GEOMETRICALLY DISTRIBUTED STIRLING WORDS AND STIRLING COMPOSITIONS

ARNOLD KNOPFMACHER[1], SHI-MEI MA, TOUFIK MANSOUR, AND STEPHAN WAGNER[2]

Abstract. A nonempty word w of finite length over the alphabet of positive integers is a *Stirling word* if for each letter i in w all entries between two consecutive occurrences of i (if these exist) are larger or equal to i. We derive an exact and also an asymptotic formula for the probability that a random geometrically distributed word of length n is a Stirling word. We also determine an asymptotic estimate for the number of compositions (called Stirling compositions) that satisfy this property. Moreover, we find generating functions and asymptotics formulas for statistics in Stirling compositions and geometrically distributed Stirling words, such as the number of distinct values and the size of the maximum part. The proofs make use of various techniques of advanced asymptotic analysis, including Mellin transforms and the saddle point method.

Keywords: Geometrically distributed Stirling words; Stirling compositions; Asymptotic formulas; Mellin transform

1. Introduction

In the last decades, several researchers have become interested in Stirling permutations and Stirling words over a finite alphabet (see [1, 4, 6]). In this paper, we derive the probability that a geometrically distributed word of length n is a Stirling word (for the case of set partitions, see [9]).

If $0 \le p \le 1$, then a discrete random variable X is said to be *geometric* if $P(X = i) = pq^{i-1}$ for all integers $i \ge 1$, where $q = 1 - p$. We will say that a word $w = w_1w_2\cdots$ over the alphabet of positive integers is *geometrically distributed* if the positions of w are independent and identically distributed geometric random variables. The research in geometrically distributed words has been a recent topic of study in enumerative combinatorics; see, e.g., [2, 3] and the references therein. A nonempty word w of finite length over the alphabet of positive integers is a *Stirling word* if for each letter i in w all entries between two consecutive occurrences of i (if these exist) are larger or equal to i.

In this paper, we study the generating functions for the probabilities of geometrically distributed Stirling words according to different statistics. In particular, we show that the probability P_n that a geometrically distributed word of length n is a Stirling word is given by

$$P_n = \frac{p^n}{1-q^n}\prod_{j=1}^{n-1}\left(1+\frac{(n+1-j)q^j}{1-q^j}\right), \tag{1}$$

see Theorem 2.5.

We also study the closely related concept of *Stirling compositions* (see [5]). For a positive integer n, a *composition of* n is a word over the alphabet $\mathbb{N}$ of positive integers whose summands (letters) add up to n. Stirling compositions are those compositions that form a Stirling word. Note that Stirling compositions are precisely the 212-avoiding compositions (a 212-avoiding composition $\sigma_1\sigma_2\cdots\sigma_m$ is a composition such that there are no indices $1 \le i < j < i' \le m$ with $\sigma_i = \sigma_{i'} > \sigma_j$).

Moreover, we find generating functions for statistics in Stirling compositions and geometrically distributed Stirling words of length n, such as the number of distinct values and the size of the maximum part. In Section 3, by various techniques of advanced asymptotic analysis, including

[1]This material is based upon work supported by the National Research Foundation under grant number 81021.
[2]This material is based upon work supported by the National Research Foundation under grant number 96236.

Mellin transforms and the saddle point method, we find an asymptotic estimate for P_n as $n \to \infty$, namely

$$P_n = n^{-\frac{\log n}{2\log q}-\frac{1}{2}}(1-q)^n \frac{\Phi(q)e^{-\frac{\pi^2}{6\log q}}}{q^{1/12}} \exp\left(\Psi(\log_{1/q} n)\right)\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right), \tag{2}$$

where Φ is the generating function for integer partitions and Ψ is an oscillating function (a precise formula is given later). Asymptotic estimates for the mean values of the aforementioned statistics in geometrically distributed Stirling words and Stirling compositions are also determined in Sections 3 and 4.

2. Main results

At first, we study the generating function for the number of Stirling compositions of a positive integer n.

Let $\mathrm{ND}(\pi)$ be the number of different summands (letters) in the word π. For instance, if $\pi = 112441$ then $\mathrm{ND}(\pi) = 3$.

Let $SC_{n,m}$ be the set of Stirling compositions of n with exactly m summands, and define $sc_m(t;u) = \sum_{n\geq m}\sum_{\pi\in SC_{n,m}} t^n u^{\mathrm{ND}(\pi)}$ to be the generating function for the number of Stirling compositions with exactly m summands according to the number of different summands. In order to study $sc_m(t;u)$, we define $sc_{m,d}(t;u)$ to be the generating function for the number of Stirling compositions with exactly m summands and largest summand d, so that $sc_m(t;u) = \sum_d sc_{m,d}(t,u)$.

Let π be a 212-avoiding composition of n with a maximal summand of d. Then π can be written as $\pi = \pi' dd\cdots d\pi''$, such that $\pi'\pi''$ is a Stirling composition of $n - dk$, where k is the number of occurrences of the letter d in π. Hence,

$$sc_{m,d}(t;u) = u\sum_{j=1}^{m}(m+1-j)t^{jd}(sc_{m-j,d-1}(t;u)+\cdots+sc_{m-j,0}(t;u))$$

with $sc_{1,d}(t;u) = ut^d$, $sc_{0,0}(t;u) = 1$ and $sc_{m,0}(t;u) = 0$ for $m \neq 0$. Thus,

$$sc_{m,d}(t;u) - t^d sc_{m-1,d}(t;u) = mut^d(sc_{m-1,d-1}(t;u)+\cdots+sc_{m-1,0}(t;u))$$

or

$$sc_{m,d}(t;u) = t^d sc_{m-1,d}(t;u) + mut^d(sc_{m-1,d-1}(t;u)+\cdots+sc_{m-1,0}(t;u)), \tag{3}$$

with $sc_{1,d}(t;u) = ut^d$.

Define $sc_m(t;u;z) = \sum_{d\geq 1} sc_{m,d}(t;u)z^d$ to be the generating function for the sequence $sc_{m,d}(t;u)$ with fixed $m \geq 1$. So, $sc_1(t;u;z) = \frac{utz}{1-tz}$ and

$$sc_m(t;u;z) = \left(1+\frac{mutz}{1-tz}\right)sc_{m-1}(t;u;tz)$$

by (3). Therefore, by induction on m, we obtain

$$sc_m(t;u;z) = \frac{ut^m z}{1-t^m z}\prod_{j=1}^{m-1}\left(1+\frac{(m+1-j)ut^j z}{1-t^j z}\right). \tag{4}$$

Hence, we can state the following result.

Theorem 2.1. *The generating function for the number of Stirling compositions, where t marks the sum of the terms, u the number of distinct summands, y the total number of summands, and z the largest part, is given by*

$$\sum_{m\geq 1} sc_m(t;u;z)y^m = \sum_{m\geq 1}\frac{ut^m y^m z}{1-t^m z}\prod_{j=1}^{m-1}\left(1+\frac{(m+1-j)ut^j z}{1-t^j z}\right).$$

In particular, the following three corollaries are immediate.

Corollary 2.2. *The generating function for Stirling compositions with respect to the number of distinct summands is*

$$\sum_{m\geq 1}\frac{ut^m}{1-t^m}\prod_{j=1}^{m-1}\left(1+\frac{(m+1-j)ut^j}{1-t^j}\right).$$

The first few terms of this generating function are

$$ut+2ut^2+(2u+2u^2)t^3+(3u+5u^2)t^4+(2u+13u^2)t^5+(3u+19u^2+6u^3)t^6+\cdots.$$

Corollary 2.3. *The generating function for Stirling compositions with respect to the total number of summands is*

$$\sum_{m\geq 1}\frac{t^my^m}{1-t^m}\prod_{j=1}^{m-1}\left(1+\frac{(m+1-j)t^j}{1-t^j}\right).$$

The first few terms of this generating function are

$$yt+(y+y^2)t^2+(y+2y^2+y^3)t^3+(y+3y^2+3y^3+y^4)t^4+(y+4y^2+5y^3+4y^4+y^5)t^5+\cdots.$$

Corollary 2.4. *The generating function for Stirling compositions with respect to the largest part is*

$$\sum_{m\geq 1}\frac{t^mz}{1-t^mz}\prod_{j=1}^{m-1}\left(1+\frac{(m+1-j)t^jz}{1-t^jz}\right).$$

The first few terms of this generating function are

$$zt+(z+z^2)t^2+(z+2z^2+z^3)t^3+(z+4z^2+2z^3+z^4)t^4+(z+6z^2+5z^3+2z^4+z^5)t^5+\cdots.$$

Generating functions for geometrically distributed Stirling words can now be derived as well: if the letters are independent geometrically distributed random variables as explained in the introduction, then the probability of a word $w_1w_2\cdots w_m$ is

$$\prod_{i=1}^{m}pq^{w_i-1}=p^mq^{-m}q^{\sum_{i=1}^m w_i}.$$

Therefore, the generating function for the probabilities of geometrically distributed words to be Stirling words can be obtained by means of the substitutions $y=px/q$ and $t=q$, which gives the following result.

