ON THE DISTRIBUTION OF THE LONGEST RUN IN NUMBER PARTITIONS

STEPHAN WAGNER

ABSTRACT. We consider the distribution of the longest run of equal elements in number partitions (equivalently, the distribution of the largest gap between subsequent elements); in a recent paper, Mutafchiev proved that the distribution of this random variable (appropriately rescaled) converges weakly. The corresponding distribution function is closely related to the generating function for number partitions. In this paper, this problem is considered in more detail—we study the behavior at the tails (especially the case that the longest run is comparatively small) and extend the asymptotics for the distribution function to the entire interval of possible values. Additionally, we prove a local limit theorem within a suitable region, i.e. when the longest run attains its typical order $n^{1/2}$, and we observe another phase transition that occurs when the largest gap is of order $n^{1/4}$: there, the conditional probability that the longest run has length d, given that it is $\leq d$, jumps from 1 to 0. Asymptotics for the mean and variance follow immediately from our considerations.

1. INTRODUCTION

The analytic theory of partitions ranges among the classical topics of number theory and combinatorics. Almost a century ago, Hardy and Ramanujan [10] proved their celebrated asymptotic formula for the partition function $p(n)$, whose main asymptotic term is

$$p(n) \sim \frac{1}{4\sqrt{3}n} \exp\left(\pi\sqrt{\frac{2n}{3}}\right).$$

Their result was subsequently improved by Rademacher [15], who gave an exact convergent series for the partition function. See [2] for an excellent exposition. A similar formula is known for the number of partitions into unequal parts (found by Hua [11]).

The first important example of a distributional result is due to Erdős and Lehner [5], who considered the length (number of parts) of a random partition. Since the length of a partition is the largest part of its conjugate, the maximum follows the same distribution law, which is an extreme value distribution (also known as Gumbel distribution) after appropriate rescaling. We will make use of this result in Section 6 to give a simple, but only heuristic argument for the main theorem of this paper. Szekeres [16, 17] refined the result of Erdős and Lehner and also studied the joint distribution of length and maximum [19] (it turns out that, around their median, the two are essentially independent of each other). He also studied the analogous problem for partitions into unequal parts [18].

Multiplicities of parts in a random partition were first considered by Erdős and Szalay [6]. In particular, they determined the distribution of the largest block size

$$\max_{1 \leq j \leq n} jM_j,$$

where M_j is the multiplicity of j in a random partition of n. This was further generalized by Fristedt [7], who considered, among other shape parameters of a partition, the i-th largest block in a partition. Corteel et al. [4] showed that a randomly selected part of a randomly selected

partition of n has multiplicity m with probability $\frac{1}{m(m+1)}$ as $n \to \infty$. The number of distinct parts in a random distribution (the number of parts with positive multiplicity) was shown to follow a normal law by Goh and Schmutz [8].

For partitions into distinct parts, distribution results are often quite different. See for instance the paper of Hwang [12], in which asymptotic normality of the length (number of parts) is shown in a very general setting.

In the present paper, we will consider the longest run (i.e. the largest occurring multiplicity) of a random partition. By the usual duality, multiplicities in a partition correspond to gaps in the conjugate, so it is equivalent to study the largest gap (or ascent) in a partition. Recently, Brennan, Knopfmacher and the author [3] showed that the number of ascents of at least a given size d follows a Gaussian law. The longest run (or largest ascent), however, has a quite unusual and interesting distribution, as recently shown by Mutafchiev [14]—the corresponding distribution function is closely related to the function that generates partitions. Finally, another recent paper by Grabner and Knopfmacher [9] that discusses various partition statistics based on gaps deserves to be mentioned.

To state the main results we need some more notation. Let $P(n, d)$ denote the number of partitions of n whose longest run has length $< d$, and let $p(n, d) = P(n, d + 1) - P(n, d)$ be the number of partitions of n whose longest run has length exactly d. It is well known [1] that $P(n, d)$ is determined by the generating function

$$G(x) = \sum_{n \geq 0} P(n, d) x^n = \prod_{j \geq 1} \frac{1 - x^{jd}}{1 - x^j} = \frac{F(x)}{F(x^d)},$$

where

$$F(x) = \sum_{n \geq 0} p(n) x^n = \prod_{j \geq 1} (1 - x^j)^{-1}$$

is the generating function for unrestricted partitions. Mutafchiev [14] was able to show that

$$P(n, d) \sim p(n) F(e^{-\alpha})^{-1}$$

for $d \sim \frac{\alpha \sqrt{6n}}{\pi}$ and any fixed positive real number α, implying weak convergence of the normalized random variable L_n, defined as the longest run in a random partition of n. In this paper, we will extend the asymptotic analysis of $P(n, d)$ to the entire range $2 \leq d \leq n$, which can be done by splitting the range into two subcases. This gives us additional information about the behavior of this random variable at its tails and allows us to prove a stronger *local limit theorem*. We also observe another phase transition when d is of order $n^{1/4}$: if $d \ll n^{1/4-\epsilon}$, the conditional probability that the longest run of a partition has length d, given that it is $\leq d$, tends to 1. On the other hand, if $d \gg n^{1/4+\epsilon}$, this probability tends to 0.

The main results can roughly be stated as follows:

Theorem. *If $d \leq C n^{1/2-\delta}$ for fixed positive constants C and δ, then*

$$P(n, d) \sim \frac{\exp\left(\sqrt{\frac{2(d-1)\pi^2 n}{3d}}\right)}{2 \sqrt[4]{6(d-1)^{-1} d^3 n^3}}$$

uniformly in d.

If, on the other hand, $d \geq C n^{1/4+\delta}$ for fixed positive constants C and δ, then

$$P(n, d) \sim p(n) F\left(\exp\left(-\frac{\pi d}{\sqrt{6n}}\right)\right)^{-1}$$

uniformly in d.

Note that the two regions overlap, thus giving a picture of the entire possible range for d.

