% of the eventual main term.

Now we can consider the remaining sum

$$\sum_{1 \leq k < n^\alpha} k\sigma(k)S(n,k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}.$$

To this end, we first add back the terms with $k \geq n^\alpha$ and estimate their sum.

For $k \leq n^{3/4}$ the expansion in $S(n,k)$ can be bounded above, $S(n,k) \leq c_1 \approx 4.372$, and for $k \geq n^{3/4}$ we have $S(n,k) \leq k^{20}/(10000n^{15})$. As for the other factors in our summands, we can bound $(k^2 - 3n + 2)(2k^2 - n)$ from above by $2k^4$. For an estimate of $\sigma(k)$ we use (5) in the range $k < n^{3/4}$, and for the remaining case of $k \geq n^{3/4}$ we use the well-known weaker bound $\sigma(k) \leq k^2$. This leaves us with

$$\begin{aligned}\sum_{n^\alpha \leq k < n^{3/4}} & k\sigma(k)S(n,k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n} \\ &\leq 2Ac_1 \log\log n \sum_{n^\alpha \leq k < n^{3/4}} k^6 e^{-k^2/n} = B_{n \geq N}\left(\frac{12553}{5000} e^{-n^{2/5}} n^{19/4} \log\log n\right),\end{aligned} \tag{9}$$

and

$$\begin{aligned}\sum_{k \geq n^{3/4}} & k\sigma(k)S(n,k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n} \\ &\leq \frac{2}{10000\,n^{15}} \sum_{k \geq n^{3/4}} k^{27} e^{-k^2/n} = B_{n \geq N}\left(\frac{3}{2000} e^{-n^{1/2}} n^{11/2}\right),\end{aligned} \tag{10}$$

where the sums have been bounded using the same integral estimate as before.

4.4 Mellin transform

Having estimated all error terms related to pruning and completing the tails of the sum, we now want to evaluate

$$\sum_{k \geq 1} k\sigma(k)S(n,k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}. \tag{11}$$

This sum is a linear combination of sums of the form

$$\sum_{k\geq 1} k^a n^b \sigma(k) e^{-k^2/n}. \tag{12}$$

In the precision chosen in our accompanying SageMath worksheet, there are 121 such summands, to be precise. Set $t = n^{-1}$, refer to the sum in (12) as $g_{a,b}(t)$ and let $d_{a,b}$ denote the coefficients such that the sum in (11) can be written as $\sum_{a,b} d_{a,b} g_{a,b}(t)$. The Mellin transform (see [4] for a general reference) of $g_{a,b}(t)$ is given by

$$g^*_{a,b}(s) = \int_0^\infty t^{s-1} \sum_{k\geq 1} k^a t^{-b} \sigma(k) e^{-k^2 t}\, dt = \zeta(2s-2b-a-1)\zeta(2s-2b-a)\Gamma(s-b).$$

By the Mellin inversion formula, the original function $g_{a,b}(t)$ can be recovered from its transform via

$$g_{a,b}(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \zeta(2s-2b-a-1)\zeta(2s-2b-a)\Gamma(s-b) t^{-s}\, ds$$

for $c > \frac{a}{2} + b + 1$. We may shift the line of integration further left as long as we collect all corresponding residues. In a first step, we shift the line of integration to $c = 3/4$. While in some summands poles occur as far right as $s = 7/2$, a straightforward computation reveals that, as mentioned in the introduction of this article, nontrivial cancellations take place: after summing all contributions, non-zero residues in the half-plane where $\Re(s) \geq 3/4$ can only be found for $s = 1$ and $s = 2$, where we collect a contribution of

$$\sum_{s_0 \in \{1,2\}} \sum_{a,b} d_{a,b} \operatorname{Res}(g_{a,b}(s), s = s_0) = -\frac{1}{8t^2} + \frac{1}{24t} = -\frac{n^2}{8} + \frac{n}{24}, \tag{13}$$

which proves the asymptotic main term given in (2).

We now need to determine an explicit error bound for these shifted integrals. To do so, we investigate, individually for each summand, how far we can shift the line of integration to the left (in half-integer units) until $\Re(s) = c_{a,b}$ without collecting any further residues.

