On the number of words with restrictions on the number of symbols

Verónica Becher Eda Cesaratto

December 23, 2021

Abstract

We show that, in an alphabet of n symbols, the number of words of length n whose number of different symbols is away from $(1-1/e)n$, which is the value expected by the Poisson distribution, has exponential decay in n. We use Laplace's method for sums and known bounds of Stirling numbers of the second kind. We express our result in terms of inequalities.

MSC 2020: 05A05, 05A10, 05A20

Keywords: Poisson distribution; Laplace method for sums; Stirling numbers of the second kind; Combinatorics on words.

Contents

1 Introduction and statement of results

Consider an alphabet of n symbols and let $\chi^{(i)}$ be the number of symbols that appear exactly i times in a word of length m. This can be seen as the allocation of m balls (the positions in a word of length m) in n bins (the n symbols of the alphabet), which determines a total of n^m allocations. When m/n is a fixed constant λ,

$$\frac{1}{n}\chi^{(i)} \text{ converges in probability to } e^{-\lambda}\frac{\lambda^i}{i!},$$

which is the Poisson formula, the proof can be read from [6, Example III.10 and Proposition V.11].

We are interested in the case when the alphabet size n equals the word length m, hence $\lambda = m/n = 1$. The number of symbols that do *not* appear in a word of length n is $\chi^{(0)}$ and its expected value is n/e. Hence, the expected number of different symbols in a word of length n is $n - n/e = (1-1/e)n$. The probability that $\chi^{(0)}$ is equal to j for $j = 0, 1, \ldots, n$ is expressible in terms of Stirling numbers of the second kind: the number $a(n,j)$ of words of length n with *exactly* j different symbols is the number of ways to choose j out of n elements times the number of surjective maps from a set of n positions to a set of j symbols. To make such a surjective map, first partition the set of n elements into j nonempty subsets and, in one of the $j!$ many ways, assign one of these subsets to each element in the set of j elements,

$$a(n,j) = \binom{n}{j} j!\ S_n^{(j)},$$

where

$$S_n^{(j)} = \frac{1}{j!}\sum_{i=0}^{j}(-1)^i\binom{j}{i}(j-i)^n.$$

Notice that

$$\sum_{j=0}^{n} a(n,j) = n^n.$$

Theorem 1 is the main result of this note and shows that in an alphabet of n symbols, the number of words of length n with exactly j symbols, has exponential decay in n when j is away from the value expected by the Poisson distribution. Precisely, Theorem 1 proves that $a(n,j)$, has exponential decay in n when j is away from $(1-1/e)n$. And this implies that for every positive $\varepsilon < 1$,

$$\sum_{n\geq 1} n^{-n}\left(\sum_{j=1}^{(1-1/e-\varepsilon)n} a(n,j) + \sum_{j=(1-1/e+\varepsilon)n}^{n} a(n,j)\right) < \infty.$$

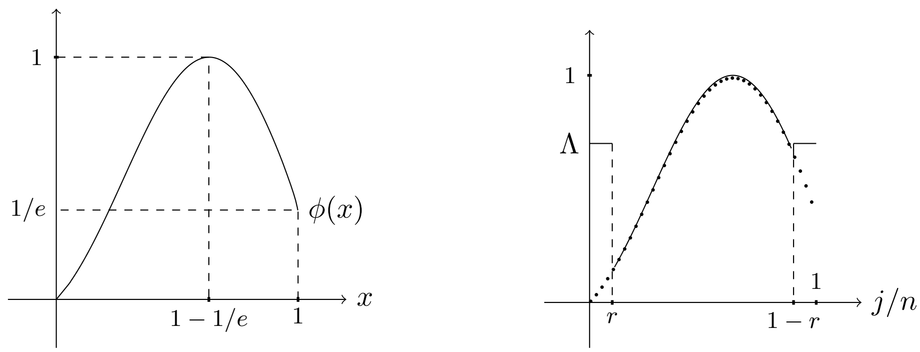

Figure 1: On the left, the graph of $\phi(x)$. On the right, the points are $\sqrt[n]{a(n,j)}n^{-1}$ for $n = 200$ and $j = 0, 5, 10, \ldots, 195, 200$ and the solid line is $\phi(j/n)$ with $r = 0.1$ and $\Lambda \approx 0.701$.

Theorem 1. *There is a function $\phi : (0,1) \mapsto \mathbb{R}$ such that $\phi(x) < 1$ for every $x \neq 1 - 1/e$, positive reals r and Λ both less than 1, and positive constants c and C satisfying the following condition: For every pair n, j of integers with $1 \leq j \leq n$,*

$$\begin{aligned} a(n,j) &\leq \begin{cases} C\sqrt{n}\Lambda^n n^n & , \text{ if } j/n \in [0,r] \cup [1-r,1] \\ C\ \phi(j/n)^n n^n & , \text{ if } j/n \in [r, 1-r] \end{cases} \\ a(n,j) &\geq\ (c/\sqrt{n})\,\phi(j/n)^n n^n\ , \text{ if } j/n \in [r, 1-r]. \end{aligned}$$

Precisely,

$$\begin{aligned} \phi : (0,1) \mapsto \mathbb{R}, &\quad \phi(x) = (e\ln(1+e^{-\delta(x)})^{-1}\varphi(x)e^{-x\delta(x)} \\ \varphi : [0,1] \mapsto \mathbb{R}, &\quad \varphi(x) = x^{-x}(1-x)^{-(1-x)},\ \varphi(0) = \varphi(1) = 1 \\ \delta : (0,1) \mapsto \mathbb{R}, &\quad \delta^{-1}(y) = \frac{1}{(1+e^y)\ln(1+e^{-y})}. \end{aligned}$$

Each of the values c, C, Λ and r in the statement of Theorem 1 can be effectively computed. Figure 1 plots the upper bound of $\sqrt[n]{a(n,j)}n^{-1}$ with the function $\phi(j/n)$ given in Theorem 1.

As a straightforward application of Theorem 1 we obtain the following.

Corollary 2. *For any positive real number ε there exist positive constants c and C and a positive real number Λ strictly less than 1 such that for every positive integers n, ℓ,*

$$\textit{if} \quad |\ell/n - (1-1/e)| \geq \varepsilon \quad \textit{then} \quad (c/\sqrt{n})\ \Lambda^n \leq n^{-n}\sum_{j=1}^{\ell} a(n,j) \leq Cn\sqrt{n}\ \Lambda^n.$$

A tail estimate is a quantification of the rate of decrease of probabilities away from the central part of a distribution. It is known that the tail of a given arbitrary discrete distribution has exponential decay if its probability generating function is analytic on a disk centered on zero and of radius greater than 1 [6, Theorem IX.3,page 627]. Theorem 1 gives, indeed, a tail estimate with exponential decay, but our methods are not analytic.