Theorem 2.5. *Let $q\in(0,1)$ be fixed and $p=1-q$. The generating function for the probability of geometrically distributed words with parameter q to be Stirling words, where x marks the length, u marks the number of distinct terms, and z the largest term, is*

$$\sum_{m\geq 1}\frac{up^mx^mz}{1-q^mz}\prod_{j=1}^{m-1}\left(1+\frac{(m+1-j)uq^jz}{1-q^jz}\right).$$

In particular, we obtain the following corollary.

Corollary 2.6. *Let $q\in(0,1)$ be fixed and $p=1-q$. The generating function for the probabilities of geometrically distributed words with parameter q and length n to be Stirling words, where u marks the number of distinct terms, and z the largest term, is*

$$\frac{up^nz}{1-q^nz}\prod_{j=1}^{n-1}\left(1+\frac{(n+1-j)uq^jz}{1-q^jz}\right).$$

In particular, setting $u=1$ and $z=1$ yields (1). Differentiating the generating function in Theorem 2.5 with respect to u and setting $u=z=1$ yields (after division by the probability P_n)

Corollary 2.7. *The average number of different parts in a geometrically distributed Stirling word of length n is given by*

$$1+\sum_{j=1}^{n-1}\frac{(n+1-j)q^j}{1+(n-j)q^j}.$$

Likewise, differentiating with respect to z and setting $u = z = 1$ leads to the following corollary.

Corollary 2.8. *The average size of the largest part in a geometrically distributed Stirling word of length n is given by*

$$\frac{1}{1-q^n} + \sum_{j=1}^{n-1} \frac{(n+1-j)q^j}{(1+(n-j)q^j)(1-q^j)}.$$

Note that following the proof of Theorem 2.1 closely, we see that Stirling compositions with maximal summand d can be constructed exactly as follows. Define $\pi^{(1)}$ to be the word $11\cdots1$ (possibly empty). Given the word $\pi^{(j)}$, we choose a position (between letters, leftmost position or rightmost position) of $\pi^{(j)}$ and insert the word $(j+1)(j+1)\cdots(j+1)$ (possibly empty) to obtain the word $\pi^{(j+1)}$, for all $j = 1, 2, \ldots, d-1$. For instance, to construct the composition 122233325552, we have $\pi^{(1)} = 1$, $\pi^{(2)} = 122222$, $\pi^{(4)} = \pi^{(3)} = 122233322$ and $\pi^{(5)} = 122233325552$.

3. Asymptotics for geometrically distributed Stirling words

We first study the asymptotic behaviour of the probabilities of a geometrically distributed Stirling word of length n, given by

$$P_n = \frac{p^n}{1-q^n} \prod_{j=1}^{n-1} \left(1 + \frac{(n+1-j)q^j}{1-q^j}\right),$$

where $q \in (0,1)$ is a fixed constant and $p = 1-q$. Taking logarithms we obtain

$$\log P_n = n \log p + \sum_{j=1}^{n-1} \log(1+(n-j)q^j) - \sum_{j=1}^{n} \log(1-q^j).$$

Now write

$$1 + (n-j)q^j = (1+nq^j)\left(1 - \frac{jq^j}{1+nq^j}\right).$$

Then

$$\begin{aligned}\sum_{j=1}^{n-1} \log(1+(n-j)q^j) &= \sum_{j=1}^{n-1} \log(1+nq^j) + \sum_{j=1}^{n-1} \log\left(1 - \frac{jq^j}{1+nq^j}\right) \\ &= \sum_{j=1}^{n-1} \log(1+nq^j) + O\left(\sum_{j=1}^{n-1} \frac{jq^j}{1+nq^j}\right).\end{aligned}$$

By the inequality between the arithmetic and geometric mean,

$$\sum_{j=1}^{n-1} \frac{jq^j}{1+nq^j} \le \sum_{j=1}^{n-1} \frac{jq^j}{2\sqrt{nq^j}} \le \frac{1}{\sqrt{n}} \sum_{j=1}^{\infty} \frac{jq^{j/2}}{2} = O\Big(\frac{1}{\sqrt{n}}\Big),$$

so it follows that

$$\sum_{j=1}^{n-1} \log(1+(n-j)q^j) = \sum_{j=1}^{n-1} \log(1+nq^j) + O\Big(\frac{1}{\sqrt{n}}\Big).$$

We next consider the behaviour of the sum

$$\sum_{j=1}^{n-1} \log(1+nq^j).$$

First of all, we extend the sum to infinity at the expense of an exponentially small error term:

$$\sum_{j=1}^{n-1} \log(1+nq^j) = \sum_{j=1}^{\infty} \log(1+nq^j) + O\Big(\sum_{j=n}^{\infty} nq^j\Big) = \sum_{j=1}^{\infty} \log(1+nq^j) + O(nq^n).$$

It remains for us to analyse the remaining infinite sum, and the *Mellin transform* is an appropriate tool to do so. Recall that the Mellin transform of a function f is defined by the integral

$$\mathcal{M}(f(t); s) = f^*(s) = \int_0^\infty t^{s-1} f(t)\, dt,$$

which converges in a certain fundamental strip (possibly empty) of the form $\langle \alpha, \beta\rangle = \{s \in \mathbb{C} : \alpha < \Re s < \beta\}$. We need the following important property of the Mellin transform for our purposes:

$$\mathcal{M}\Big(\sum_j \mu_j f(\lambda_j t); s\Big) = \sum_j \mu_j \lambda_j^{-s} \mathcal{M}(f(t); s).$$

In our case, we have $\mu_j = 1$ and $\lambda_j = q^j$, and the Mellin transform of $\log(1+t)$ is well known to be given by $\frac{\pi}{s\sin(\pi s)}$, with fundamental strip $\langle -1, 0\rangle$. Therefore, the Mellin transform of

$$F(t) = \sum_{j=1}^\infty \log(1 + tq^j)$$

is given by

$$F^*(s) = \mathcal{M}(F(t), s) = \sum_{j=1}^\infty q^{-js} \cdot \frac{\pi}{s\sin(\pi s)} = \frac{1}{q^s - 1} \cdot \frac{\pi}{s\sin(\pi s)}.$$

Now we follow the classical paper of Flajolet, Gourdon and Dumas [10]:

Lemma 3.1 ([10, Corollary 1 to Theorem 4]). *Let $F(t)$ be a continuous function on $(0,\infty)$ with Mellin transform $f^*(s)$ having nonempty fundamental strip $\langle\alpha,\beta\rangle$. Assume that $F^*(s)$ admits a meromorphic continuation to $\langle\alpha,\gamma\rangle$ for some $\gamma > \beta$, and that it is analytic for $\Re s = \gamma$. assume further that the following growth condition holds for some $r > 1$ and $\eta \in (\alpha, \beta)$:*

$$F^*(s) = O(|s|^{-r})$$

when $\Re s \in \{\eta, \gamma\}$ and $|s| \to \infty$, and also when $\eta \le \Re s \le \gamma$ and $|s| \to \infty$ along a denumerable set of horizontal segments $|\Im s| = T_j$, $T_j \to \infty$.