In the following section, we will describe the general saddle point approach (the classical tool to treat problems of this type), and specialize it to the case of small d (Section 3) and large d (Section 4). The limit distribution as well as asymptotics for mean and variance (Section 5) follow immediately from these considerations. We conclude with some heuristics and potential generalizations.

2. THE GENERAL SADDLE POINT APPROACH

The saddle point method is the most common way to treat problems of such a type, see for instance [19] or [14]. It starts with the observation that we have

$$P(n, d) = [x^n]G(x) = \frac{1}{2\pi i} \int_{\mathcal{C}} x^{-n-1} G(x) \, dx,$$

by the residue theorem, where the integral is taken over any circle $\mathcal{C} = \{re^{it} : -\pi \le t \le \pi\}$ of radius $r < 1$ around 0. Write $r = e^{-\tau}$, and substitute $x = e^{-\tau+it}$ to obtain

$$P(n, d) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{\tau n - int} G\left(e^{-\tau+it}\right) \, dt.$$

Setting $G(e^{-t}) = \exp g(t)$, this can also be written as

$$(1) \qquad P(n, d) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(g(\tau - it) + n(\tau - it)\right) \, dt.$$

In the following, we will always choose τ to be a saddle point, i.e. in such a way that

$$(2) \qquad \frac{d}{dt}\left(g(t) + nt\right) = g'(t) + n = 0$$

at $t = \tau$. We start by collecting a few properties (partly well known, yet important) of τ and the function g; note first that

$$g(t) = \sum_{j \ge 1} \log\left(1 - e^{-djt}\right) - \sum_{j \ge 1} \log\left(1 - e^{-jt}\right) = f(t) - f(dt),$$

where

$$f(t) = -\sum_{j \ge 1} \log(1 - e^{-jt}).$$

Furthermore, writing the function g as

$$g(t) = -\sum_{\substack{j \ge 1 \\ d \nmid j}} \log\left(1 - e^{-jt}\right) = \sum_{\substack{j \ge 1 \\ d \nmid j}} \sum_{k \ge 1} \frac{1}{k} e^{-jkt},$$

it becomes immediately clear that g and all its derivatives are monotone on the entire set of reals, and that the odd-order derivatives are negative, while the even-order derivatives are positive. Hence, the aforementioned saddle point is unique (if it exists, which will be shown later).

Now, let us take a closer look at the function

$$f(t) = -\sum_{j \ge 1} \log(1 - e^{-jt})$$

and its derivatives

$$f'(t) = \sum_{j \ge 1} \frac{j}{1 - e^{jt}},$$

$$f''(t) = \sum_{j \ge 1} \frac{j^2 e^{jt}}{(1 - e^{jt})^2},$$

and generally

$$f^{(k)}(t) = \sum_{j \geq 1} \frac{j^k p_k(e^{jt})}{(1 - e^{jt})^k},$$

where $p_k(x)$ is a polynomial of degree $k - 1$. Note that the Mellin transform of $\frac{d^k}{dt^k}\left(-\log(1 - e^t)\right)$ is given by

$$(-1)^k \Gamma(s) \zeta(s + 1 - k),$$

which is an easy induction. Therefore, the Mellin transform of

$$f^{(k)}(t) = \sum_{j \geq 1} \frac{d^k}{dt^k}\left(-\log(1 - e^{-jt})\right)$$

is given by $(-1)^k \Gamma(s) \zeta(s - k) \zeta(s + 1 - k)$. This function has poles at $s = k$ and $s = k + 1$, and also at $s = k - 1$ if $k \leq 1$. It follows immediately from the Mellin inversion formula that we have, for $k > 1$,

$$f^{(k)}(t) = (-1)^k \left(\frac{\pi^2}{6 t^{k+1}} \Gamma(k + 1) - \frac{1}{2 t^k} \Gamma(k)\right) + O(t^K) = (-1)^k \left(\frac{k! c^2}{t^{k+1}} - \frac{(k - 1)!}{2 t^k}\right) + O(t^K)$$

as $t \to 0$, where we set $c^2 = \zeta(2) = \frac{\pi^2}{6}$, and the implied constant only depends on K (which can be an arbitrary positive number). Additionally,

$$f(t) = \frac{c^2}{t} + \frac{1}{2} \log\left(\frac{t}{2\pi}\right) - \frac{t}{24} + O(t^K)$$

and

$$f'(t) = -\frac{c^2}{t^2} + \frac{1}{2t} - \frac{1}{24} + O(t^K).$$

Note also that $f^{(k)}(t)$ tends to 0 at an exponential rate if $t \to \infty$. Therefore, we have

$$f^{(k)}(t) = O(t^{-k-1})$$

for any k-th derivative of f on the entire set of reals. Hence,

$$g'(t) = f'(t) - df'(dt) \leq -c^2 \left(1 - \frac{1}{d}\right) t^{-2} + O(t^{-1}) \leq -\frac{c^2}{2t^2} + O(t^{-1})$$

and

$$g'(t) = f'(t) - df'(dt) \geq f'(t) = -\frac{c^2}{t^2} + O(t^{-1})$$

uniformly in d, which implies the following lemma:

Lemma 1. *There are absolute constants c_1 and c_2 such that the saddle point τ (i.e., the solution of (2)) satisfies*

$$\frac{c_1}{\sqrt{n}} \leq \tau \leq \frac{c_2}{\sqrt{n}}$$

for all n and d.