Our proof of Theorem 1 is elementary except for the estimates for Stirling numbers of the second kind that we use as a black box. We follow the principles of Laplace's method for sums, which is useful for sums of positive terms which increase to a certain point and then decrease. For a general explanation with examples we refer to Flajolet and Sedgewick's book [6, p.761], see [10] for a rigorous application to an hypergeometric-type series. However, we do not use the exp-log transformation to build the approximation function.

Specifically, to prove Theorem 1 we give a smooth function ϕ so that $\phi(j/n)^n$ bounds $a(n,j)n^{-n}$ from above and below (up to multiplicative sequences that increase or decrease slowly). We consider the ratio between j and n. When j is near to 0 or near to n we use the classical upper bound of Stirling numbers of the second kind given by Rennie and Dobson [11]. When j is not near to 0 nor near to n we use Bender's approximation of Stirling numbers of the second kind [2] as a black box. This approximation comes from analytic combinatorics methods and it was initially devised by Laplace, then proved by Moser and Wyman [9] and later sharpened by Bender, see also [8]. Our two choices are motivated by the comparison of bounds on Stirling numbers by Rennie-Dobson [11], Arratia and DeSalvo [1], and also a trivial bound, given in Section 2.

The approach we use in the proof of Theorem 1 was previously used by one of the authors in two different problems. In [3] it is used to estimate $n!\prod_{i=1}^{k} p_i^{j_i s}/j_i!$ where each p_i is the probability of the symbol i in an alphabet of k elements, s is a real number in $(0,1)$ and the integers j_i sum up n and $\sum_{i=1}^{k} i j_i \leq Mn$ for a fixed $M > 1$. In [4, Remark 4.3] the same approach is used to obtain an upper bound for $\binom{n}{j}/j!$ when n is fixed and j varies. Besides, the asymptotic behavior of these quantities when n tends to infinity was studied using a similar technique in [7].

We crossed the problem solved in the present note when studying the set $\mathcal{S}$ of infinite binary sequences with too many of too few, with respect to the expected by the Poisson distribution, different words of length $\lfloor \log n \rfloor$, counted with no overlapping in their initial segment of length $n\lfloor \log n \rfloor$, for infinitely many ns. Corollary 2 allows us to prove that the Lebesgue measure of this set $\mathcal{S}$ is zero, as follows. For simplicity, let n be a power of 2 and let log be the logarithm in base 2. Identify the binary words of length $\log n$ with integers from 0 to $n-1$. Thus, each binary word of length $n \log n$ is identified with a with a word of n integers from 0 to $n-1$. Notice that there are $2^{n \log n} = n^n$ many of these binary words. Corollary 2 assumes an alphabet of n symbols and gives an upper bound for the proportion of words of length n having a number of different symbols away from $(1-1/e)n$, which is the quantity expected by the Poisson distribution. By the identification we made, this yields an upper bound of the proportion of binary words of length $n \log n$ having too many or too few different binary blocks with respect to what is expected by the Poisson distribution. Since this upper bound has exponential decay in n, we can apply Borel-Cantelli lemma to show that the sum, for every n, of these bounds is finite. Consequently, the Lebesgue measure of the set $\mathcal{S}$ is zero. A different proof of this result follows from the metric theorem given by Benjamin Weiss and Yuval Peres in [13] where they show that the set of Poisson generic sequences on a finite alphabet has Lebesgue measure 1. Their proof is probabilistic, with a randomized part and a concentration part.

2 On different bounds on Stirling numbers of second kind

We compare four estimates on Stirling numbers of the second kind $S_n^{(j)}$. When j/n belongs to $(0,1)$, we consider a trivial bound, Rennie and Dobson's bound [11] and Arratia and DeSalvo's bounds [1]. When j/n belongs to a closed interval included in $(0,1)$, we consider Bender's estimate [2]. We start by giving bounds for the binomial coefficients.

2.1 Binomial coefficients

Consider the following bounds for the factorial which are consequence of the classical Stirling's formula for the factorial, see [12],

$$n! = \sqrt{2\pi} n^{n+1/2} e^{-n+r_n}, \quad \frac{1}{12n+1} \leq r_n \leq \frac{1}{12n}.$$

Then, for any $n \geq 1$,

$$\sqrt{2\pi} n^{n+1/2} e^{-n} \leq n! \leq \sqrt{2\pi} e^{1/12} n^{n+1/2} e^{-n}. \tag{1}$$

In the sequel we write $a \approx b$ to indicate that the two numbers a and b coincide up to the precision explicitly indicated, but they may differ in the fractional part that is not exhibited. For example, $\pi \approx 3.14159$. From this approximation of the factorial, we obtain bounds for the binomial coefficient that involve the following functions,

$$\begin{aligned} &\varphi : [0,1] \mapsto \mathbb{R}, \quad \varphi(x) = x^{-x}(1-x)^{-(1-x)}, \ \varphi(0) = \varphi(1) = 1; \qquad (2)\\ &\gamma : (0,1) \mapsto \mathbb{R}, \quad \gamma(x) = (x - x^2)^{1/2} \end{aligned}$$

There exist constants c_0 and C_0 such that for any pair of integers n, j where $n \geq 2$ and $1 \leq j \leq n-1$,

$$\frac{c_0}{\sqrt{n}\gamma(j/n)} \varphi(j/n)^n \leq \binom{n}{j} \leq \frac{C_0}{\sqrt{n}\gamma(j/n)} \varphi(j/n)^n .$$

The constants c_0 and C_0 can be chosen as $c_0 = (\sqrt{2\pi} e^{1/6})^{-1} \approx 0.33$ and $C_0 = e^{1/12}(\sqrt{2\pi})^{-1} \approx 0.43$. From (1), it follows that

$$\binom{n}{j} \leq e^{1/12} (\sqrt{2\pi})^{-1} \left(\frac{n}{j(n-j)} \right)^{1/2} \frac{n^n}{j^j (n-j)^{n-j}}. \tag{3}$$

First, notice that

$$\left(\frac{n}{j(n-j)} \right)^{1/2} = \frac{n^{1/2}}{n(j/n(1-j/n))^{1/2}} = \frac{1}{\sqrt{n}\gamma(j/n)}.$$

Now, we deal with the last factor of (3). The following holds:

$$\frac{n^n}{j^j (n-j)^{n-j}} = \frac{n^n}{n^n (j/n)^j (1-j/n)^{n-j}} = \left((j/n)^{-j/n} (1-j/n)^{-(1-j/n)} \right)^n = \varphi(j/n)^n.$$

This proves the upper bound on the binomial coefficient. The proof of the lower bound is similar, except that the factor $e^{1/12}$ appears twice in the denominator.