If $F^*(s)$ admits the singular expansion

$$F^*(s) \asymp \sum_{(\xi,\ell)\in A} d_{\xi,\ell} \frac{1}{(s-\xi)^\ell}$$

in the strip $\langle\alpha,\gamma\rangle$ (i.e., if the right side is subtracted from $F^*(s)$, the remaining function is analytic in this strip), then $F(t)$ has the following asymptotic expansion as $t\to\infty$:

$$F(t) = \sum_{(\xi,\ell)\in A} \frac{(-1)^\ell}{(\ell-1)!} d_{\xi,\ell} t^{-\xi} (\log t)^{\ell-1} + O(t^{-\gamma}).$$

Note that $|\sin(\pi s)|$ increases exponentially as $\Im s \to \pm\infty$, and that $\big|\frac{1}{q^s-1}\big|$ is bounded by a constant if we take $\Re s \in \{-1, \frac{1}{2}\}$, and also if $-1 \le \Re s \le \frac{1}{2}$ and $\Im s = \pm\frac{(2k+1)\pi}{\log q}$ ($k \in \mathbb{Z}$). Hence the technical conditions of Lemma 3.1 are satisfied. Inside the strip $\langle -1, \frac{1}{2}\rangle$, $F^*(s)$ has a triple pole at 0 and further poles at $\chi_k = 2\pi i k/\log q$ for $k \in \mathbb{Z}\setminus 0$. We have the singular expansion

$$F^*(s) \asymp \frac{1}{(\log q)s^3} - \frac{1}{2s^2} + \frac{\frac{\pi^2}{6\log q} + \frac{\log q}{12}}{s} - \sum_{k\in\mathbb{Z},k\neq 0} \frac{1}{2k\sinh(2k\pi^2/\log q)(s-\chi_k)},$$

so we find that

$$F(t) = -\frac{\log^2 t}{2\log q} - \frac{\log t}{2} - \frac{\log q}{12} - \frac{\pi^2}{6\log q} + \sum_{k\in\mathbb{Z},k\neq 0} \frac{1}{2k\sinh(2k\pi^2/\log q)} t^{-\chi_k} + O\Big(\frac{1}{\sqrt{t}}\Big).$$

The infinite sum can be regarded as a Fourier series:

$$\sum_{k\in\mathbb{Z},k\neq 0} \frac{1}{2k\sinh(2k\pi^2/\log q)} t^{-\chi_k} = \sum_{k\in\mathbb{Z},k\neq 0} \frac{1}{2k\sinh(2k\pi^2/\log q)} \exp\Big(-\frac{2\pi i k\log t}{\log q}\Big) = \sum_{k=1}^{\infty} \frac{\cos(2k\pi\log_{1/q} t)}{k\sinh(2k\pi^2/\log q)}.$$

In conclusion, we have

$$\sum_{j=1}^{\infty} \log(1+tq^j) = -\frac{\log^2 t}{2\log q} - \frac{\log t}{2} - \frac{\log q}{12} - \frac{\pi^2}{6\log q} + \Psi(\log_{1/q} t) + O\Big(\frac{1}{\sqrt{t}}\Big)$$

as $t\to\infty$, where Ψ is a periodic function with period 1 that has mean value 0 and is given by

$$\Psi(u) = \sum_{k=1}^{\infty} \frac{\cos(2k\pi u)}{k\sinh(2k\pi^2/\log q)}. \tag{5}$$

Consequently, we also have

$$\sum_{j=1}^{n-1} \log(1+(n-j)q^j) = -\frac{\log^2 n}{2\log q} - \frac{\log n}{2} - \frac{\log q}{12} - \frac{\pi^2}{6\log q} + \Psi(\log_{1/q} n) + O\Big(\frac{1}{\sqrt{n}}\Big)$$

as $n\to\infty$.

Moreover,

$$-\sum_{j=1}^{n} \log(1-q^j) = -\sum_{j=1}^{\infty} \log(1-q^j) + O\Big(\sum_{j=n+1}^{\infty} q^j\Big) = \log\Phi(q) + O(q^n),$$

where

$$\Phi(q) = \prod_{j=1}^{\infty} \frac{1}{1-q^j} \tag{6}$$

is the well-known generating function for the number of integer partitions.

In conclusion, we have

$$\log P_n = n\log p - \frac{\log^2 n}{2\log q} - \frac{\log n}{2} - \frac{\log q}{12} - \frac{\pi^2}{6\log q} + \Psi(\log_{1/q} n) + \log\Phi(q) + O\left(\frac{1}{\sqrt{n}}\right),$$

and hence the following theorem.

Theorem 3.2. *As $n\to\infty$, the probability P_n for a geometrically distributed word of length n to be a Stirling word is*

$$P_n = n^{-\frac{\log n}{2\log q}-\frac{1}{2}}(1-q)^n \frac{\Phi(q)e^{-\frac{\pi^2}{6\log q}}}{q^{1/12}} \exp\left(\Psi(\log_{1/q} n)\right)\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right),$$

where Φ is given by (6) *and Ψ is given by* (5).

The term $\Psi(\log_{1/q} n)$, plotted below for $q=1/2$, implies that P_n has tiny oscillations, although these do grow in size for q close enough to 0.

3.1. **Approximations to the partition generating function.** We see that the asymptotic estimate for P_n depends on the values of the partition generating function $\Phi(q)$. Highly accurate approximations can be deduced from the well-known functional equation

$$\Phi(e^{-t}) = \sqrt{t/(2\pi)}\exp\big(\pi^2/(6t) - t/24\big)\Phi(e^{-4\pi^2/t}).$$

Setting $t = -\log q$, we obtain the approximation

$$\Phi(q) = \exp\left(-\frac{\pi^2}{6\log q}\right)\sqrt{-\frac{\log q}{2\pi}}q^{1/24}\Phi(e^{4\pi^2/\log q}) \approx \exp\left(-\frac{\pi^2}{6\log q}\right)\sqrt{-\frac{\log q}{2\pi}}q^{1/24}.$$

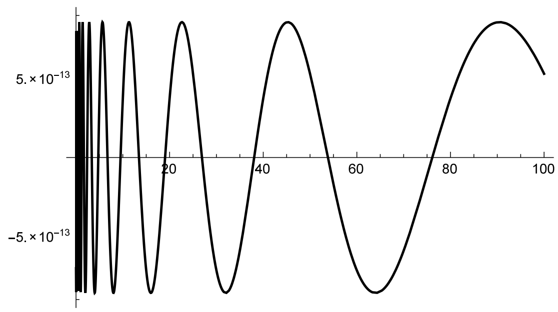

FIGURE 1. $\Psi(\log_{1/q} n)$ *for* $q = 1/2$.

For example, when $q = 1/2$ the approximation is $3.462746619455063611537956 7\cdots$, as compared to the exact value $P_n = 3.4627466194550636115379573\cdots$, a difference of only 6.4×10^{-25}. For the partition function that occurs on the right side of the identity, one can use the estimate

$$1 \le \Phi(x) \le 1 + x/(1 - 2x),$$

which follows from the fact that the number of partitions of n is bounded above for each n, by the number of compositions of n. Hence the relative error of the approximation is bounded by $1/(e^{-4\pi^2/\log q} - 2)$. This gives, for example, 1.8×10^{-25} when $q = 1/2$. When q is very small, the accuracy is lower, but in such cases one can simply approximate Φ by the first few terms of its series expansion.

3.2. **Asymptotics for parameters in geometrically distributed Stirling words.** Now that we have an asymptotic estimate for the probabilities P_n, we continue with the asymptotic study of different statistics. The first one is the number of distinct parts.

Theorem 3.3. *The average number of distinct parts in a geometrically distributed Stirling word of length n is given by*

$$1 + \sum_{i=1}^{n-1} \frac{(n+1-i)q^i}{1+(n-i)q^i} = \log_{1/q} n + \frac{1}{2} + \Psi_1(\log_{1/q} n) + O\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$*, where* Ψ_1 *is a periodic function with period* 1 *given by*

$$\Psi_1(u) = \frac{2\pi}{\log q} \sum_{k=1}^{\infty} \frac{\sin(2k\pi u)}{\sinh(2k\pi^2/\log q)}.$$

Proof. In view of Corollary 2.7, we need to estimate the sum

$$1 + \sum_{i=1}^{n-1} \frac{(n+1-i)q^i}{1+(n-i)q^i}.$$

We have

$$\frac{(n+1-i)q^i}{1+(n-i)q^i} = \frac{nq^i}{1+nq^i} + \frac{q^i}{1+nq^i} + \frac{iq^i(q^i-1)}{(1+(n-i)q^i)(1+nq^i)}, \tag{7}$$

and the sums

$$\sum_{i=1}^{n-1} \frac{q^i}{1+nq^i} \qquad \text{and} \qquad \sum_{i=1}^{n-1} \frac{iq^i(q^i-1)}{(1+(n-i)q^i)(1+nq^i)}$$

are both found to be $O(\frac{1}{\sqrt{n}})$ in the same way as in the proof of Theorem 3.2.