In a similar manner,

$$c^2 \left(1 - \frac{1}{d}\right) t^{-k-1} + O(t^{-k}) \leq |g^{(k)}(t)| = |f^{(k)}(t) - d^k f^{(k)}(dt)| \leq c^2 t^{-k-1} + O(t^{-k})$$

uniformly in d, which shows that $g^{(k)}(\tau)$ is always of order $n^{(k+1)/2}$. Now, we consider the integral representation (1), where τ is taken to be the saddle point. We split the integral into two parts: the inner section between $-n^{-5/7}$ and $n^{-5/7}$, and the rest. It turns out, as common in applications of the saddle point method, that the latter is negligible. To this end, we need an estimate for $g(\tau - it)$ if $|t| \geq n^{-5/7}$. Such estimates are well known for f (see e.g. [14] and the references therein), but the additional term $-f(dt)$ can make things a little more intricate; for the sake of completeness, we include the following lemma with a short proof:

Lemma 2. *Assume that there are constants c_1 and c_2 such that $\frac{c_1}{\sqrt{n}} \leq \tau \leq \frac{c_2}{\sqrt{n}}$. Then there is a positive constant c_3 such that*

$$\left| \frac{G(e^{-\tau+it})}{G(e^{-\tau})} \right| \leq \exp\left(-c_3 n^{1/14} \right)$$

whenever $n^{-5/7} \leq |t| \leq \pi$.

Proof. First, we obtain

$$\left| \frac{G(e^{-\tau})}{G(e^{-\tau+it})} \right|^2 = \prod_{\substack{j \geq 1 \\ d \nmid j}} \frac{(1 - e^{-j\tau}\cos(jt))^2 + (e^{-j\tau}\sin(jt))^2}{(1 - e^{-j\tau})^2} = \prod_{\substack{j \geq 1 \\ d \nmid j}} \left(1 + \frac{2(1 - \cos(jt))e^{-j\tau}}{(1 - e^{-j\tau})^2} \right)$$

$$\geq \prod_{\substack{\sqrt{n} \leq j \leq 2\sqrt{n} \\ d \nmid j}} \left(1 + 2(1 - \cos(jt))e^{-j\tau} \right) \geq \prod_{\substack{\sqrt{n} \leq j \leq 2\sqrt{n} \\ d \nmid j}} \left(1 + A_1(1 - \cos(jt)) \right),$$

where $A_1 = 2e^{-2c_2}$ is a positive constant. Now consider two cases: if $|t| \leq \frac{\pi}{2\sqrt{n}}$, we have $n^{-3/14} \leq |tj| \leq \pi$ for all $\sqrt{n} \leq j \leq 2\sqrt{n}$, which allows us to make use of the inequality $1 - \cos u \geq \frac{2u^2}{\pi^2}$ $(0 \leq u \leq \pi)$ to deduce

$$\prod_{\substack{\sqrt{n} \leq j \leq 2\sqrt{n} \\ d \nmid j}} (1 + A_1(1 - \cos(jt))) \geq \prod_{\substack{\sqrt{n} \leq j \leq 2\sqrt{n} \\ d \nmid j}} \left(1 + A_2 t^2 j^2 \right) \geq (1 + A_2 t^2 n)^{\sqrt{n}/2 + O(1)}$$

for some constant $A_2 > 0$. Furthermore, there is a constant $A_3 > 0$ such that $1 + u \geq e^{A_3 u}$ holds for all $0 \leq u \leq A_2 \pi^2 / 4$. Since we have $A_2 t^2 n \leq A_2 \pi^2 / 4$, it follows that

$$(1 + A_2 t^2 n)^{\sqrt{n}/2 + O(1)} \geq \exp\left(A_2 A_3 t^2 n^{3/2}/2 + O(t^2 n) \right) \geq \exp\left(A_4 n^{1/14} \right)$$

for some constant $A_4 > 0$. On the other hand, let now $|t| \geq \frac{\pi}{2\sqrt{n}}$. We consider the set

$$\left| \{ \sqrt{n} \leq j \leq 2\sqrt{n} : |tj - 2k\pi| \leq n^{-1/12} \text{ for some integer } k \} \right|.$$

There are at most $\frac{\sqrt{n}|t|}{2\pi} + 1$ possible values for k, since we have $|t|\sqrt{n} \leq |tj| \leq 2|t|\sqrt{n}$. Furthermore, there are at most $\frac{2n^{-1/12}}{|t|} + 1$ different values of j belonging to any given k. Therefore, the cardinality of this set is at most

$$\left(\frac{2n^{-1/12}}{|t|} + 1 \right)\left(\frac{\sqrt{n}|t|}{2\pi} + 1 \right) = \frac{\sqrt{n}|t|}{2\pi} + O(n^{5/12}) \leq \frac{\sqrt{n}}{2} + O(n^{5/12}).$$

Hence, if $d \geq 3$, we have $\cos(tj) \leq \cos n^{-1/12}$ for at least $\sqrt{n}\left(\frac{1}{2} - \frac{1}{d} + O(n^{5/12}) \right)$ values j for which $\sqrt{n} \leq j \leq 2\sqrt{n}$ and $d \nmid n$. We obtain the estimate

$$\prod_{\substack{\sqrt{n} \leq j \leq 2\sqrt{n} \\ d \nmid j}} (1 + A_1(1 - \cos(jt))) \geq (1 + A_1(1 - \cos n^{-1/12}))^{\sqrt{n}/6 + O(n^{5/12})}$$

$$= (1 + A_1 n^{-1/6}/2 + O(n^{-1/3}))^{\sqrt{n}/6 + O(n^{5/12})}$$

$$= \exp(A_5 n^{1/3} + O(n^{1/4})) \geq \exp(A_6 n^{1/14})$$

for certain positive constants A_5 and A_6, which proves the lemma in the case $d \geq 3$. However, for $d = 2$, we can write

$$G(x) = \prod_{j \geq 1} (1 + x^j)$$

and obtain

$$\left| \frac{G(e^{-\tau})}{G(e^{-\tau+it})} \right| = \prod_{j\geq 1} \left(\frac{(1+e^{-j\tau}\cos(jt))^2 + (e^{-j\tau}\sin(jt))^2}{(1+e^{-j\tau})^2} \right)^{-1} = \prod_{j\geq 1} \left(1 - \frac{2(1-\cos(jt))e^{-j\tau}}{(1+e^{-j\tau})^2} \right)^{-1}$$

$$\geq \prod_{j\geq 1} \left(1 + \frac{2(1-\cos(jt))e^{-j\tau}}{(1+e^{-j\tau})^2} \right) \geq \prod_{j\geq 1} \left(1 + A_7(1-\cos(jt)) \right)$$

for yet another constant $A_7 = e^{-2c_2}/2$ and proceed as for $d \geq 3$, now without the restriction that j has to be odd. ∎