Finally, we remark that for any pair of positive integers n, j such that $n \geq 2$ and $1 \leq j \leq n-1$, we have $\min\{j(n-j) : 1 \leq j \leq n-1\} = n-1$ (this value is attained at $j=1$ or $j=n-1$). Also $n-1 \geq n/2$ for $n \geq 2$. Hence,

$$\gamma(j/n) = \left(\frac{j(n-j)}{n^2}\right)^{1/2} \geq \left(\frac{n}{2n^2}\right)^{1/2} = \frac{\sqrt{2}}{2} n^{-1/2}, \text{ and}$$
$$\gamma(j/n) \leq \max\{\gamma(x) : x \in [0,1]\} = \max\{(x-x^2)^{1/2} : x \in [0,1]\} \leq 1/2.$$

Thus, multiplying by $\sqrt{n}$,

$$\sqrt{2}/2 \leq \sqrt{n}\gamma(j/n) \leq (1/2)\sqrt{n},$$

which implies

$$\frac{2}{\sqrt{n}} \leq \frac{1}{\sqrt{n}\gamma(j/n)} \leq \sqrt{2}.$$

We have that $2c_0 \approx 0.67 > 1/2$ and $\sqrt{2}C_0 \approx 0.61 < 1$. This shows the following inequalities, for every positive $n \geq 2$ and every j such that $1 \leq j \leq n-1$,

$$\frac{1}{2\sqrt{n}}\varphi(j/n)^n \leq \frac{2c_0}{\sqrt{n}}\varphi(j/n)^n \leq \binom{n}{j} \leq \sqrt{2}C_0\varphi(j/n)^n \leq \varphi(j/n)^n. \tag{4}$$

2.2 A trivial bound on Stirling numbers the second kind

The simplest upper bound takes just the first term of the alternating sum that defines $S_n^{(j)}$,

$$S_n^{(j)} \leq j^n/j!.$$

This upper bound appears explicitly taking just one term in Bonferroni inequalities, see [5, Section 4.7]. First remark that the upper bound given in (1) for the factorial yields

$$\frac{j^n}{j!} \leq \frac{j^n}{\sqrt{2\pi j}j^j e^{-j}} = \frac{1}{\sqrt{2\pi j}} \frac{n^{n-j}(j/n)^n e^j}{(j/n)^j} = \frac{1}{\sqrt{2\pi j}} \left(n^{1-j/n}(j/n)^{1-j/n}e^{j/n}\right)^n. \tag{5}$$

The same lines together with the lower bound for (1) give a lower bound for $j^n/j!$.

Let $\theta : [0,1] \mapsto \mathbb{R}$,

$$\theta(x) = x^{1-x}e^x. \tag{6}$$

It follows that

$$\frac{1}{e^{1/12}\sqrt{2\pi j}} \left(n^{1-j/n}\theta(j/n)\right)^n \leq j^n/j! \leq \frac{1}{\sqrt{2\pi j}} \left(n^{1-j/n}\theta(j/n)\right)^n.$$

Consequently,

$$S_n^{(j)} \leq \frac{1}{\sqrt{2\pi j}} \left(n^{1-j/n}\theta(j/n)\right)^n. \tag{7}$$

2.3 Rennie and Dobson's bound

The following is the classsical upper bound of Stirling numbers of the second kind given by Rennie and Dobson [11], which holds for every positive n and every j such that $1 \leq j \leq n-1$,

$$S_n^{(j)} \leq \frac{1}{2}\binom{n}{j} j^{n-j}. \tag{8}$$

Let $\eta : [0,1] \mapsto \mathbb{R}$,

$$\eta(x) = x^{1-x}\varphi(x), \tag{9}$$

where φ is defined in (2). Since $j^{n-j} = (n^{1-j/n}(j/n)^{1-j/n})^n$, the bounds on the binomial given in (4) imply

$$\frac{1}{2\sqrt{n}}\left(n^{1-j/n}\eta(j/n)\right)^n \le \binom{n}{j} j^{n-j} \le \left(n^{1-j/n}\eta(j/n)\right)^n. \tag{10}$$

2.4 Arratia and DeSalvo's bound

Arratia and DeSalvo [1, Theorems 5 and 6] give these bounds for $n \ge 3$ and $1 \le j \le n-2$,

$$\begin{aligned} S_n^{(j)} &\le A_5(n,j) \\ S_n^{(j)} &\le A_6(n,j) \end{aligned}$$

where

$$\begin{aligned}
A_5(n,j) :=& \binom{N}{n-j} e^{-2\mu_5(n,j)} \left(1 + e^{2\mu_5(n,j)} D_5(n,j)\right) \\
A_6(n,j) :=& \frac{N^{n-j}}{(n-j)!} e^{-\mu_6(n,j)} \left(1 + e^{\mu_6(n,j)} D_6(n,j)\right) \\
N :=& \binom{n}{2} \\
\mu_5(n,j) :=& \binom{(n-j)}{2}\binom{n}{3} / \binom{N}{2} \\
\mu_6(n,j) :=& \binom{(n-j)}{2} \frac{n(n-1)(4n-5)}{6N^2} \\
d_5(n,j) :=& P + Q + (1-Q)((n-j)-2)(R+S+T) \text{ where} \\
& P := \frac{2\binom{n}{3}}{\binom{N}{2}} \\
& Q := \frac{13 - 12(n-j) + 3(n-j)^2}{\binom{N}{2}} \\
& R := \frac{8\binom{n}{3}}{\binom{N}{2}} \\
& S := \frac{6\binom{n}{4}}{\binom{n}{3}(N-2)} \\
& T := \frac{1}{(N-2)}\left(\frac{5n-11}{4}\right) \\
d_6(n,j) :=& U + 2(V + W + X) \text{ where} \\
& U := \frac{n(n-1)(4n-5)}{6N^2} \\
& V := 4((n-j)-2)\frac{n(n-1)(2n-1)}{6N^2}
\end{aligned}$$

$$W := \frac{3((n-j)-2)n(n-1)}{(4n-5)N}$$
$$Z := \frac{2((n-j)-2)(2n-1)(n+1)}{(4n-5)N}$$
$$D_5(n,j) := \min\{d_5(n,j), 2\mu_5(n,j)d_5(n,j), 1\}$$
$$D_6(n,j) := \min\{(d_6(n,j), 2\mu_6(n,j)d_6(n,j), 1\}.$$

The goal of this section is to give bounds on $A_5(n,j)$ and $A_6(n,j)$ from below and above. They are displayed in Proposition 5, and the proofs of these bounds rely on Lemma 3 and Lemma 4. In the sequel when we write $A_{5,6}$ we denote two statements, one about A_5 and one about A_6. Similarly for $D_{5,6}$ and $\mu_{5,6}$

Lemma 3. *For any $n \geq 3$ and $1 \leq j \leq n-2$,*

$$\frac{1}{2n^2} \leq e^{-\mu_{5,6}(n,j)}\left(1 + e^{\mu_{5,6}(n,j)}D_{5,6}(n,j)\right) \leq 2.$$

Proof. By definition, $D_{5,6}(n,j) \leq 1$, and clearly $\mu_{5,6}(n,j) \geq 0$, then

$$e^{-\mu_{5,6}(n,j)}\left(1 + e^{\mu_{5,6}(n,j)}D_{5,6}(n,j)\right) = e^{-\mu_{5,6}(n,j)} + D_{5,6}(n,j) \leq 2.$$

To obtain a lower bound for $e^{-\mu_{5,6}(n,j)} + D_{5,6}(n,j)$, it suffices to bound the quantities $D_{5,6}(n,j)$.