Moreover,

$$\sum_{i=n}^{\infty} \frac{q^i}{1+nq^i} \le \sum_{i=n}^{\infty} q^i = O(q^n),$$

so it remains to analyze the sum

$$\sum_{i=1}^{\infty} \frac{nq^i}{1+nq^i}.$$

To this end, we consider the Mellin transform of the function

$$G(t) = \sum_{i=1}^{\infty} \frac{tq^i}{1+tq^i},$$

in the same way as in the proof of Theorem 3.2. The Mellin transform of $\frac{t}{1+t}$ is $-\frac{\pi}{\sin(\pi s)}$ for $-1 < \Re s < 0$, so the Mellin transform of the sum that defines $G(t)$ is given by

$$-\frac{\pi}{\sin(\pi s)(q^s - 1)},$$

with fundamental strip $-1 < \Re s < 0$. In this case, the singular expansion is

$$G^*(s) \asymp -\frac{1}{(\log q)s^2} + \frac{1}{2s} + \sum_{k \in \mathbb{Z}, k \neq 0} \frac{i\pi}{(\log q)\sinh(2k\pi^2/\log q)(s - \chi_k)},$$

which gives

$$G(t) = -\frac{\log t}{\log q} - \frac{1}{2} - \sum_{k \in \mathbb{Z}, k \neq 0} \frac{i\pi}{(\log q)\sinh(2k\pi^2/\log q)} t^{-\chi_k} + O(t^{-1/2}).$$

The series can be simplified as

$$\frac{2\pi}{\log q} \sum_{k=1}^{\infty} \frac{\sin(2k\pi \log_{1/q} t)}{\sinh(2k\pi^2/\log q)},$$

so putting everything together gives us

$$1 + \sum_{i=1}^{n-1} \frac{(n+1-i)q^i}{1+(n-i)q^i} = 1 + G(n) + O\Big(\frac{1}{\sqrt{n}}\Big) = \log_{1/q} n + \frac{1}{2} + \Psi_1(\log_{1/q} n) + O\Big(\frac{1}{\sqrt{n}}\Big),$$

where

$$\Psi_1(u) = \frac{2\pi}{\log(1/q)} \sum_{k=1}^{\infty} \frac{\sin(2k\pi u)}{\sinh(2k\pi^2/\log(1/q))}.$$

This completes the proof. ∎

Theorem 3.4. *The average size of the largest part in a geometrically distributed Stirling word of length n is given by*

$$\frac{1}{1-q^n} + \sum_{i=1}^{n-1} \frac{(n+1-i)q^i}{(1+(n-i)q^i)(1-q^i)} = \log_{1/q} n + \sum_{j=1}^{\infty} \frac{q^j}{1-q^j} + \frac{1}{2} + \Psi_1(\log_{1/q} n) + O\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$, where Ψ_1 is the same periodic function as in Theorem 3.3.

Proof. We can follow the same steps as in the proof of the previous theorem to show that

$$\frac{1}{1-q^n} + \sum_{i=1}^{n-1} \frac{(n+1-i)q^i}{(1+(n-i)q^i)(1-q^i)} = 1 + \sum_{i=1}^{\infty} \frac{nq^i}{(1+nq^i)(1-q^i)} + O\left(\frac{1}{\sqrt{n}}\right).$$

Now we focus on the sum on the right side of this equation:

$$H(t) = \sum_{i=1}^{\infty} \frac{tq^i}{(1+tq^i)(1-q^i)}.$$

The Mellin transform associated with this function is

$$-\frac{\pi}{\sin(\pi s)} \sum_{i=1}^{\infty} \frac{q^{-is}}{1-q^i}.$$

Making use of the geometric series, we can rewrite this as

$$-\frac{\pi}{\sin(\pi s)}\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}q^{-is}q^{ij} = -\frac{\pi}{\sin(\pi s)}\sum_{j=0}^{\infty}\frac{1}{q^{s-j}-1}.$$

The sum converges locally uniformly and exhibits poles at $j+2k\pi i/\log q$ ($j,k\in\mathbb{Z}$, $j\geq 0$). Now we can apply Lemma 3.1 again as in the previous proof. The singular terms at 0 are

$$-\frac{1}{(\log q)s^2}+\frac{1}{2s}-\frac{1}{s}\sum_{j=1}^{\infty}\frac{q^j}{1-q^j},$$

the last term being the only difference compared to Theorem 3.3. The singular terms at the poles χ_k remain the same, and all poles with positive real part only contribute to the error term. ∎

Remark 3.5. *We observe that the average size of the largest part in a geometrically distributed Stirling word of length n is only a little larger than the average number of distinct parts. In the limit as $n\to\infty$, the difference is the constant*

$$\sum_{j=1}^{\infty}\frac{q^j}{1-q^j}.$$

This means that geometrically distributed Stirling words are typically "almost" gap-free: there are only few values (or none) missing between 1 and the maximum.

4. Asymptotics for Stirling compositions

In this section we determine an asymptotic estimate for the number of Stirling compositions of a positive integer n and also study different statistics again. Recall that the generating function for Stirling compositions of length m is given by

$$t^m\prod_{j=1}^{m}\frac{1+(m-j)t^j}{1-t^j},$$

see Theorem 2.1. As it turns out, the main contribution to the asymptotic behaviour stems from the product $\prod_{j=1}^{m}(1+(m-j)t^j)$. This contribution is quantified in the following lemma.

Lemma 4.1. *For all positive integers m and N, we have*

$$\log\Big([t^N]\prod_{j=1}^{m}(1+(m-j)t^j)\Big)\leq\sqrt{2N}\log m+O(\sqrt{N}).$$

This holds uniformly in m, i.e. the constant in the O-term is independent of m.

Proof. Let $\mathcal{Q}(N,m)$ be the set of all subsets of $\{1,2,\ldots,m\}$ whose sum is N (equivalently, one can think of partitions of N into distinct elements of $\{1,2,\ldots,m\}$). We have

$$[t^N]\prod_{j=1}^{m}(1+(m-j)t^j)=\sum_{J\in\mathcal{Q}(N,m)}\prod_{j\in J}(m-j).$$

Suppose that $J\in\mathcal{Q}(N,m)$ has k elements. Then the sum of all terms in J is at least $1+2+\cdots+k$, so

$$N\geq 1+2+\cdots+k=\frac{k(k+1)}{2}\geq\frac{k^2}{2},$$

which implies that $k\leq\sqrt{2N}$. Hence

$$\begin{aligned}[t^N]\prod_{j=1}^{m}(1+(m-j)t^j) &= \sum_{J\in\mathcal{Q}(N,m)}\prod_{j\in J}(m-j)\\ &\leq \sum_{J\in\mathcal{Q}(N,m)} m^{|J|}\\ &\leq m^{\sqrt{2N}}|\mathcal{Q}(N,m)|.\end{aligned}$$

The cardinality $|\mathcal{Q}(N,m)|$ is certainly less or equal to the number $q(N)$ of partitions of N into distinct summands. It is well known that $\log q(N) = O(\sqrt{N})$ (see e.g. [11]), so the desired inequality follows immediately upon taking the logarithm. ∎

If we suitably restrict m, it is possible to make the formula more precise. This is achieved by means of the saddle point method. Our goal is the following result.

Proposition 4.2. *Let $\delta > 0$ be fixed. For all positive integers m and N that satisfy $N^{1/2+\delta} \leq m \leq N$, we have*

$$\log\Big([t^N]\prod_{j=1}^{m}(1+(m-j)t^j)\Big) = \sqrt{2N}\Big(\log^2 m + \frac{\pi^2}{3}\Big)^{1/2} + O\Big(\frac{N}{m} + \log N\Big).$$

Again, this holds uniformly in m.

In order to prove this statement, we need some additional ingredients, which are provided in the following lemma.