This shows that the part $|t| \geq n^{-5/7}$ in the integral representation (1) is negligible: we have

$$\left| \frac{1}{2\pi} \int_{n^{-5/7}}^{\pi} \exp\left(g(\tau-it) + n(\tau-it)\right) dt \right| \leq \frac{1}{2} \exp\left(g(\tau) + n\tau - c_3 n^{1/14} \right)$$

and an analogous estimate for $t \leq -n^{-5/7}$. Thus we concentrate on the remaining part

$$\frac{1}{2\pi} \int_{-n^{-5/7}}^{n^{-5/7}} \exp\left(g(\tau-it) + n(\tau-it)\right) dt.$$

In order to find the asymptotics of this integral, we expand the exponent into a series around $t = 0$. Note that since τ is known to be of order $n^{-1/2}$ by Lemma 1, $g^{(k)}(\tau)$ is of order $n^{(k+1)/2}$ for all $k \geq 2$. Hence we obtain

$$g(\tau-it) + n(\tau-it) = g(\tau) + n\tau + \sum_{k=2}^{\ell-1} \frac{1}{k!} g^{(k)}(\tau)(-it)^k + O\left(n^{(\ell+1)/2-5\ell/7}\right)$$

for any $\ell \geq 2$ and $|t| \leq n^{-5/7}$, since the first derivative vanishes by the definition of τ. Note further that $g^{(k)}(\tau)t^k \ll n^{(k+1)/2-5k/7} = n^{(7-3k)/14}$ tends to 0 for all $k \geq 3$, and so we have

$$\exp\left(g(\tau-it) + n(\tau-it)\right) = \exp\left(g(\tau) + n\tau - g''(\tau)t^2/2\right)$$
$$\cdot \left(1 + \frac{ig^{(3)}(\tau)t^3}{3!} + \frac{g^{(4)}(\tau)t^4}{4!} - \frac{ig^{(5)}(\tau)t^5}{5!} - \frac{g^{(3)}(\tau)^2 t^6}{2 \cdot 3!^2} - \frac{g^{(6)}(\tau)t^6}{6!} + \cdots \right)$$

for $|t| \leq n^{-5/7}$. Finally, we remark that we can extend the range of the integral to the entire set of reals at the expense of only an exponentially small error term: we have

$$\int_{-n^{-5/7}}^{n^{-5/7}} t^k \exp\left(-g''(\tau)t^2/2\right) dt = \int_{-\infty}^{\infty} t^k \exp\left(-g''(\tau)t^2/2\right) dt - 2\int_{n^{-5/7}}^{\infty} t^k \exp\left(-g''(\tau)t^2/2\right) dt$$

$$= \Gamma\left(\frac{k+1}{2}\right)\left(\frac{g''(\tau)}{2}\right)^{-(k+1)/2} - 2\int_{n^{-5/7}}^{\infty} t^k \exp\left(-g''(\tau)t^2/2\right) dt$$

for even integers k (for odd k, the integral vanishes anyway), and

$$\int_{n^{-5/7}}^{\infty} t^k \exp\left(-g''(\tau)t^2/2\right) dt \leq \int_{n^{-5/7}}^{\infty} t^k \exp\left(-n^{-5/7}g''(\tau)t/2\right) dt$$

$$= \left(\frac{2n^{5/7}}{g''(\tau)}\right)^{k+1} \int_{n^{-10/7}g''(\tau)/2}^{\infty} u^k e^{-u} du$$

$$= \left(\frac{2n^{5/7}}{g''(\tau)}\right)^{k+1} \exp\left(-n^{-10/7}g''(\tau)/2\right) \sum_{j=0}^{k} \frac{k!}{j!}\left(\frac{g''(\tau)}{2n^{10/7}}\right)^j$$

$$\leq \exp\left(-c_4 n^{1/14} + O(\log n)\right)$$

for some absolute positive constant $c_4 = c_4(k)$ (depending only on k), since $g''(\tau) \gg n^{3/2}$ uniformly in d. Putting everything together, we obtain the following theorem:

Theorem 3. *For all $2 \leq d \leq n$, we have the asymptotic expansion*

$$(3) \quad P(n, d) = \frac{1}{\sqrt{2\pi g''(\tau)}} \exp\left(g(\tau) + n\tau\right)$$

$$\cdot \left(1 + \frac{g^{(4)}(\tau)}{8g''(\tau)^2} - \frac{5g^{(3)}(\tau)^2}{24g''(\tau)^3} - \frac{g^{(6)}(\tau)}{48g''(\tau)^3} + \frac{35g^{(4)}(\tau)^2}{384g''(\tau)^4} + \frac{7g^{(3)}(\tau)g^{(5)}(\tau)}{48g''(\tau)^4} + \frac{g^{(8)}(\tau)}{384g''(\tau)^4} + \dots\right),$$

uniformly in d, where τ is the saddle point defined by equation (2).

In the following two sections, we will deduce the precise asymptotics for $P(n, d)$ in two cases: if d is "small" ($d \ll n^{1/2-\delta}$) and if d is "large" ($d \gg n^{1/4+\delta}$). Together, these two cases cover the entire range of possibilities.