Lower bound for $D_5(n,j)$. First we consider $\mu_5(n,j)$. The equality

$$\binom{N}{2} = \frac{1}{2}\frac{n(n-1)}{2}\left(\frac{n(n-1)}{2} - 1\right) = \frac{(n+1)n(n-1)(n-2)}{8} \tag{11}$$

yields

$$\mu_5(n,j) = \frac{2}{3}\frac{(n-j)(n-j-1)}{n+1}.$$

This quantity, $\mu_5(n,j)$, takes its minimum when $j = n-2$. It follows that

$$\mu_5(n,j) \geq \frac{1}{n}.$$

We claim that the quantity $d_5(n,j) = P + Q + (1-Q)((n-j)-2)(R+S+T)$ satisfies that

$$d_5(n,j) \geq P \quad \text{for any} \quad 1 \leq j \leq n-2.$$

It is clear that Q and $((n-j)-2)(R+S+T)$ are nonnegative. It only remains to prove that $1-Q$ is nonnegative. In fact, $Q \leq 1/2$ for the values of n and j under consideration. To prove that, first, we complete squares and apply (11); then we take $j = 1$, and finally, we maximize over over n to obtain the last inequality,

$$Q = 8\frac{3((n-j)-2)^2 + 1}{(n+1)n(n-1)(n-2)} \leq 8\frac{3(n-3)^2 + 1}{(n+1)n(n-1)(n-2)} \leq \frac{1}{2}.$$

Finally,

$$d_5(n,j) \geq P = 2\frac{\binom{n}{3}}{\binom{N}{2}} = \frac{8}{3(n+1)} \geq \frac{1}{n}.$$

From the last lower bound and the bound $\mu_5(n,j) \geq 1/n$, we get the following:

$$D_5(n,j) := \min\left(d_5(n,j), 2\mu_5(n,j)d_5(n,j), 1\right) \geq \min\left(\frac{1}{n}, \frac{2}{n^2}, 1\right) \geq \frac{1}{n^2}.$$

Lower bound for $D_6(n,j)$. First we consider $\mu_6(n,j)$. By definition, $N = \binom{n}{2}$. It turns out that for every $n \geq 3$ and j such that $1 \leq j \leq n-2$,

$$\mu_6(n,j) = \binom{(n-j)}{2}\frac{n(n-1)(4n-5)}{6N^2} = \frac{1}{3}\frac{(n-j)(n-j-1)(4n-5)}{n(n-1)} \geq \frac{1}{n}.$$

All the terms involved in the sum defining $d_6(n,j)$ are non-negative. Hence,

$$d_6(n,j) \geq U = \frac{n(n-1)(4n-5)}{6N^2} = \frac{2}{3}\frac{4n-5}{n(n-1)} \geq \frac{2}{3n}.$$

Finally, the following holds and completes the proof of this lemma.

$$D_6(n,j) := \min\left(d_6(n,j), 2\mu_6(n,j)d_6(n,j), 1\right) \geq \min\left(\frac{2}{3n}, \frac{4}{3n^2}, 1\right) \geq \frac{1}{2n^2}.$$

□

Let κ be the map from $[0,1]$ to $\mathbb{R}$ given by

$$\kappa(x) = (e/2)^{1-x}(1-x)^{-(1-x)}, \quad \kappa(1) = 1. \tag{12}$$

Lemma 4. *For any $n \geq 3$ and j with $1 \leq j \leq n-2$, the following holds*

$$\frac{e^{-2}}{2\sqrt{n(n-1)}}\left(n^{1-j/n}\kappa(j/n)\right)^n \leq \binom{N}{n-j} \leq \left(n^{1-j/n}\kappa(j/n)\right)^n, \tag{13}$$

$$\frac{1}{4\sqrt{2\pi}\sqrt{n}}\left(n^{1-j/n}\kappa(j/n)\right)^n \leq \frac{N^{n-j}}{(n-j)!} \leq \frac{1}{\sqrt{2\pi}}\left(n^{1-j/n}\kappa(j/n)\right)^n. \tag{14}$$

Proof. We start by proving inequality (13). With the bounds given for the binomial coefficients in (4), the following holds

$$\frac{1}{2\sqrt{N}}\varphi((n-j)/N)^N \leq \binom{N}{n-j} \leq \varphi((n-j)/N)^N \tag{15}$$

with $\varphi(x) = x^{-x}(1-x)^{-(1-x)}$, for any $n \geq 3$ and $1 \leq j \leq n-2$. The expression $\varphi((n-j)/N)^N$ has two factors, the first one corresponds to x^{-x} and the second one corresponds to $(1-x)^{1-x}$. We replace N by $n(n-1)/2$ only in the first factor. The exponent of the second factor is multiplied and divided by $\frac{N}{n-j}$. This leads to the following equality

$$\varphi\left(\frac{(n-j)}{N}\right)^N = n^{n-j}\left(1-\frac{1}{n}\right)^{n-j}\left(2\left(1-\frac{j}{n}\right)\right)^{-(n-j)}\left(\left(1-\frac{n-j}{N}\right)^{\frac{N}{n-j}}\right)^{-(n-j)+\frac{(n-j)^2}{N}}.$$

The right side of this equality is the product of four factors. We leave the first and the third as they are. We deal with the second and the fourth. We define $b(n,j)$ and $c(n,j)$ as follows

$$b(n,j) = \left(\left(1-\frac{n-j}{N}\right)^{\frac{N}{n-j}}\right)^{-(n-j)+\frac{(n-j)^2}{N}} \quad \text{and} \quad c(n,j) = (1-1/n)^{n-j}b(n,j). \tag{16}$$

The factor $(1-1/n)^{n-j}$ satisfies

$$e^{-1} \leq \left(1-\frac{1}{n}\right)^{n-1} \leq \left(1-\frac{1}{n}\right)^{n-j} \leq 1. \tag{17}$$

The right hand side inequality is due to the fact that $1-1/n \leq 1$. The left hand side inequality is due to the fact that $(1-1/n)^{n-1}$ decreases towards its limit as $n \to \infty$.