In a central region of $c_{a,b} + iw$ for $|w| \leq 100$ we use rigorous integration via interval arithmetic to determine the value of the shifted integrals. Outside, for $|w| > 100$, we determine a suitable upper bound of the integrand. For $\Gamma(c_{a,b} + iw)$ where $c_{a,b} > 0$ we use [3, (5.6.9)], and when $c_{a,b} < 0$ we first shift the argument to the right via the functional equation $\Gamma(s) = \frac{1}{s}\Gamma(s+1)$ and then proceed as before. For $\zeta(c_{a,b} + iw)$ we bound the modulus from above by $\zeta(3/2)$ if $c_{a,b} \geq 3/2$. When $c_{a,b} \leq -1/2$ we first apply the reflection formula [3, (25.4.1)]; the resulting factors can all be estimated directly. For the special case of $c_{a,b} = 1/2$ we use the bound proved by Hiary, Patel, and Yang in [9, Theorem 1.1] to obtain

$$|\zeta(1/2 + iw)| \leq 0.618\, t^{1/6} \log t \leq 0.618\, t^{1/2}$$

for $t \geq 100$. Letting a computer collect and combine all of these contributions then yields

$$\begin{aligned} &\left| \sum_{a,b} \frac{1}{2\pi i} \int_{3/4-i\infty}^{3/4+i\infty} g^*_{a,b}(s) t^{-s}\ ds \right| \\ &\qquad \leq \frac{1}{2\pi} \sum_{a,b} n^{c_{a,b}} \int_{-\infty}^{\infty} |g^*_{a,b}(c_{a,b} + iw)|\ dw = B_{n\geq N}\Big(\frac{406531}{100} n^{3/4}\Big). \end{aligned} \tag{14}$$

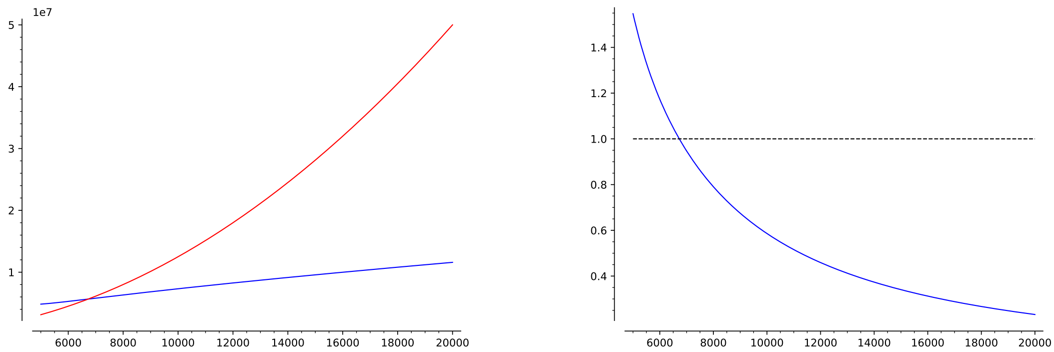

Figure 1 Comparison of the absolute value of the asymptotic main term $-n^2/8 + n/24$ (red) against the accumulated total error (blue) on the left. The right plot depicts the ratio of the error bound to the main term.

5 Conclusion

Throughout Section 4 we have accumulated several explicit error terms. They are given in (6), (7), (8), (9), (10), and (14). Combining them using crude estimates such as $\log\log n \le n^{1/10}$ for $n \ge N$ proves the following theorem.

▶ **Theorem 2.** *For $n \ge 10000$, the binomial sum $F(n)$ satisfies the asymptotic formula*

$$F(n) = \binom{2n}{n}\Big(-\frac{n^2}{8} + \frac{n}{24} + B_{n\ge N}\Big(\frac{38755553}{5000}n^{3/4}\Big)\Big).$$

Observe that for $n = 10000$ the certified error is already only approximately 62.1% of the absolute value of the exact main term. Together with the direct verification for $5 \le n < N$ this settles Problem 1. See Figure 1 for an illustration of the behavior of the total error compared to the main term.

To conclude this paper, we briefly discuss an alternative approach that was kindly pointed out to us by a referee. Recall that the task is to prove the inequality (1), i.e.,

$$F(n) = \sum_{k=1}^{n} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k} < 0.$$

Now one can use the well-known generating function identity

$$\sum_{n=k}^{\infty}\binom{2n}{n-k}x^n = x^k\sum_{m=0}^{\infty}\binom{2m+2k}{m}x^m = x^k\frac{1}{\sqrt{1-4x}}C(x)^{2k},$$

where $C(x) = \frac{1-\sqrt{4x}}{2x}$ is the generating function for the Catalan numbers, see e.g. [6, (5.72)]. This gives an expression for $F(n)$ in terms of coefficients of functions involving $C(x)$. Specifically, we have

$$F(n) = 3n^2\sum_{k=1}^{n}k\sigma(k)\binom{2n}{n-k} - n\sum_{k=1}^{n}k(7k^2+2)\sigma(k)\binom{2n}{n-k} + \sum_{k=1}^{n}2k^3(k^2+2)\sigma(k)\binom{2n}{n-k}$$