Lemma 4.3. *Let*

$$G_m(u) = \sum_{j=1}^{m} \log(1+(m-j)e^{-\tau j + iju}),$$

where u is a real variable. If $m^{\epsilon-1} \leq \tau \leq m^{-\epsilon}$ for some fixed $\epsilon > 0$, then we have, as $m \to \infty$,

$$\begin{aligned}
G_m(0) &= \frac{1}{\tau}\Big(\frac{\log^2 m}{2} + \frac{\pi^2}{6}\Big) + O\Big(\frac{\log^2 m}{\tau^2 m} + \log m\Big),\\
G'_m(0) &= \frac{i}{\tau^2}\Big(\frac{\log^2 m}{2} + \frac{\pi^2}{6}\Big) + O\Big(\frac{\log^2 m}{\tau^3 m} + \frac{\log m}{\tau}\Big),\\
G''_m(0) &= -\frac{1}{\tau^3}\Big(\log^2 m + \frac{\pi^2}{3}\Big) + O\Big(\frac{\log^2 m}{\tau^4 m} + \frac{\log^2 m}{\tau^2}\Big),\\
G'''_m(u) &= O\Big(\frac{\log^3 m}{\tau^4}\Big) \text{ if } |u| \leq \tau/\log^2 m.
\end{aligned}$$

These estimates hold uniformly in τ and u (provided they satisfy the respective inequalities).

Proof. We first consider

$$G_m(0) = \sum_{j=1}^{m} \log(1+(m-j)e^{-\tau j}).$$

This can be rewritten as

$$G_m(0) = \sum_{j=1}^{m} \log(1+me^{-\tau j}) + \sum_{j=1}^{m} \log\Big(1 - \frac{j}{m+e^{\tau j}}\Big). \tag{8}$$

A simple application of the Euler-Maclaurin sum formula gives us

$$\sum_{j=1}^{m} \log(1+me^{-\tau j}) = \int_0^m \log(1+me^{-\tau v})\,dv + O(\log m).$$

Now we use the substitution $me^{-\tau v} = w$:

$$\begin{aligned}
\int_0^m \log(1+me^{-\tau v})\,dv &= \frac{1}{\tau}\int_{me^{-\tau m}}^{m} \frac{\log(1+w)}{w}\,dw\\
&= \frac{1}{\tau}\int_0^m \frac{\log(1+w)}{w}\,dw + O\Big(\frac{m}{\tau}e^{-\tau m}\Big)\\
&= -\frac{1}{\tau}\operatorname{Li}_2(-m) + O\Big(\frac{m}{\tau}e^{-\tau m}\Big),
\end{aligned}$$

where Li_2 is the dilogarithm:

$$\operatorname{Li}_2(x) = -\int_0^x \frac{\log(1-t)}{t}\,dt = \sum_{n=1}^{\infty} \frac{x^n}{n^2},$$

where the series representation converges for $|x| \le 1$. Since the dilogarithm satisfies the functional equation

$$\mathrm{Li}_2(x) + \mathrm{Li}_2(1/x) = -\frac{\pi^2}{6} - \frac{1}{2}\log^2(-x),$$

we have

$$-\mathrm{Li}_2(-m) = \frac{\log^2 m}{2} + \frac{\pi^2}{6} + O\Big(\frac{1}{m}\Big).$$

The error term $O(\tau^{-1} m e^{-\tau m})$ is smaller than any power of m by our assumptions on τ. Thus

$$\sum_{j=1}^{m} \log(1 + m e^{-\tau j}) = \frac{1}{\tau}\Big(\frac{\log^2 m}{2} + \frac{\pi^2}{6}\Big) + O(\log m).$$

It remains to estimate the second sum in (8):

$$\sum_{j=1}^{m} \log\Big(1 - \frac{j}{m + e^{\tau j}}\Big) = O\Big(\sum_{j=1}^{m} \frac{j}{m + e^{\tau j}}\Big) = O\Big(\int_0^\infty \frac{v}{m + e^{\tau v}}\,dv\Big)$$
$$= O\Big(-\frac{1}{\tau^2 m}\,\mathrm{Li}_2(-m)\Big) = O\Big(\frac{\log^2 m}{\tau^2 m}\Big).$$

In conclusion, we find

$$G_m(0) = \frac{1}{\tau}\Big(\frac{\log^2 m}{2} + \frac{\pi^2}{6}\Big) + O\Big(\frac{\log^2 m}{\tau^2 m} + \log m\Big),$$

as claimed. The proofs of the asymptotic formulas for $G_m'(0)$ and $G_m''(0)$ are analogous, so we only consider the last estimate (for the third derivative) in the remainder of this proof. We have

$$G_m'''(u) = -\sum_{j=1}^{m} \frac{ij^3(m-j)e^{\tau j + iju}(e^{\tau j} - e^{iju}(m-j))}{(e^{\tau j} + e^{iju}(m-j))^3},$$

so

$$|G_m'''(u)| \le \sum_{j=1}^{m} \frac{j^3 m e^{\tau j}(e^{\tau j} + m - j)}{|e^{\tau j} + e^{iju}(m-j)|^3}.$$

Now consider the following two cases:

- If $|ju| \le \frac{\pi}{3}$, then we have
$$|e^{\tau j} + e^{iju}(m-j)| \ge \Re(e^{\tau j} + e^{iju}(m-j)) \ge e^{\tau j} + \frac{1}{2}(m-j) \ge \frac{1}{2}(e^{\tau j} + m - j).$$
- On the other hand, if $|ju| > \frac{\pi}{3}$, then by the assumption that $|u| \le \tau/\log^2 m$, we have $\tau j \ge |ju|\log^2 m \ge \frac{\pi}{3}\log^2 m$, so (for sufficiently large m, to be precise for $m \ge 5$)
$$e^{\tau j} \ge 3m \ge 3(m-j)$$
and thus
$$|e^{\tau j} + e^{iju}(m-j)| \ge e^{\tau j} - (m-j) \ge \frac{1}{2}(e^{\tau j} + m - j).$$

Hence, we get

$$|G_m'''(u)| \le \sum_{j=1}^{m} \frac{8j^3 m e^{\tau j}(e^{\tau j} + m - j)}{(e^{\tau j} + m - j)^3} = \sum_{j=1}^{m} \frac{8j^3 m e^{\tau j}}{(e^{\tau j} + m - j)^2},$$

which no longer depends on u. Now we can proceed as in the estimate for $G_m(0)$. ∎

We also need a tail estimate, which is provided in the following lemma.

Lemma 4.4. *Suppose that $m^{\epsilon - 1} \le \tau \le m^{-\epsilon}$ for some fixed $\epsilon > 0$, and set $z = e^{-\tau + iu}$, $u \in [-\pi, \pi]$. For sufficiently large m, we have*

$$\frac{\Big|\prod_{j=1}^{m}(1 + (m-j)z^j)\Big|}{\prod_{j=1}^{m}(1 + (m-j)e^{-\tau j})} \le \exp\Big(-\frac{u^2}{30\tau^3} + O(1)\Big)$$

if $|u| \leq 3\tau$*, and*

$$\frac{\left|\prod_{j=1}^{m}(1+(m-j)z^j)\right|}{\prod_{j=1}^{m}(1+(m-j)e^{-\tau j})} \leq \exp\Big(-\frac{1}{6\tau}\Big)$$

otherwise.