We study $b(n,j)$, defined in (16). First, we use the classical inequality

$$-x-x^2 \leq \ln(1-x) \leq -x \quad (0 < x \leq 2/3).$$

After multiplying by $1/x$ and taking powers, we get

$$e^{-1-x} \leq (1-x)^{1/x} \leq e^{-1} \quad (0 < x \leq 2/3). \tag{18}$$

Observe that, for $j \geq 1$,

$$\frac{n-j}{N} = \frac{2(n-j)}{n(n-1)} \leq \frac{2}{n}.$$

Notice that $0 < (n-j)/N \leq 2/3$ since $n \geq 3$. This allows us to replace x by $(n-j)/N$ in (18). It turns out that

$$e^{-1-(\frac{n-j}{N})} \leq \left(1-\frac{n-j}{N}\right)^{\frac{N}{n-j}} \leq e^{-1}.$$

To obtain $b(n,j)$, consider the previous expressions to the power $-(n-j)+(n-j)^2/N$. With our bound on $(n-j)/N$, the exponent of the left hand side satisfies

$$\left(-1-\frac{n-j}{N}\right)\left(-(n-j)+\frac{(n-j)^2}{N}\right) = (n-j)-\frac{(n-j)^3}{N^2} \geq (n-j)-\frac{4}{n^2}(n-1) \geq (n-j)-1.$$

Finally,

$$e^{-1}e^{(n-j)} \leq b(n,j) \leq e^{(n-j)-\frac{(n-j)^2}{N}} \leq e^{(n-j)}. \tag{19}$$

From inequalities (17) and (19), it follows that $c(n,j)e^{-(n-j)}$ takes values in $[e^{-2},1]$ for which

$$\varphi\left(\frac{(n-j)}{N}\right)^N = c(n,j)n^{n-j}\left(2\left(1-\frac{j}{n}\right)\right)^{-(n-j)} = c(n,j)e^{-(n-j)}\left(n^{1-j/n}\kappa(j/n)\right)^n.$$

To end the proof of inequality (13) consider inequality (15) together with the fact that $N = n(n-1)/2$.

Proof of inequality (14). Approximating the factorial by (1); extracting n as a common factor in $(n-j)^{n-j}$, in $(n-1)^{n-j}$, and in $(n-1)^{n-j}$; and writing the final expression as an n-th power (similarly to what it is done in (5)), we get

$$\frac{N^{n-j}}{(n-j)!} \leq \frac{n^{n-j}(n-1)^{n-j}}{2^{n-j}}\frac{e^{n-j}}{\sqrt{2\pi(n-j)}(n-j)^{n-j}} = \frac{(1-1/n)^{n-j}}{\sqrt{2\pi(n-j)}}\left(n^{1-j/n}\kappa(j/n)\right)^n.$$

We obtain the lower bound similarly,

$$\frac{N^{n-j}}{(n-j)!} \geq e^{-1/12} \frac{(1-1/n)^{n-j}}{\sqrt{2\pi(n-j)}} \left(n^{1-j/n}\kappa(j/n)\right)^n.$$

Finally, with inequality (17), and since $1 \leq n-j \leq n$, we obtain the bounds

$$\frac{1}{4\sqrt{2\pi n}} \leq \frac{e^{-1-1/12}}{\sqrt{2\pi n}} \leq \frac{(1-1/n)^{n-j}}{\sqrt{2\pi(n-j)}} \leq \frac{1}{\sqrt{2\pi}}$$

that prove the estimates on $N^{n-j}/(n-j)!$. □

The next Proposition 5 is a direct consequence of Lemmas 3 and 4. Recall that $\kappa : [0,1] \to \mathbb{R}$ defined in (12), $\kappa(x) = (e/2)^{1-x}(1-x)^{-(1-x)}, \quad \kappa(1) = 1$.

Proposition 5. *For any $n \geq 3$ and $1 \leq j \leq n-2$,*

$$\frac{e^{-2}}{4n^3} \left(n^{1-j/n}\kappa(j/n)\right)^n \leq A_{5,6}(n,j) \leq 2\left(n^{1-j/n}\kappa(j/n)\right)^n. \tag{20}$$

Proof. Lemma 3 proves that, for any $n \geq 3$ and $1 \leq j \leq n-2$,

$$\frac{1}{2n^2} \leq e^{-\mu_{5,6}(n,j)} \left(1 + e^{\mu_{5,6}(n,j)} D_{5,6}(n,j)\right) \leq 2.$$

Then,

$$\begin{aligned} \frac{1}{2n^2}\binom{N}{n-j} \leq &\quad A_5(n,j) &\leq 2\binom{N}{n-j}, \\ \frac{1}{2n^2}\frac{N^{n-j}}{(n-j)!} \leq &\quad A_6(n,j) &\leq 2\frac{N^{n-j}}{(n-j)!}. \end{aligned}$$

Lemma 4 provides us bounds on the terms involving combinatorials and factorials and gives

$$\begin{aligned} \frac{e^{-2}}{2\sqrt{n(n-1)}} \left(n^{1-j/n}\kappa(j/n)\right)^n &\leq \binom{N}{n-j} \leq \left(n^{1-j/n}\kappa(j/n)\right)^n, \\ \frac{1}{4\sqrt{2\pi}\sqrt{n}} \left(n^{1-j/n}\kappa(j/n)\right)^n &\leq \frac{N^{n-j}}{(n-j)!} \leq \frac{1}{\sqrt{2\pi}} \left(n^{1-j/n}\kappa(j/n)\right)^n. \end{aligned}$$

Finally,

$$\begin{aligned} \frac{e^{-2}}{4n^3} \left(n^{1-j/n}\kappa(j/n)\right)^n &\leq A_5(n,j) \leq 2\left(n^{1-j/n}\kappa(j/n)\right)^n, \\ \frac{1}{8\sqrt{2\pi}n^2\sqrt{n}} \left(n^{1-j/n}\kappa(j/n)\right)^n &\leq A_6(n,j) \leq \frac{2}{\sqrt{2\pi}} \left(n^{1-j/n}\kappa(j/n)\right)^n. \end{aligned}$$

Combining both inequalities, Proposition 5 follows. □

2.5 Bender's estimate

The notation $r_n \sim s_n$ indicates that $\lim_{n\to\infty} r_n/s_n = 1$ when $n \to \infty$. Bender [2] establishes that for any real number r such that $0 < r < 1/2$, then

$$S_n^{(j)} \sim \frac{n! e^{-\alpha j}}{j! \rho^{n+1}(1+e^{\alpha})\sigma\sqrt{2\pi n}}$$

uniformly for $j/n \in [r, 1-r]$, where α is such that

$$\frac{n}{j} = (1+e^{\alpha})\ln(1+e^{-\alpha})$$

and

$$\rho = \ln(1+e^{-\alpha}),$$
$$\sigma^2 = \left(\frac{j}{n}\right)^2 \left(1 - e^{\alpha}\ln(1+e^{-\alpha})\right).$$

We introduce two functions to describe the behavior of $S_n^{(j)}$ in terms of j/n,

$$\psi : (0,1) \mapsto \mathbb{R}, \qquad \psi(x) = \frac{e^{-((1-x)+x\delta(x))}}{x^x \ln(1+e^{-\delta(x)})} \tag{21}$$
$$\mu : (0,1) \mapsto \mathbb{R}, \qquad \mu(x) = \left(x(1 - e^{\delta(x)}\ln(1+e^{-\delta(x)}))\right)^{1/2}$$

where $\delta : (0,1) \mapsto \mathbb{R}$ is defined by

$$\delta^{-1}(y) = \frac{1}{(1+e^{y})\ln(1+e^{-y})}. \tag{22}$$

The next lemma rephrases Bender's estimate using $\psi(j/n)$ and $\mu(j/n)$.