$$\begin{aligned}
&= 3n^2[x^n]\frac{1}{\sqrt{1-4x}}\sum_{k=1}^{\infty} k\sigma(k)(xC(x)^2)^k \\
&\quad - n[x^n]\frac{1}{\sqrt{1-4x}}\sum_{k=1}^{\infty} k(7k^2+2)\sigma(k)(xC(x)^2)^k \\
&\quad + [x^n]\frac{1}{\sqrt{1-4x}}\sum_{k=1}^{\infty} 2k^3(k^2+2)\sigma(k)(xC(x)^2)^k \\
&= 3n^2[x^n]\frac{1}{\sqrt{1-4x}}f_1(H(x)) - n[x^n]\frac{1}{\sqrt{1-4x}}f_2(H(x)) + [x^n]\frac{1}{\sqrt{1-4x}}f_3(H(x)),
\end{aligned}$$

where $H(x) = xC(x)^2 = \frac{1-2x-\sqrt{1-4x}}{2x}$, and f_1, f_2, f_3 are given by the series

$$f_1(z) = \sum_{k=1}^{\infty} k\sigma(k)z^k, \quad f_2(z) = \sum_{k=1}^{\infty} k(7k^2+2)\sigma(k)z^k, \quad f_3(z) = \sum_{k=1}^{\infty} 2k^3(k^2+2)\sigma(k)z^k.$$

At the singularity $x = \frac{1}{4}$, $H(x)$ has the expansion

$$1 - 2\sqrt{1-4x} + 2(1-4x) + \cdots,$$

so we need the behavior of $f_1(z), f_2(z), f_3(z)$ around $z = 1$. This can be determined by means of the Mellin transform: setting $z = e^{-t}$, we obtain for instance

$$f_1(e^{-t}) = \sum_{k=1}^{\infty} k\sigma(k)e^{-kt},$$

whose Mellin transform is $\Gamma(s)\zeta(s-1)\zeta(s-2)$. Applying the inverse Mellin transform in the same way as in Section 4.4 (though now with complex parameter t) yields

$$f_1(e^{-t}) = \frac{\pi^2}{3t^3} - \frac{1}{2t^2} + O(t^K)$$

for any positive real K. This and analogous asymptotic formulas for f_2 and f_3 give us the behavior of $f_1(H(x))$, $f_2(H(x))$ and $f_3(H(x))$ at the dominant singularity $\frac{1}{4}$, from which the asymptotic formula (2) can be obtained by means of contour integration and singularity analysis. Carrying all of this out with explicit error terms comes with its own challenges, though, as one now has to deal with complex asymptotics.

References

1. Miklós Bóna and Elijah DeJonge. Pattern avoiding permutations with a unique longest increasing subsequence. *Electron. J. Combin.*, 27(4):Paper No. 4.44, 11, 2020. doi:10.37236/9506.
2. Mireille Bousquet-Mélou and Yann Ponty. Culminating paths. *Discrete Math. Theor. Comput. Sci.*, 10(2):125–152, 2008.
3. *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.1.12 of 2023-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. URL: https://dlmf.nist.gov/.
4. Philippe Flajolet, Xavier Gourdon, and Philippe Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theor. Comput. Sci.*, 144(1-2):3–58, 1995. doi:10.1016/0304-3975(95)00002-E.

5 Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009. doi:10.1017/CBO9780511801655.
6 Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics: a foundation for computer science*. Amsterdam: Addison-Wesley Publishing Group, 2nd ed. edition, 1994.
7 Benjamin Hackl, Clemens Heuberger, and Daniel Krenn. Asymptotic expansions in SageMath. http://trac.sagemath.org/17601, 2015. module in SageMath 6.10.
8 Benjamin Hackl, Clemens Heuberger, Helmut Prodinger, and Stephan Wagner. Analysis of bidirectional ballot sequences and random walks ending in their maximum. *Ann. Comb.*, 20(4):775–797, 2016. doi:10.1007/s00026-016-0330-0.
9 Ghaith A. Hiary, Dhir Patel, and Andrew Yang. An improved explicit estimate for $\zeta(1/2+it)$. *J. Number Theory*, 256:195–217, 2024. doi:10.1016/j.jnt.2023.09.003.
10 G. Robin. Grandes valeurs de la fonction somme des diviseurs et hypothèse de Riemann. *J. Math. Pures Appl. (9)*, 63:187–213, 1984.
11 The SageMath Developers. SageMath, March 2023. doi:10.5281/zenodo.593563.