Proof. An easy calculation shows that

$$\frac{|1+(m-j)z^j|^2}{(1+(m-j)e^{-\tau j})^2} = 1 + \frac{2(m-j)e^{-\tau j}}{(1+(m-j)e^{-\tau j})^2}(\cos(ju)-1) \leq 1.$$

We focus on those values of j for which $|\log m - \tau j| \leq 1$. By our assumptions on τ, this means that $j = O(\tau^{-1}\log m) = O(m^{1-\epsilon}\log m)$. Note also that there are $\frac{2}{\tau} + O(1)$ such vaues. If j satisfies the condition, then

$$\frac{1}{e}\Big(1 + O\Big(\frac{\log m}{m^{\epsilon}}\Big)\Big) \leq (m-j)e^{-\tau j} \leq e,$$

which in turn implies

$$\frac{2(m-j)e^{-\tau j}}{(1+(m-j)e^{-\tau j})^2} \geq \frac{1}{3}$$

for sufficiently large m. Thus

$$\begin{aligned}\frac{\left|\prod_{j=1}^{m}(1+(m-j)z^j)\right|^2}{\prod_{j=1}^{m}(1+(m-j)e^{-\tau j})^2} &\leq \prod_{|\log m - \tau j| \leq 1}\Big(1 + \frac{1}{3}(\cos(ju)-1)\Big)\\ &\leq \exp\Big(\frac{1}{3}\sum_{|\log m - \tau j| \leq 1}(\cos(ju)-1)\Big).\end{aligned}$$

Making use of the geometric series, we find that the sum evaluates to

$$\begin{aligned}\sum_{|\log m - \tau j| \leq 1}(\cos(ju)-1) &= -\frac{2}{\tau} + \frac{\cos(u\log m/\tau)\sin(u/\tau)}{\sin(u/2)} + O(1)\\ &= -\frac{2}{\tau} + \frac{2\cos(u\log m/\tau)\sin(u/\tau)}{u} + O(1).\end{aligned}$$

If $\pi \geq |u| > 3\tau$, then

$$-\frac{2}{\tau} + \frac{2\cos(u\log m/\tau)\sin(u/\tau)}{u} + O(1) \leq -\frac{2}{\tau} + \frac{2}{u} + O(1) < -\frac{1}{\tau}$$

for sufficiently large m. This yields

$$\frac{\left|\prod_{j=1}^{m}(1+(m-j)z^j)\right|}{\prod_{j=1}^{m}(1+(m-j)e^{-\tau j})} \leq \exp\Big(\frac{1}{6}\sum_{|\log m - \tau j| \leq 1}(\cos(ju)-1)\Big) \leq \exp\Big(-\frac{1}{6\tau}\Big).$$

On the other hand, if $|u| \leq 3\tau$, then $|u/\tau| \leq 3$. If then follows that

$$\frac{\sin(u/\tau)}{u/\tau} \leq 1 - \frac{(u/\tau)^2}{10},$$

and consequently

$$-\frac{2}{\tau} + \frac{2\cos(u\log m/\tau)\sin(u/\tau)}{u} + O(1) \leq -\frac{2}{\tau} + \frac{2}{\tau}\Big(1 - \frac{(u/\tau)^2}{10}\Big) + O(1) = -\frac{u^2}{5\tau^3} + O(1),$$

and we conclude in the same fashion as before. ∎

Now we can apply the saddle point method to prove Proposition 4.2.

Proof of Proposition 4.2. By Cauchy's integral formula, we have

$$[t^N]\prod_{j=1}^m (1+(m-j)t^j) = \frac{1}{2\pi i}\oint_{|z|=R} z^{-N-1}\prod_{j=1}^m (1+(m-j)z^j)\,dz$$

for every $R > 0$. We set $R = e^{-\tau}$ and perform the substitution $z = e^{-\tau+iu}$ to obtain

$$[t^N]\prod_{j=1}^m (1+(m-j)t^j) = \frac{e^{\tau N}}{2\pi}\int_{-\pi}^{\pi} e^{-iuN}\prod_{j=1}^m (1+(m-j)e^{-\tau j+iju})\,du. \tag{9}$$

Our first goal is to determine τ in such a way that the derivative of the expression

$$F_{m,N}(u) = -iuN + \sum_{j=1}^m \log(1+(m-j)e^{-\tau j+iju}) = -iuN + G_m(u)$$

vanishes for $u = 0$: we have

$$F'_{m,N}(0) = -iN + \sum_{j=1}^m \frac{ij(m-j)}{m-j+e^{\tau j}} = -iN + G'_m(0),$$

so τ has to satisfy

$$\sum_{j=1}^m \frac{j(m-j)}{m-j+e^{\tau j}} = N. \tag{10}$$

Since the left side of this equation is decreasing in τ, the solution (if it exists) is unique. Moreover, the assumptions on m imply that the left side of the equation is

$$\sum_{j=1}^m \frac{j(m-j)}{m-j+1} = \frac{m^2}{2} + O(m\log m) > N$$

for $\tau = 0$ (and sufficiently large N), and it tends to 0 as $\tau \to \infty$. Therefore, a solution of (10) must exist by the intermediate value theorem.

Now we can use the asymptotic information from Lemma 4.3 to estimate τ: since we need to have

$$G'_m(0) = iN,$$

we get

$$\frac{1}{\tau^2}\Big(\frac{\log^2 m}{2} + \frac{\pi^2}{6}\Big) + O\Big(\frac{\log^2 m}{\tau^3 m} + \frac{\log m}{\tau}\Big) = N,$$

so

$$\tau \sim \frac{\log m}{\sqrt{2N}}$$

and more precisely, with the assumption that $m \geq N^{1/2+\delta}$,

$$\tau = \frac{\sqrt{3\log^2 m + \pi^2}}{\sqrt{6N}} + O\Big(\frac{\log m}{N} + \frac{1}{m}\Big). \tag{11}$$

Note here that τ does indeed satisfy the conditions imposed in Lemma 4.3 and Lemma 4.4 by the assumptions made on m.

As is typical for applications of the saddle point method, we split the integral (9) into a central part $[-T, T]$, where T is chosen to be $N^{-5/7}$ here, and the rest. Let us first estimate the tails, i.e. the integrals where $|u| > T$: by Lemma 4.4, we have

$$\begin{aligned}\Big|\int_T^{\pi} e^{-iuN}\prod_{j=1}^m (1+(m-j)e^{-\tau j+iju})\,du\Big| &\leq K\prod_{j=1}^m (1+(m-j)e^{-\tau j})\exp\Big(-\frac{T^2}{30\tau^3}\Big)\\ &= Ke^{G_m(0)}\exp\Big(-\frac{T^2}{30\tau^3}\Big)\end{aligned}$$

for some positive constant K. Since $\frac{T^2}{30\tau^3} \sim \frac{\sqrt{2}N^{1/14}}{15\log^3 m}$, the final factor goes to 0 faster than any power of N. The same estimate applies to the integral over the interval $[-\pi, -T]$, so we focus on the remaining integral

$$\int_{-T}^{T} e^{-iuN} \prod_{j=1}^{m} (1 + (m-j)e^{-\tau j + iju})\, du = \int_{-T}^{T} \exp\Big(-iuN + \sum_{j=1}^{m} \log(1 + (m-j)e^{-\tau j + iju})\Big)\, du$$
$$= \int_{-T}^{T} \exp\Big(-iuN + G_m(u)\Big)\, du.$$

Taylor expansion gives us, combined with the last item of Lemma 4.3,

$$-iuN + G_m(u) = G_m(0) + \frac{u^2}{2} G_m''(0) + O\Big(\frac{T^3 \log^3 m}{\tau^4}\Big) = G_m(0) + \frac{u^2}{2} G_m''(0) + O(N^{-1/7}).$$

Note that the linear term vanishes by our choice of τ. Finally, since $-G_m''(0)T^2 \sim \frac{2\sqrt{2}N^{1/14}}{\log m} \to \infty$, we have the standard estimate

$$\int_{-T}^{T} \exp\Big(\frac{u^2}{2} G_m''(0)\Big)\, du = \frac{\sqrt{2\pi}}{\sqrt{-G_m''(0)}} + O\Big(e^{G_m''(0)T^2/2}\Big).$$

Putting everything together gives us

$$[t^N] \prod_{j=1}^{m} (1 + (m-j)t^j) = \frac{e^{\tau N + G_m(0)}}{\sqrt{-2\pi G_m''(0)}} \Big(1 + O(N^{-1/7})\Big).$$

Plugging in the asymptotic formula (11) for τ and invoking Lemma 4.3 to estimate $G_m(0)$ and its second derivative, we end up with the following formula:

$$\log\Big([t^N] \prod_{j=1}^{m} (1 + (m-j)t^j)\Big) = \tau N + G_m(0) + O(\log N)$$
$$= 2\tau N + O\Big(\frac{N}{m} + \log N\Big)$$
$$= \sqrt{2N}\Big(\log^2 m + \frac{\pi^2}{3}\Big)^{1/2} + O\Big(\frac{N}{m} + \log N\Big),$$

as required. ∎

Now we are able to prove the main asymptotic formula for the number of Stirling compositions.

Theorem 4.5. *Let $S(n,m)$ be the number of Stirling compositions of n with m parts. The following estimates hold:*

- *For all positive integers n, m with $m \le n$, we have*
$$\log S(n,m) \le \sqrt{2(n-m)} \log m + O(\sqrt{n}).$$
- *Fix a constant $\delta > 0$. For all positive integers n, m with $n^{1/2+\delta} \le m \le n/3$, we have*
$$\log S(n,m) = \sqrt{2(n-m)} \log m + O\Big(\frac{\sqrt{n}}{\log n}\Big).$$

In both formulas, the error terms are uniform in m.