Lemma 6. *For any positive real number r such that $0 < r < 1/2$ and for any real number $C > 1$ there exists an integer $n_0 = n_0(r, C) \geq 2$ such that for every integer $n \geq n_0$ and for every integer j with $1 \leq j \leq n-1$ and $j/n \in [r, 1-r]$.*

$$e^{-1/12}\frac{1}{C\sqrt{2\pi n}\mu(j/n)}\left(n^{1-j/n}\psi(j/n)\right)^n \leq S_n^{(j)} \leq e^{1/12}\frac{C}{\sqrt{2\pi n}\mu(j/n)}\left(n^{1-j/n}\psi(j/n)\right)^n.$$

Proof. Observe that

$$(1+e^{\alpha})\rho\sigma = \left(1 - e^{\alpha}\ln(1+e^{-\alpha})\right)^{1/2}.$$

Thus, Bender's estimate implies that for any r with $0 < r < 1/2$ and for any $C > 1$ there exists $n_0 = n_0(r, C)$ such that for any pair of positive integers n, j, with $n \geq n_0$ and $j/n \in [r, 1-r]$,

$$\frac{1}{C} T_\alpha(n,j) \leq S_n^{(j)} \leq C T_\alpha(n,j) \tag{23}$$

where

$$T_\alpha(n,j) = \frac{n!}{j!}\frac{e^{-\alpha j}}{\rho^n(1 - e^{\alpha}\ln(1+e^{-\alpha}))^{1/2}\sqrt{2\pi n}}.$$

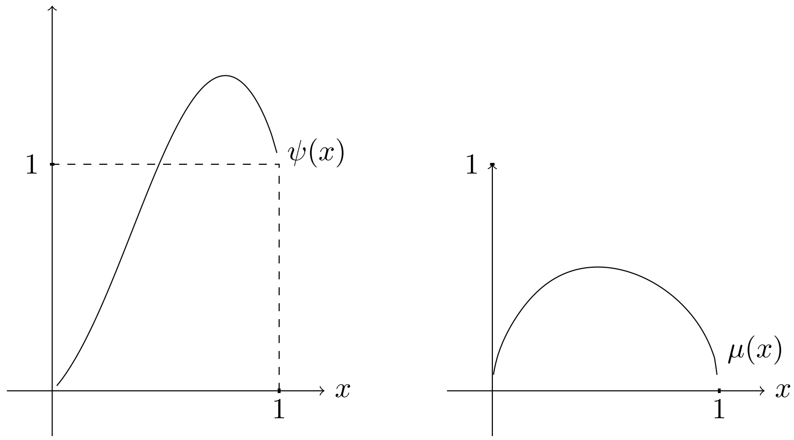

Figure 2: Graphs of $\psi(x)$ and $\mu(x)$.

Using (1) we have

$$e^{-1/12}e^{j-n}\frac{\sqrt{n}}{\sqrt{j}}\frac{n^n}{j^j} \le \frac{n!}{j!} \le e^{1/12}e^{j-n}\frac{\sqrt{n}}{\sqrt{j}}\frac{n^n}{j^j}.$$

We remark that

$$e^{j-n}\frac{n^n}{j^j} = \left(e^{-(1-j/n)}\,(j/n)^{-j/n}\right)^n.$$

Then, using the expressions for $\psi(n/j)$ and $\mu(j/n)$,

$$e^{-1/12}\frac{1}{\sqrt{2\pi n}\mu(j/n)}\left(n^{(n-j)/n}\psi(j/n)\right)^n \le T_\alpha(n,j) \le e^{1/12}\frac{1}{\sqrt{2\pi n}\mu(j/n)}\left(n^{(n-j)/n}\psi(j/n)\right)^n.$$

Combining these inequalities with (23) we obtain the wanted result. □

The functions $\psi(x)$ and $\mu(x)$ are smooth and concave in the open interval $(0,1)$. The function $\delta^{-1}(y)$ is increasing and

$$\lim_{x\to 0^+}\delta(x) = -\infty \text{ and } \lim_{x\to 1^-}\delta(x) = +\infty.$$

From this, it is clear that $\lim_{x\to 0^+}\psi(x) = 0$, $\lim_{x\to 1^-}\psi(x) = 1$, and $\lim_{x\to 0^+}\mu(x) = \lim_{x\to 1^-}\mu(x) = 0$. Then, the bounds given in Lemma 6 become indeterminate when j/n is near 0 or 1. This is why j/n must be in a central interval in $(0,1)$.

The next corollary is a straightforward consequence of Lemma 6 and the fact that $\mu(x)$ is uniformly bounded on any closed interval included in $(0,1)$. The constants c_1 and C_1 in the statement of Corollary 7 can be chosen as the minimum and maximum values of $\{\mu(x) : x \in [r,1-r]\}$.

Corollary 7. *For any positive real number r such that $0 < r < 1/2$, there exist c_1 and C_1 such that for every pair of positive integers n, j with $j/n \in [r, 1-r]$ we have*

$$e^{-1/12}\frac{c_1}{\sqrt{2\pi n}}\left(n^{1-j/n}\psi(j/n)\right)^n \le S_n^{(j)} \le e^{1/12}\frac{C_1}{\sqrt{2\pi n}}\left(n^{1-j/n}\psi(j/n)\right)^n. \quad (24)$$

2.6 A plot

The four upper bounds given in (7), (10), (20) and (24) are of the form

$$S_n^{(j)} \leq n^{n-j}\texttt{bound}$$

In order to visualize them we divide both sides by n^{n-j} and we take n-th root in both sides.

$$\left(S_n^{(j)}/n^{n-j}\right)^{1/n} \leq \texttt{bound}^{1/n}$$

In the four cases $\texttt{bound}^{1/n}$ is of the form

$$\texttt{expression}^{1/n}\,(f^n)^{1/n},$$

where $\texttt{expression}^{1/n}$ goes to 1 as n goes to infinity and f is either θ, η, κ or ψ. Thus, we ignore $\texttt{expression}^{1/n}$. Figure 3 plots the following:

In dotted blue, the exact value
$\widehat{S}_n^{(j)} = (S_n^{(j)}/n^{n-j})^{1/n}$.