3. Analysis if d is small

In this section, we assume that d is small, i.e. $d \ll n^{1/2-\delta}$ for some fixed $\delta > 0$. In this case, it is guaranteed that $d\tau \ll n^{-\delta}$ tends to 0 for our saddle point τ, which allows us to expand $f(d\tau)$ and thus $g(\tau)$ into a series. Then (2) becomes

$$g'(\tau) + n = f'(\tau) - df'(d\tau) + n = -\frac{c^2}{\tau^2}\left(1 - \frac{1}{d}\right) + \frac{d-1}{24} + n + O\left(n^{-\delta K}\right) = 0,$$

where the implied constant only depends on K. Solving for τ, this yields

$$\tau = \sqrt{\frac{24c^2(d-1)}{d(24n+d-1)}} + O(n^{-\delta K - 3/2}).$$

We also know that

$$g(\tau) = f(\tau) - f(d\tau) = \frac{c^2}{\tau}\left(1 - \frac{1}{d}\right) - \frac{1}{2}\log d + \frac{(d-1)\tau}{24} + O\left(n^{-\delta K}\right)$$

and

$$g^{(k)}(\tau) = f^{(k)}(\tau) - d^k f^{(k)}(d\tau) = (-1)^k\left(1 - \frac{1}{d}\right) \cdot \frac{k!c^2}{\tau^{k+1}} + O\left(n^{-\delta K}\right)$$

for all $k \geq 2$. Using all these asymptotic expansions in (3), we obtain

Theorem 4. *If $d \leq Cn^{1/2-\delta}$ for fixed positive constants C and δ, then*

$$P(n, d) = \frac{\exp\left(\sqrt{\frac{2(d-1)\pi^2 n}{3d}}\right)}{2\sqrt[4]{6(d-1)^{-1}d^3n^3}}\left(1 + \frac{(d-1)^2\pi^2 - 27d}{24\pi\sqrt{6d(d-1)n}} + \frac{(d-1)^4\pi^4 - 270d(d-1)^2\pi^2 - 1215d^2}{6912\pi^2d(d-1)n} + \dots\right)$$

uniformly in d.

Remark. If we consider $p(n, d)$, the number of partitions whose longest run has length d (rather than $< d$), we observe a phase transition at $n^{1/4}$: if

- $d \ll n^{1/4-\epsilon}$, then

$$\frac{p(n, d)}{P(n, d+1)} = 1 - \frac{P(n, d)}{P(n, d+1)} \to 1,$$

  i.e. almost every partition whose longest run is of length $\leq d$ has longest run $= d$.
- $d \sim \alpha n^{1/4}$, the ratio of partitions whose longest run has length d among partitions with longest run $\leq d$ converges to a number between 0 and 1:

$$\frac{p(n, d)}{P(n, d+1)} \to 1 - \exp\left(-\frac{\pi}{\sqrt{6}}\alpha^{-2}\right).$$

- $d \gg n^{1/4+\epsilon}$, then

$$\frac{p(n,d)}{P(n,d+1)} = 1 - \frac{P(n,d)}{P(n,d+1)} \to 0,$$

i.e. almost every partition whose longest run is of length $\leq d$ has longest run $< d$, and the number of partitions whose longest run has length d is asymptotically

$$p(n,d) \sim \frac{\pi}{2 \cdot 6^{3/4} n^{1/4} d^{5/2}} \exp\left(\sqrt{\frac{2d\pi^2 n}{3(d+1)}}\right).$$

Finally, let us remark that we obtain the number of partitions into distinct summands as the special case $d = 2$: in this case,

$$p(n,1) = P(n,2) \sim \frac{\exp\left(\pi\sqrt{n/3}\right)}{4 \cdot (3n^3)^{1/4}}.$$

4. Analysis if d is large

Let us now consider the case $d > n^{1/4+\delta}$ (i.e., d is "relatively large") in more detail: in this case, it is not guaranteed any longer that $d\tau$ tends to 0; however, we know now that

$$\frac{d^k}{dt^k} f(dt)\Big|_{t=\tau} = d^k f^{(k)}(d\tau) \ll d^k (d\tau)^{-(k+1)} = d^{-1}\tau^{-(k+1)} \ll n^{(k+1)/2-1/4-\delta} \ll n^{-1/4-\delta} f^{(k)}(\tau)$$

for all k. In particular, $df'(d\tau)$ is asymptotically smaller than $f'(\tau)$, and so (2) yields

$$f'(\tau) = -n + O\left(\frac{n}{d}\right),$$

and further

$$\tau \sim \frac{c}{\sqrt{n}}.$$

In fact, we can be more precise: first of all, we use the expansion of f' around 0, which gives us

$$-\frac{c^2}{\tau^2} + \frac{1}{2\tau} - \frac{1}{24} + O(\tau^K) = -n + O\left(\frac{n}{d}\right),$$

from which it follows immediately that

$$\tau = \frac{c}{\sqrt{n}} + O\left(\frac{1}{n} + \frac{1}{d\sqrt{n}}\right).$$

Now, we expand $f'(dt)$ around $t = \frac{c}{\sqrt{n}}$ to obtain

$$df'(d\tau) = df'\left(\frac{cd}{\sqrt{n}}\right) + d^2 f''\left(\frac{cd}{\sqrt{n}}\right)\left(\tau - \frac{c}{\sqrt{n}}\right) + \dots$$

$$= df'\left(\frac{cd}{\sqrt{n}}\right) + O\left(n^{3/2}d^{-1}\left(\frac{1}{n} + \frac{1}{d\sqrt{n}}\right)\right)$$

$$= df'\left(\frac{cd}{\sqrt{n}}\right) + O\left(\frac{\sqrt{n}}{d} + \frac{n}{d^2}\right).$$

So we get

$$-\frac{c^2}{\tau^2} + \frac{1}{2\tau} - df'\left(\frac{cd}{\sqrt{n}}\right) + n + O\left(1 + \frac{\sqrt{n}}{d} + \frac{n}{d^2}\right) = 0,$$

from which it follows that

$$\tau = \frac{c}{\sqrt{n}} + \frac{cd}{2n^{3/2}} f'\left(\frac{cd}{\sqrt{n}}\right) - \frac{1}{4n} + O\left(\frac{1}{d^2\sqrt{n}} + \frac{1}{dn} + \frac{1}{n^{3/2}}\right).$$