Proof. We have

$$S(n,m) = [t^n] t^m \prod_{j=1}^{m} \frac{1 + (m-j)t^j}{1 - t^j}$$
$$= \sum_{N_1 + N_2 = n-m} \Big([t^{N_1}] \prod_{j=1}^{m} (1 + (m-j)t^j)\Big)\Big([t^{N_2}] \prod_{j=1}^{m} (1 - t^j)^{-1}\Big). \tag{12}$$

We start with the unconditional estimate: by Lemma 4.1, we have

$$\log\Big([t^{N_1}]\prod_{j=1}^{m}(1+(m-j)t^j)\Big) \le \sqrt{2N_1}\log m + O(\sqrt{N_1}) \le \sqrt{2(n-m)}\log m + O(\sqrt{n})$$

for all $N_1 \le n-m$. Moreover, we note that $[t^{N_2}]\prod_{j=1}^{m}(1-t^j)^{-1}$ is the number of partitions of N_2 whose maximum (or equivalently length) is at most m. Clearly, this is at most equal to the total number of partitions of N_2, so the celebrated Hardy-Ramanujan-Rademacher formula for the partition function (see [11]) gives us

$$\log\Big([t^{N_2}]\prod_{j=1}^{m}(1-t^j)^{-1}\Big) = O(\sqrt{N_2}) = O(\sqrt{n}).$$

Combining the two, we immediately obtain that

$$\log S(n,m) \le \sqrt{2(n-m)}\log m + O(\sqrt{n}).$$

Now we make the additional assumption that $n^{1/2+\delta} \le m \le n/3$. Consider first the term corresponding to $N_2 = 0$ in (12): in this case, $N_1 = n-m$, so

$$N_1^{1/2+\delta} \le n^{1/2+\delta} \le m \le n-m = N_1,$$

which means that we can apply Proposition 4.2. This gives us

$$\begin{aligned}\log\Big([t^{N_1}]\prod_{j=1}^{m}(1+(m-j)t^j)\Big) &= \sqrt{2N_1}\Big(\log^2 m + \frac{\pi^2}{3}\Big)^{1/2} + O\Big(\frac{N_1}{m} + \log N_1\Big)\\ &= \sqrt{2(n-m)}\log m + O\Big(\frac{\sqrt{n}}{\log n}\Big),\end{aligned}$$

hence

$$\log S(n,m) \ge \sqrt{2(n-m)}\log m + O\Big(\frac{\sqrt{n}}{\log n}\Big).$$

Now we derive a matching upper bound. If $N_1 < m$, we also have $N_1 < (n-m)/2$, thus

$$\log\Big([t^{N_1}]\prod_{j=1}^{m}(1+(m-j)t^j)\Big) \le \sqrt{2N_1}\log m + O(\sqrt{N_1}) \le \sqrt{n-m}\log m + O(\sqrt{n}),$$

while still

$$\log\Big([t^{N_2}]\prod_{j=1}^{m}(1-t^j)^{-1}\Big) = O(\sqrt{n}).$$

This renders all terms in (12) with $N_1 < m$ asymptotically irrelevant. If, however, $N_1 \ge m$, then the conditions of Proposition 4.2 are met again, so

$$\log\Big([t^{N_1}]\prod_{j=1}^{m}(1+(m-j)t^j)\Big) = \sqrt{2N_1}\log m + O\Big(\frac{\sqrt{n}}{\log n}\Big)$$

as before in the case $N_1 = n-m$. On the other hand, it is well known that almost all partitions of N have length and maximum part (the two are equivalent by conjugation of the Ferrers diagram) $\frac{\sqrt{6N}}{2\pi}\log N(1+o(1))$, see [12]. This implies that almost all partitions of N_2 have maximum at most m, so by the Hardy-Ramanujan-Rademacher formula

$$\log\Big([t^{N_2}]\prod_{j=1}^{m}(1-t^j)^{-1}\Big) = \log\big(p(N_2)(1+o(1))\big) = \pi\sqrt{\frac{2N_2}{3}} + O(\log N_2).$$

Combining the two estimates, we obtain

$$\log\Big([t^{N_1}]\prod_{j=1}^{m}(1+(m-j)t^j)\Big) + \log\Big([t^{N_2}]\prod_{j=1}^{m}(1-t^j)^{-1}\Big) = \sqrt{2N_1}\log m + \pi\sqrt{\frac{2N_2}{3}} + O\Big(\frac{\sqrt{n}}{\log n}\Big). \tag{13}$$

Subject to the condition that $N_1 + N_2 = n - m$, the maximum of the expression without the O-term is

$$\sqrt{2(n-m)}\log m + O\Big(\frac{\sqrt{n}}{\log n}\Big),$$

which means that the logarithm of each term in the sum (12) for $S(n,m)$ is bounded above by $\sqrt{2(n-m)}\log m + O\Big(\frac{\sqrt{n}}{\log n}\Big)$. This is exactly the matching upper bound that we needed to complete the proof of Theorem 4.5. ∎

In order to find an asymptotic estimate for the total number of Stirling compositions of a positive integer n, we determine where the maximum of $S(n,m)$ occurs (as a function of m). This will yield the following theorem.

Theorem 4.6. *Let $S(n)$ denote the total number of Stirling compositions of n. We have*

$$\log S(n) = \sqrt{2n}\Big(\log n - \log\log n + \log 2 - 1\Big) + O\Big(\frac{\sqrt{n}\log\log n}{\log n}\Big).$$

Proof. We first estimate the contribution of "very small" and "very large" m: if $m \leq n^{1/2+\delta}$, then by the first statement of Theorem 4.5, we have

$$\log S(n,m) \leq \sqrt{2n}\log m + O(\sqrt{n}) \leq \big(\tfrac{1}{2}+\delta\big)\sqrt{2n}\log n + O(\sqrt{n}).$$

If on the other hand $m \geq \frac{4n\log\log n}{\log n}$, then the same statement gives us

$$\log S(n,m) \leq \sqrt{2(n-m)}\log n + O(\sqrt{n}) \leq \sqrt{2n}(\log n - 2\log\log n + O(1)).$$

Now, we focus on the case that $n^{1/2+\delta} \leq m \leq \frac{4n\log\log n}{\log n}$. By the second part of Theorem 4.5, we have

$$\log S(n,m) = \sqrt{2(n-m)}\log m + O\Big(\frac{\sqrt{n}}{\log n}\Big).$$

The expression $\sqrt{2(n-m)}\log m$ increases up to a point

$$m_0 = \frac{2n}{\log n}\Big(1 + \frac{\log\log n}{\log n} + O\Big(\frac{1}{\log n}\Big)\Big),$$

then decreases. Indeed, if $m \leq \frac{2n}{\log n}(1-(\log n)^{-1/3})$ or $m \geq \frac{2n}{\log n}(1+(\log n)^{-1/3})$, then we have

$$\log S(n,m) \leq \sqrt{2n}\Big(\log n - \log\log n + \log 2 - 1 - \frac{1}{2(\log n)^{2/3}} + O\Big(\frac{\log\log n}{\log n}\Big)\Big). \tag{14}$$

On the other hand, the maximum attained at m_0 is

$$\log S(n,m_0) = \sqrt{2n}\Big(\log n - \log\log n + \log 2 - 1 + O\Big(\frac{\log\log n}{\log n}\Big)\Big).$$

This readily proves the desired formula, since the summation over m can only contribute at most $O(\log n)$ to $\log S(n)$. ∎

Remark 4.7. *While it might be possible to determine further terms in an asymptotic expansion of $\log S(n)$, it seems that an actual asymptotic formula for $S(n)$ is difficult, if not impossible, to obtain.*

Finally, we consider the asymptotic behaviour of statistics associated with Stirling compositions. The proof of Theorem 4.6 immediately gives us the following result.

Theorem 4.8. *Let X_n be the length of a uniformly random Stirling composition of n. We have*

$$\frac{X_n}{n/\log n} \to 2,$$

both in expectation and in probability.