In red, the trivial bound
$\widehat{S}_n^{(j)} \leq \dfrac{1}{(\sqrt{2\pi j})^{1/n}}\theta(j/n),$ where $\theta(x)$ is given in (6).

In green, Rennie and Dobson's bound
$\widehat{S}_n^{(j)} \leq \dfrac{1}{2^{1/n}}\eta(j/n),$ where $\eta(x)$ is given in (9).

In blue, Arratia and DeSalvo's bound
$\widehat{S}_n^{(j)} \leq 2^{1/n}\kappa(j/n),$ where $\kappa(x)$ is given in (12).

In black, Bender's estimate
$\widehat{S}_n^{(j)} \leq \left(e^{1/12}\dfrac{C_1}{\sqrt{2\pi n}}\right)^{1/n}\psi(j/n),$ where $\psi(x)$ is given in (21) and C_1 in Corollary 7, with $j/n \in [r, 1-r]$ for any real r such that $0 < r < 1/2$.
The constant C_1 depends on r. In the plot of Figure 3, $r = 0.1$.

3 Application to our problem

For the proof of Theorem 1 we must give an upper bounds of $a(n,j)$, which is always a positive term. Since $a(n,j) = \binom{n}{j} j!\; S_n^{(j)}$, we can use upper bounds for the Stirling numbers of the second kind. We choose Rennie and Dobson's bound in the case j/n is near 0 or 1, and the bound originated in Bender's estimate when j/n is in $[1/r, 1-1/r]$, for $r > 0$.

3.1 When the ratio j/n is near 0 or 1

The next lemma expresses this bound in terms of the ratio j/n with the help of the function

$$\nu : [0,1] \to \mathbb{R}, \quad \nu(x) = x\; e^{-x}\varphi(x)^2, \tag{25}$$

where $\varphi(x)$ is defined in (2).

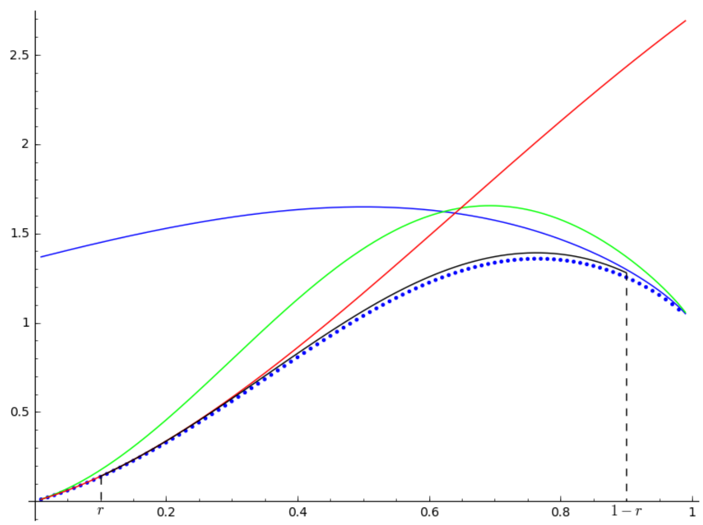

Figure 3: Comparison of four estimates for the normalized Stirling numbers of the second kind, normalized, $\widehat{S}_n^{(j)} = (S_n^{(j)}/n^{n-j})^{1/n}$ for $n = 100$ and $j = 1, .., 100$.

Lemma 8. *For any pair of positive integers n, j such that $n \geq 1$ and $1 \leq j \leq n-1$,*

$$a(n,j)n^{-n} \leq \sqrt{j}\,\nu(j/n)^n.$$

Proof. Recall that $a(n,j) = \binom{n}{j} j!\, S_n^{(j)}$. Rennie and Dobson's upper bound (8) for $S_n^{(j)}$ yields

$$a(n,j) \leq \frac{1}{2}\binom{n}{j}^2 j! j^{n-j}.$$

We apply the estimates (1) for the factorial. Then we use the upper bound for the binomial coefficient given in (4) that involves the constant $C_0 = e^{1/12}(\sqrt{2\pi})^{-1}$, which yields

$$\begin{aligned}\frac{1}{2}\binom{n}{j}^2 j! j^{n-j} &\leq \frac{1}{2}(\sqrt{2}C_0)^2 e^{1/12}\sqrt{2\pi}\sqrt{j}\varphi(j/n)^2 e^{-j} j^n \\ &\leq \frac{e^{1/4}}{\sqrt{2\pi}}\sqrt{j}\,\nu(j/n)^n \\ &\leq \sqrt{j}\,\nu(j/n)^n.\end{aligned}$$

□

The function $\nu(x)$ is smooth and concave, $\nu(0) = 0$, and $\nu(1) = e^{-1}$. The bound given in Lemma 8 is tight when j/n is near 0 or 1. However, it is not good when j/n takes values in middle of the interval $[0, 1]$. In fact, this bound satisfies $\sqrt{j}(\nu(1/2))^n \geq \sqrt{j}(1.1)^n > 1$ but we know that $n^{-n}a(n,j) \leq 1$ for any choice of j and n. This leads us to consider the only two real numbers x_0 and x_1 in $[0, 1]$ for which $\nu(x_0) = \nu(x_1) = 1$ and $x_0 < x_1$. These numbers are $x_0 \approx 0.387$ and $x_1 \approx 0.790$. Figure 4 displays the graphs of $\nu(x)$ and $\varphi(x)$.

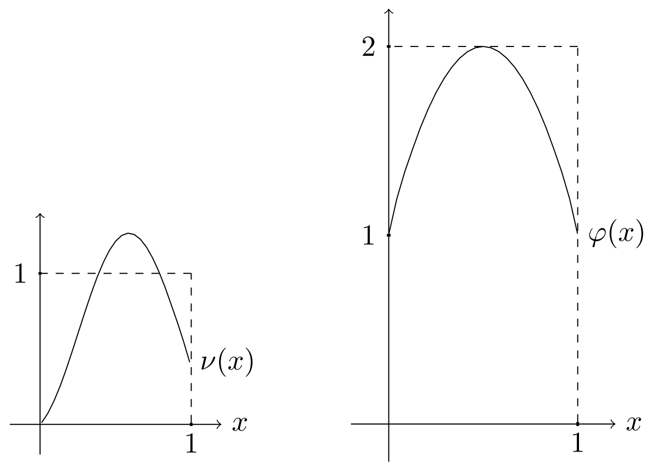

Figure 4: Graphs of functions $\nu(x)$ and $\varphi(x)$.