All that remains is to determine an asymptotic expansion for $g^{(k)}(\tau)$, which can be done by the very same means: one has

$$g^{(k)}(\tau) = f^{(k)}(\tau) - d^k f^{(k)}(d\tau),$$

and the two parts can be written as

$$f^{(k)}(\tau) = \begin{cases} \frac{c^2}{\tau} + \frac{1}{2}\log\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{24} + O(\tau^K) & k = 0 \\ (-1)^k \left(\frac{k!c^2}{\tau^{k+1}} - \frac{(k-1)!}{2\tau^k}\right) + O(\tau^K) & k \geq 2 \end{cases}$$

for arbitrary K, where the implied constant only depends on K, and

$$d^k f^{(k)}(d\tau) = \sum_{j=k}^{\ell-1} \frac{d^j}{(j-k)!} f^{(j)}\left(\frac{cd}{\sqrt{n}}\right)\left(\tau - \frac{c}{\sqrt{n}}\right)^{j-k} + O\left(d^\ell \left(\frac{d}{\sqrt{n}}\right)^{-\ell-1}\left(\frac{1}{n} + \frac{1}{d\sqrt{n}}\right)^{\ell-k}\right)$$

$$= \sum_{j=k}^{\ell-1} \frac{d^j}{(j-k)!} f^{(j)}\left(\frac{cd}{\sqrt{n}}\right)\left(\tau - \frac{c}{\sqrt{n}}\right)^{j-k} + O\left(d^{-1} n^{(k+1)/2}\left(\frac{1}{\sqrt{n}} + \frac{1}{d}\right)^{\ell-k}\right)$$

for any $\ell \geq k \geq 0$, where the implied constant only depends on k and ℓ. Using all this in Theorem 3 yields

Theorem 5. *If $d \geq Cn^{1/4+\delta}$ for fixed positive constants C and δ, then*

$$P(n,d) = \frac{\exp\left(2c\sqrt{n} - f\left(\frac{cd}{\sqrt{n}}\right)\right)}{4\sqrt{3}n}$$

$$\cdot \left(1 - \frac{72+\pi^2}{24\pi\sqrt{6n}} + \frac{d}{n}f'\left(\frac{cd}{\sqrt{n}}\right) - \frac{\pi d^2}{4\sqrt{6}n^{3/2}}\left(f'\left(\frac{cd}{\sqrt{n}}\right)^2 - f''\left(\frac{cd}{\sqrt{n}}\right)\right) + O\left(\frac{n}{d^4} + \frac{1}{n}\right)\right)$$

$$\sim p(n)\exp\left(-f\left(\frac{cd}{\sqrt{n}}\right)\right)$$

uniformly in d.

This shows that if L_n is the random variable "longest run in a random integer partition of n", then the normalized variable $cn^{-1/2}L_n$ converges weakly to the distribution whose distribution function is given by

$$(4) \qquad\qquad D(y) = \exp(-f(y)) = \prod_{j\geq 1}\left(1 - e^{-jy}\right).$$

However, as another immediate corollary, we obtain the following *local limit theorem*, which is a stronger statement than the convergence in distribution:

Corollary 6. *If $Cn^{1/4+\delta} \leq d \leq \left(\frac{1}{2c} - \epsilon\right)\sqrt{n}\log n$ for fixed positive constants C, δ and ϵ, then the number of partitions of n whose longest run has length d is asymptotically*

$$p(n,d) \sim -\frac{c}{\sqrt{n}}f'\left(\frac{cd}{\sqrt{n}}\right)\exp\left(-f\left(\frac{cd}{\sqrt{n}}\right)\right)p(n),$$

uniformly in d.

Remark. Note that the upper bound for n is necessary to ensure that

$$-\frac{c}{\sqrt{n}}f'\left(\frac{cd}{\sqrt{n}}\right)$$

is asymptotically larger than the error term, since $-f'(t)$ behaves like e^{-t} as $t \to \infty$. However, if one expands further, it is possible to improve on the constant $\frac{1}{2c}$ (not on the order of magnitude, though).

5. Mean and variance

It is not difficult to determine the asymptotics of mean and variance of the longest run in a random integer partition of n from our theorems (note, however, that weak convergence of the normalized random variable does not imply convergence of moments). The mean of the longest run is given by

$$\mathbb{E}(L_n) = \sum_{d=1}^{n} \frac{dp(n,d)}{p(n)} = \sum_{d=1}^{n} \frac{d(P(n,d+1) - P(n,d))}{p(n)} = n - \sum_{d=1}^{n} \frac{P(n,d)}{p(n)} = \sum_{d=1}^{n} \left(1 - \frac{P(n,d)}{p(n)}\right),$$

since $P(n, n+1) = p(n)$. Now we apply Theorem 5 to obtain

$$\mathbb{E}(L_n) = \sum_{d=\lfloor n^{1/3} \rfloor}^{n} \left(1 - \frac{P(n,d)}{p(n)}\right) + O(n^{1/3})$$

$$= \sum_{d=\lfloor n^{1/3} \rfloor}^{n} \left(1 - \exp\left(-f\left(\frac{cd}{\sqrt{n}}\right)\right)\right) + O(n^{1/3}).$$

Since $f(t)$ decreases exponentially as $t \to \infty$, we can extend the range of summation to obtain

$$\mathbb{E}(L_n) = \sum_{d=1}^{\infty} \left(1 - \exp\left(-f\left(\frac{cd}{\sqrt{n}}\right)\right)\right) + O(n^{1/3}),$$

and the Euler-Maclaurin formula gives us

$$\mathbb{E}(L_n) = \int_0^{\infty} \left(1 - \exp\left(-f\left(\frac{ct}{\sqrt{n}}\right)\right)\right) dt + O(n^{1/3})$$

$$= \frac{\sqrt{n}}{c} \int_0^{\infty} \left(1 - \exp\left(-f(u)\right)\right) du + O(n^{1/3}).$$