Proof. From the proof of Theorem 4.6, it can be seen that values of m that are "too far away" from $2n/\log n$ only play a negligible role. Specifically, (14) implies that

$$P\Big(\Big|\frac{X_n \log n}{2n} - 1\Big| \geq \frac{1}{(\log n)^{1/3}}\Big) \leq \exp\Big(- \frac{\sqrt{2n}}{2(\log n)^{2/3}} + O\Big(\frac{\sqrt{n}\log\log n}{\log n}\Big)\Big),$$

and the statement of the theorem readily follows. ∎

The number of distinct summands satisfies a similar theorem, which is shown in the following result.

Theorem 4.9. *Let Y_n be the number of distinct summands of a uniformly random Stirling composition of n. We have*

$$\frac{Y_n}{\sqrt{n}} \to \sqrt{2},$$

both in expectation and in probability.

Proof. First of all, if r is the number of different summands, then we trivially have

$$\frac{r^2}{2} < \frac{r^2+r}{2} = 1 + 2 + \cdots + r \leq n,$$

so $r \leq \sqrt{2n}$ (compare the proof of Lemma 4.1).

On the other hand, we estimate the probability that the number of different summands is "much" less than $\sqrt{2n}$. To this end, we use the standard saddle point bound (see [13, Proposition IV.1]) in the version

$$[u^r t^n]F(t,u) \leq u_0^{-r}[t^n]F(t,u_0),$$

valid for every positive real u_0. We apply it to the bivariate generating function

$$\sum_{m\geq 1} \frac{ut^m}{1-t^m} \prod_{j=1}^{m-1} \Big(1 + \frac{(m+1-j)ut^j}{1-t^j}\Big)$$

(see Theorem 2.1), with $u_0 = 1/m$ for every m:

$$\begin{aligned}[u^r t^n] \sum_{m\geq 1} \frac{ut^m}{1-t^m} \prod_{j=1}^{m-1} \Big(1 + \frac{(m+1-j)ut^j}{1-t^j}\Big) &\leq [t^n] \sum_{m\geq 1} \frac{m^r t^m}{m(1-t^m)} \prod_{j=1}^{m-1} \Big(1 + \frac{(m+1-j)t^j/m}{1-t^j}\Big) \\ &\leq [t^n] \sum_{m\geq 1} m^{r-1} \frac{t^m}{1-t^m} \prod_{j=1}^{m-1} \Big(1 + \frac{t^j}{1-t^j}\Big) \\ &\leq n^{r-1}[t^n] \sum_{m\geq 1} t^m \prod_{j=1}^{m} \frac{1}{1-t^j}.\end{aligned}$$

The latter sum is exactly equal to the generating function for partitions (m representing the maximum or equivalently the length), so this gives us an upper bound of $n^{r-1}p(n)$, where $p(n)$ is the number of partitions of n. As before, we use the estimate $\log p(n) = O(\sqrt{n})$. Combined with Theorem 4.6, this gives

$$P(Y_n = r) \leq \frac{n^{r-1}p(n)}{S(n)} \leq \exp\Big(\log n(r - \sqrt{2n}) + O(\sqrt{n}\log\log n)\Big)$$

for every r, thus

$$P\Big(Y_n \leq \sqrt{2n}\big(1 - (\log n)^{-1/2}\big)\Big) \leq \exp\Big(- \sqrt{2n\log n} + O(\sqrt{n}\log\log n)\Big),$$

from which the desired statement follows. ∎

Finally, we consider the largest summand. Interestingly, its asymptotic behaviour differs from that of the number of distinct summands only by a constant factor of $\frac{3}{2}$.

Theorem 4.10. *Let Z_n be the largest summand of a uniformly random Stirling composition of n. We have*

$$\frac{Z_n}{\sqrt{n}} \to \frac{3}{\sqrt{2}},$$

both in expectation and in probability.

Proof. Note that the difference between the largest part and the number of distinct parts is governed by the factor

$$\prod_{j=1}^{m} \frac{1}{1-t^j z}$$

in the generating function, see Corollaries 2.2 and 2.4. This can be interpreted as the generating function for the number of partitions of a positive integer into parts of size at most m, where the second variable z marks the length.

We need to revisit the proof of Theorem 4.5 in order to determine which values of N_2 contribute most to the total number of Stirling compositions. By Theorem 4.8 and its proof, we may focus on values of m for which

$$m = \frac{2n}{\log n}(1+o(1)) \tag{15}$$

holds. Looking back at (13), we find that the maximum of the expression

$$\sqrt{2N_1}\log m + \pi\sqrt{\frac{2N_2}{3}},$$

subject to the condition that $N_1 + N_2 = n - m$, occurs when

$$N_2 = \frac{(n-m)\pi^2}{3\log^2 m + \pi^2} \sim \frac{\pi^2 n}{3\log^2 n}.$$

Following the same line of proof as in Theorem 4.8, we find that only terms where N_2 is close to this value, in the sense that

$$N_2 = \frac{\pi^2 n}{3\log^2 n}(1+o(1)), \tag{16}$$

make an asymptotically relevant contribution to $S(n,m)$, while the others are negligibly small.

Fixing m and N_2, we can regard the difference between the largest summand and the number of distinct summands as the length of a random partition of N_2 into parts of size at most m. As mentioned before, almost all partitions of N_2 are known to have length and maximum part $\frac{\sqrt{6N_2}}{2\pi}\log N_2(1+o(1))$. In view of (15) and (16), this is asymptotically much less than m, so the restriction on the largest part is in fact irrelevant. Hence we find that the difference between the largest part and the number of distinct parts is concentrated around

$$\frac{\sqrt{6N_2}}{2\pi}\log N_2 \sim \sqrt{\frac{n}{2}}.$$

Combining this with Theorem 4.9, we arrive at the desired statement. ∎

References

[1] , M. Bóna, Real zeros and normal distribution for statistics on Stirling permutations defined by Gessel and Stanley, *SIAM J. Discrete Math.* **23:1** (2008/09) 401–406.

[2] C. Brennan and A. Knopfmacher, The first and last ascents of size d or more in samples of geometric random variables, *Quaest. Math.* **28:4** (2005) 487–500.

[3] C. Brennan and A. Knopfmacher, The distribution of ascents of size d or more in samples of geometric random variables, *2005 International Conference on Analysis of lgorithms*, *Discrete Math. Theor. Comput. Sci. Proc., AD* (2005) 343–351.

[4] I. Gessel and R.P. Stanley, Stirling polynomials, *J. Combinatorial Theory Ser. A* **24:1** (1978) 24–33.

[5] S. Heubach and T. Mansour, Toufik, Combinatorics of compositions and words, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2010.

[6] S.-M. Ma and T. Mansour, Toufik, Pattern restricted Stirling k-ary words, the plateau statistic and the kernel method, *Discrete Appl. Math.* **206** (2016) 100–108.

[7] T. Mansour, M. Shattuck, and S. Wagner, Counting subwords in flattened partitions of sets, *Discrete Math.* **338:11** (2015) 1989–2005.
[8] T. Mansour and M. Shattuck, Some results on the avoidance of vincular patterns by multisets, *Discrete Appl. Math.* **181** (2015) 201–208.
[9] T. Mansour and M. Shattuck, Set partitions as geometric words, *Australas. J. Combin.* **53** (2012) 31–39.
[10] P. Flajolet, X. Gourdon, and P. Dumas, Mellin transforms and asymptotics: harmonic sums, *Theoret. Comput. Sci.* **144** (1995) 3–58.
[11] G.E. Andrews, The theory of partitions, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1998.
[12] P. Erdös and J. Lehner, The distribution of the number of summands in the partitions of a positive integer, *Duke Math. J.* **8** (1941) 335–345.
[13] P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge University Press, Cambridge, 2009.

A. Knopfmacher, The John Knopfmacher Centre for Applicable Analysis and Number Theory, School of Mathematics, University of the Witwatersrand, Private Bag 3, Wits 2050, Johannesburg, South Africa
E-mail address: Arnold.Knopfmacher@wits.ac.za

S.-M. Ma, School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Hebei 066004, P.R. China
E-mail address: shimeimapapers@163.com

T. Mansour, Department of Mathematics, University of Haifa, 3498838 Haifa, Israel
E-mail address: tmansour@univ.haifa.ac.il

S. Wagner, Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa
E-mail address: swagner@sun.ac.za