Lemma 9. *Let x_0 and x_1 be such that $0 < x_0 < x_1 < 1$ and $\nu(x_0) = \nu(x_1) = 1$. For any pair of real numbers r_0 and r_1 such that $0 < r_0 < x_0$ and $x_1 < r_1 < 1$ there exists a real number Λ less than 1, such that for every positive integer n,*

$$n^{-n}a(n,j) \leq \sqrt{n}\Lambda^n, \text{ if } j/n \in [0,r_0] \cup [r_1,1].$$

Proof. Lemma 8 says that $a(n,j)n^{-n} \leq \sqrt{j}\nu(j/n)^n$. The function $\nu(x)$ is smooth and concave with $\nu(0) = 0$, and $\nu(1) = e^{-1}$. This implies the existence of unique points x_0 and x_1 such that $0 < x_0 < x_1 < 1$ and $\nu(x_0) = \nu(x_1) = 1$. Fix r_0 and r_1 such that $0 < r_0 < x_0$ and $x_1 < r_1 < 1$. Necessarily, $\nu(r_0) < 1$ and $\nu(r_1) < 1$. Let $\Lambda_0 = \nu(r_0)$ and $\Lambda_1 = \nu(r_1)$. If $j/n \in [0, r_0]$ then

$$\nu(j/n) \leq \max\{\nu(x) : x \in [0,r_0]\} \leq \Lambda_0.$$

Similarly, if $j/n \in [r_1, 1]$, we have $\nu(j/n) \leq \Lambda_1$. Taking $\Lambda = \max\{\Lambda_0, \Lambda_1\}$, the lemma is proved. □

Example: The choice $r_0 = 0.1$ yields $\Lambda_0 \approx 0.173$, and $r_1 = 0.9$ yields $\Lambda_1 \approx 0.701$. In Figure 1, the value of Λ equals the maximum between the approximations of Λ_0 and Λ_1.

3.2 When the ratio j/n is not near 0 nor 1

We introduce the function

$$\phi : (0,1) \mapsto \mathbb{R}, \qquad \phi(x) = (e\ln(1 + e^{-\delta(x)}))^{-1}\varphi(x)e^{-x\delta(x)} \tag{26}$$

where $\varphi(x)$ is defined in (2) and $\delta(x)$ is defined in (22).

Lemma 10. *Consider the constants c_1 and C_1 in Corollary 7. For any real number r such that $0 < r < 1/2$, and for any pair of positive integers n, j such that $j/n \in [r, 1-r]$*

$$\frac{e^{-1/6}c_1}{\sqrt{2\pi(n-j)}}\phi(j/n)^n \leq n^{-n}a(n,j) \leq \frac{e^{1/6}C_1}{\sqrt{2\pi(n-j)}}\phi(j/n)^n.$$

Proof. Write $a(n,j) = S_n^{(j)} n!/(n-j)!$, then use Stirling estimates (1) for the factorial, apply Corollary 7 and use the definition of $\varphi(x)$ given in (2). $\square$

The function $\phi(x)$ is displayed in Figure 1. It is smooth, concave, $\phi(0) = 0$ and $\phi(1) = e^{-1}$. The auxiliary function $\delta(x)$ takes the value $-\ln(e-1)$ at $x = 1-1/e$ and then, $\phi(1-1/e) = 1$. This value is the maximum of $\phi(x)$ because the lower bound given in Lemma 10 implies that $\phi(x) \leq 1$ for $x \in (0,1)$.

3.3 Proofs of Theorem 1 and Corollary 2

Theorem 1 considers the ratio between j and n. The proof combines the two cases we just studied: when j/n is near 0 or 1, and when j/n is in a central interval away from 0 and 1.

Proof of Theorem 1. Consider the function ν given in (25). Pick numbers x_0 and x_1 such that $0 < x_0 < x_1 < 1$ and $\nu(x_0) = \nu(x_1) = 1$. Take any $r \in (0, 1/2)$ so that $r \leq \max\{x_0, 1-x_1\}$. If $j/n \in [r, 1-r]$ apply Lemma 10. Otherwise, apply Lemma 9. $\square$

The proof of Corollary 2 is immediate from the statement of Theorem 1.

Proof of Corollary 2. The result is a direct application of Theorem 1 because

$$\max\{n^{-n}a(n,j), 1 \leq j \leq \ell\} \leq n^{-n} \sum_{j=1}^{\ell} a(n,j) \leq n \max\{n^{-n}a(n,j), 1 \leq j \leq \ell\}.$$

$\square$

Acknowledgements. We thank an anonymous referee for noticing an error in the lower bound given in Theorem 1 in a previous version of this paper and for many valuable suggestions. The authors are members of LIA SINFIN (ex-INFINIS), Université de Paris-CNRS/Universidad de Buenos Aires-CONICET and they are part of the STIC AMSUD project 20STIC-06. Becher is supported by PICT2018-02315 and Cesaratto is partially supported by Grant UNGS 30/3307.

References

[1] R. Arratia and S. DeSalvo. Completely effective error bounds for Stirling numbers of the first and second kinds via Poisson approximation. *Annals of Combinatorics*, 21:1–24, 2017.

[2] E. A. Bender. Central and local limit theorems applied to asymptotics enumeration. *Journal of Combinatorial Theory, Series A*, 15:91–111, 1973.

[3] E. Cesaratto. *Dimensión de Hausdorff y esquemas de representación de números*. PhD thesis, Universidad de Buenos Aires, 2006.

[4] E. Cesaratto, G. Matera, M. Pérez, and M. Privitelli. On the value set of small families of polynomials over a finite field, I. *Journal of Combinatorial Theory Series A*, 124(4):203–227, 2014.

[5] L. Comtet. *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. Springer Netherlands, 1974.

[6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[7] V. Lifschitz and B. Pittel. The number of increasing subsequences of the random permutation. *Journal of Combinatorial Theory, Series A*, 31(1):1 – 20, 1981.

[8] G. Louchard. Asymptotics of the Stirling numbers of the second kind revisited: A saddle point approach. *Applicable Analysis and Discrete Mathematics*, 7(2):193–210, 2013.

[9] L. Moser and M. Wyman. Stirling numbers of the second kind. *Duke Math. Journal*, 25(1):29–43, 1958.

[10] R. B. Paris. The discrete analogue of Laplace's method. *Computers and Mathematics with Applications*, 61(10):3024 – 3034, 2011.

[11] B. C. Rennie and A. J. Dobson. On Stirling numbers of the second kind. *Journal of Combinatorial Theory*, 7(2):116–121, 1969.

[12] H. Robbins. A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955.

[13] B. Weiss. Poisson generic points, 23-27 November 2020. Conference Diophantine Problems, Determinism and Randomness - CIRM- organized by Joël Rivat and Robert Tichy.
`https://library.cirm-math.fr/Record.htm?idlist=13&record=19287810124910050929`.

Verónica Becher
Departamento de Computación, Facultad de Ciencias Exactas y Naturales & ICC
Universidad de Buenos Aires & CONICET - `vbecher@dc.uba.ar`

Eda Cesaratto
Instituto del Desarrollo Humano,
Universidad Nacional de General Sarmiento & CONICET - `ecesaratto@campus.ungs.edu.ar`