Now recall that

$$1 - \exp\left(-f(u)\right) = 1 - \prod_{j \geq 1}(1 - e^{-ju}) = \sum_{k \geq 1}(-1)^{k-1}\left(e^{-\frac{k(3k-1)}{2}u} + e^{-\frac{k(3k+1)}{2}u}\right)$$

by virtue of Euler's pentagonal number theorem. Hence,

$$\mathbb{E}(L_n) \sim \frac{\sqrt{n}}{c} \int_0^{\infty} \sum_{k \geq 1}(-1)^{k-1}\left(e^{-\frac{k(3k-1)}{2}u} + e^{-\frac{k(3k+1)}{2}u}\right) du$$

$$= \frac{\sqrt{n}}{c} \sum_{k \geq 1}(-1)^{k-1}\left(\frac{2}{k(3k-1)} + \frac{2}{k(3k+1)}\right)$$

$$= \frac{\sqrt{n}}{c} \sum_{k \geq 1}(-1)^{k-1}\left(\frac{6}{3k-1} - \frac{6}{3k+1}\right)$$

$$= \frac{6\sqrt{n}}{\sqrt{3}ci} \sum_{k \geq 2} \frac{\zeta^k - \zeta^{-k}}{k},$$

where $\zeta = \exp(\pi i/3)$ is a sixth root of unity. The sum reduces to

$$\mathbb{E}(L_n) \sim \frac{6\sqrt{n}}{\sqrt{3}ci}\left(-i\sqrt{3} - \log(1-\zeta) + \log(1-\zeta^{-1})\right) = \left(4\sqrt{2} - \frac{6\sqrt{6}}{\pi}\right)\sqrt{n} \approx 0.978673\sqrt{n}.$$

The same method can be applied to determine the asymptotics of the variance. In this case, the identity

$$\mathbb{E}(L_n^2) = \sum_{d=1}^{n}(2d-1)\left(1 - \frac{P(n,d)}{p(n)}\right)$$

needs to be used, and analogous reasoning yields

$$\mathbb{V}(L_n) \sim \left(\frac{1080}{\pi^2} - \frac{96\sqrt{3}}{\pi} - 56\right) n \approx 0.499304\, n.$$

6. Concluding remarks

Remark. It was noted in Mutafchiev's paper [14] that Theorem 5 implies weak convergence of the distribution of the normalized random variable "longest run in a random partition of n" to the distribution of the random variable

$$X = \max_{j \geq 1} \frac{X_j}{j}, \tag{5}$$

where X_j are mutually independent random variables, all following an exponential distribution with common mean 1. The distribution function of this random variable is indeed given by (4). This agrees with a simple heuristic argument: if M_j is the multiplicity of j in a random partition of n for some fixed j, then $\frac{c}{\sqrt{n}} M_j$ follows the distribution of $\frac{X_j}{j}$, and assuming asymptotic independence, one can expect the distribution of the normalized longest run to be given by (5).

It is tempting to try a similar heuristic argument for the largest gap. By the well-known result of Erdős and Lehner [5], the size of the largest part in a random partition is asymptotically

$$\frac{\sqrt{n}}{2c} \log n - \frac{\sqrt{n}}{c} V_1,$$

where $X_1 = \frac{\exp V_1}{c}$ follows an exponential distribution with mean 1. The distribution of the second part is the same with a new random variable V_2 (if we remove the largest part, we obtain a partition of some n', where $n' \sim n$ for almost all partitions), with the obvious additional condition that $V_2 \geq V_1$. However, since the conditional distribution of an exponential random variable, given that it is larger than some constant, is still an exponential distribution, we obtain

$$V_1 \sim \log(cX_1), \qquad V_2 \sim \log(cX_1 + cX_2)$$

for independent exponentially distributed random variables X_1, X_2 with mean 1. Likewise, the j-th largest part is asymptotically

$$\frac{\sqrt{n}}{2c} \log n - \frac{\sqrt{n}}{c} V_j$$

with

$$V_j \sim \log(cX_1 + cX_2 + \ldots + cX_j),$$

which implies that the gaps have sizes

$$\frac{\sqrt{n}}{c}\Big(\log(cX_1 + cX_2 + \ldots + cX_{j+1}) - \log(cX_1 + cX_2 + \ldots + cX_j)\Big) = \frac{\sqrt{n}}{c} \log \frac{X_1 + \ldots + X_{j+1}}{X_1 + \ldots + X_j}$$

for $j \geq 1$, and one would expect the random variable Y that is defined by

$$Y = \log\left(\max_{j \geq 1} \frac{X_1 + \ldots + X_{j+1}}{X_1 + \ldots + X_j}\right)$$

to follow the same law as the normalized largest gap. Indeed, we will show now that this is true. The distribution function of Y is given by

$$\mathbb{P}(Y \leq y) = \mathbb{P}\left(\frac{X_1 + \ldots + X_{j+1}}{X_1 + \ldots + X_j} \leq e^y \text{ for all } j\right) = \mathbb{P}\left(X_{j+1} \leq (e^y - 1)(X_1 + \ldots + X_j) \text{ for all } j\right).$$

Now let us compute the probability that $X_{j+1} \leq (e^y - 1)(X_1 + \ldots + X_j)$ holds for all $j < n$. We use a as an abbreviation for $e^y - 1$. Then this probability is given by

$$\int_0^\infty \int_0^{ax_1} \int_0^{a(x_1+x_2)} \ldots \int_0^{a(x_1+\ldots+x_{n-1})} e^{-(x_1+\ldots+x_n)} \, dx_n \ldots dx_1$$

The easiest way to determine this integral is to use induction to show that

$$\int_0^\infty \int_0^{ax_1} \int_0^{a(x_1+x_2)} \cdots \int_0^{a(x_1+\ldots+x_{n-1})} e^{-b(x_1+\ldots+x_n)} \, dx_n \ldots dx_1 = b^{-n} \prod_{j=1}^{n-1} \left(1 - \frac{1}{(a+1)^j} \right).$$

This is obviously true for $n = 1$. Now, simply note that

$$\int_0^\infty \int_0^{ax_1} \int_0^{a(x_1+x_2)} \cdots \int_0^{a(x_1+\ldots+x_n)} e^{-b(x_1+\ldots+x_{n+1})} \, dx_{n+1} \ldots dx_1$$

$$= \frac{1}{b} \int_0^\infty \int_0^{ax_1} \int_0^{a(x_1+x_2)} \cdots \int_0^{a(x_1+\ldots+x_{n-1})} e^{-b(x_1+\ldots+x_n)} \left(1 - e^{-ab(x_1+\ldots+x_n)} \right) \, dx_n \ldots dx_1$$

$$= \frac{1}{b} \left(b^{-n} \prod_{j=1}^{n-1} \left(1 - \frac{1}{(a+1)^j} \right) - ((a+1)b)^{-n} \prod_{j=1}^{n-1} \left(1 - \frac{1}{(a+1)^j} \right) \right)$$

$$= b^{-n-1} \prod_{j=1}^{n} \left(1 - \frac{1}{(a+1)^j} \right).$$

Taking $b = 1$ and letting $n \to \infty$, we obtain the desired result, namely that the distribution function is indeed

$$D(y) = \prod_{j \geq 1} \left(1 - e^{-jy} \right).$$

Remark. Our asymptotic analysis is not restricted to the case of ordinary integer partitions. In fact, it can be carried out within a far more general framework known as Meinardus' scheme [1, 13], see also [12]: given a sequence $a_1, a_2, \ldots$ of nonnegative integers, consider the generating function

$$F_a(x) = \sum_{n \geq 0} p_a(n) x^n = \prod_{j \geq 1} (1 - x^j)^{-a_j},$$

which counts integer partitions with the additional assumption that there are a_j distinguishable "types" of the integer j (one might think of different versions of the same integer, distinguished by colors). Then, if the Dirichlet generating function $A(s) = \sum_{j \geq 1} a_j j^{-s}$ has only a simple pole at $s = \sigma_0 > 0$, can be extended to the half-plane $\operatorname{Re} s > -\sigma_1$ for some $\sigma_1 > 0$, and satisfies some more technical conditions, then a similar analysis can be carried out.

REFERENCES

[1] G. E. Andrews. *The theory of partitions.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1998.

[2] T. M. Apostol. *Modular functions and Dirichlet series in number theory*, volume 41 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1990.

[3] C. Brennan, A. Knopfmacher, and S. Wagner. The distribution of ascents of size d or more in partitions of n. *Combinatorics, Probability and Computing*, 2008. To appear.

[4] S. Corteel, B. Pittel, C. D. Savage, and H. S. Wilf. On the multiplicity of parts in a random partition. *Random Structures Algorithms*, 14(2):185–197, 1999.

[5] P. Erdős and J. Lehner. The distribution of the number of summands in the partitions of a positive integer. *Duke Math. J.*, 8:335–345, 1941.

[6] P. Erdős and M. Szalay. On the statistical theory of partitions. In *Topics in classical number theory, Vol. I, II (Budapest, 1981)*, volume 34 of *Colloq. Math. Soc. János Bolyai*, pages 397–450. North-Holland, Amsterdam, 1984.

[7] B. Fristedt. The structure of random partitions of large integers. *Trans. Amer. Math. Soc.*, 337(2):703–735, 1993.

[8] W. M. Y. Goh and E. Schmutz. The number of distinct part sizes in a random integer partition. *J. Combin. Theory Ser. A*, 69(1):149–158, 1995.

[9] P. J. Grabner and A. Knopfmacher. Analysis of some new partition statistics. *Ramanujan J.*, 12(3):439–454, 2006.

[10] G. H. Hardy and S. Ramanujan. Asymptotic formulæ in combinatory analysis [Proc. London Math. Soc. (2) **16** (1917), Records for 1 March 1917]. In *Collected papers of Srinivasa Ramanujan*, page 244. AMS Chelsea Publ., Providence, RI, 2000.

[11] L. K. Hua. On the number of partitions of a number into unequal parts. *Trans. Amer. Math. Soc.*, 51:194–201, 1942.

[12] H.-K. Hwang. Limit theorems for the number of summands in integer partitions. *J. Combin. Theory Ser. A*, 96(1):89–126, 2001.

[13] G. Meinardus. Asymptotische Aussagen über Partitionen. *Math. Z.*, 59:388–398, 1954.

[14] L. R. Mutafchiev. On the maximal multiplicity of parts in a random integer partition. *Ramanujan J.*, 9(3):305–316, 2005.

[15] H. Rademacher. On the expansion of the partition function in a series. *Ann. of Math. (2)*, 44:416–422, 1943.

[16] G. Szekeres. An asymptotic formula in the theory of partitions. *Quart. J. Math., Oxford Ser. (2)*, 2:85–108, 1951.

[17] G. Szekeres. Some asymptotic formulae in the theory of partitions. II. *Quart. J. Math., Oxford Ser. (2)*, 4:96–111, 1953.

[18] G. Szekeres. Asymptotic distribution of the number and size of parts in unequal partitions. *Bull. Austral. Math. Soc.*, 36(1):89–97, 1987.

[19] G. Szekeres. Asymptotic distribution of partitions by number and size of parts. In *Number theory, Vol. I (Budapest, 1987)*, volume 51 of *Colloq. Math. Soc. János Bolyai*, pages 527–538. North-Holland, Amsterdam, 1990.

Stephan Wagner, Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

E-mail address: swagner@sun.ac.za