Discrete Mathematics and Theoretical Computer Science DMTCS vol. **17:1**, 2015, 123–144

*Bootstrapping and double-exponential limit laws**

Helmut Prodinger† Stephan Wagner‡

Department of Mathematical Sciences, Stellenbosch University, South Africa.

received 15th May 2012, revised 6th Mar. 2015, accepted 9th Mar. 2015.

We provide a rather general asymptotic scheme for combinatorial parameters that asymptotically follow a discrete double-exponential distribution. It is based on analysing generating functions $G_h(z)$ whose dominant singularities converge to a certain value at an exponential rate. This behaviour is typically found by means of a bootstrapping approach. Our scheme is illustrated by a number of classical and new examples, such as the longest run in words or compositions, patterns in Dyck and Motzkin paths, or the maximum degree in planted plane trees.

Keywords: double-exponential distribution, generating functions, bootstrapping, asymptotic scheme

1 Introduction

In many combinatorial problems, one encounters the situation that one obtains a sequence of generating functions $G_h(z)$, depending on a parameter h whose distribution is to be studied, such that the dominant singularity ζ_h of G_h converges to a value ζ as $h \to \infty$, and $\zeta_h - \zeta$ decreases exponentially with h. The archetypical example is probably the distribution of the longest sequence of 1's in a random 0-1-sequence: Let $G_h(z)$ denote the generating function for 0-1-sequences with the property that there is no sequence of more than h consecutive 1's. Such a word can be symbolically described as (using the notation of [5])

$$\textsc{Seq}(\textsc{Seq}_{0..h}(1) \times 0) \times \textsc{Seq}_{0..h}(1), \tag{1}$$

which translates directly to the following expression for the generating function:

$$G_h(z) = \frac{1}{1 - \frac{z(1-z^{h+1})}{1-z}} \cdot \frac{1 - z^{h+1}}{1-z} = \frac{1 - z^{h+1}}{1 - 2z + z^{h+2}}.$$

Note that $\lim_{h\to\infty} G_h(z) = \frac{1}{1-2z}$ is the generating function for all 0-1-sequences. The dominant singularity ζ_h of this rational function is a pole close to $1/2$, the unique positive solution of the equation $2z - z^{h+2} = 1$. By means of bootstrapping, one finds that $\zeta_h = 1/2 + 2^{-h-3} + O(h2^{-2h})$. Such

*This material is based upon work supported financially by the National Research Foundation of South Africa under grant numbers 2053748 and 70560.

†Email: hproding@sun.ac.za

‡Email: swagner@sun.ac.za

behaviour in the dominant singularity typically leads to a discrete limit law analogous to the Gumbel distribution, and to periodic fluctuations. This was probably first observed by Knuth [17] in his work on carry propagation. In the present example, we find:

- If L_n is the length of the longest sequence of consecutive 1's in a random 0-1-sequence of length n, then the mean of L_n is

$$\mathbb{E}(L_n) = \log_2 n + \frac{\gamma}{\log 2} - \frac{3}{2} + \psi(\log_2 n) + o(1)$$

  for a 1-periodic (i.e., periodic with period 1) function ψ that will be specified later.

- The random variable L_n follows a distribution given by

$$\mathbb{P}(L_n \le \log_2 n + x) \sim \exp(-2^{-x-2}),$$

  provided that $\log_2 n + x$ is an integer. This double-exponential distribution can be seen as a discrete analogue of the Gumbel distribution.

General schemes treating the case of rational generating functions [14] and functions with a square root singularity [25] have been developed in the past. Gourdon [8] provides a general framework for the distribution of the largest component in combinatorial structures obtained by a substitution process, which leads to a double-exponential distribution in the so-called *super-critical* case, and more recently Bender and Gao [2] obtained general results for the maximum part in supercritical compositional structures. However, there remain many examples that are not covered by these papers. Our aim is to provide a reasonably general framework of families of generating functions whose structure leads to the same behaviour as in our first example (double-exponential limit law, fluctuations in the moments). The main result reads as follows:

Theorem 1 *Let $G_h(z) = \sum_{n\ge 0} a_{h,n} z^n$ ($h \ge 0$) be a sequence of generating functions with nonnegative coefficients such that $a_{h,n}$ is nondecreasing in h and*

$$\lim_{h\to\infty} G_h(z) = G(z) = \sum_{n\ge 0} a_n z^n,$$

and let X_n denote the sequence of random variables with support $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ defined by

$$\mathbb{P}(X_n \le h) = \frac{a_{h,n}}{a_n}.$$

Assume, moreover, that each generating function G_h has a singularity at $\zeta_h \in \mathbb{R}$, such that

- *$\zeta_h = \zeta + c\rho^h + o(\rho^h)$ as $h \to \infty$ for some constants $\zeta > 0$, $c > 0$ and $\rho \in (0, 1)$.*
- *$G_h(z)$ can be continued analytically to the domain*

$$\{z \in \mathbb{C} : |z| \le (1+\delta)|\zeta_h|,\ |\arg(z/\zeta_h - 1)| > \phi\}$$

  for some fixed $\delta > 0$ and $\phi \in (0, \pi/2)$, and

$$G_h(z) = A_h(z) + C_h(1 - z/\zeta_h)^\alpha + o\big((1 - z/\zeta_h)^\alpha\big)$$

holds within this domain, uniformly in h, where $A_h(z)$ is analytic and uniformly bounded in h within the aforementioned region, $\alpha \in \mathbb{R} \setminus \mathbb{N}_0$, and C_h is a constant depending on h such that $\lim_{h\to\infty} C_h = C$. Finally,

$$G(z) = A(z) + C(1 - z/\zeta)^\alpha + o\big((1 - z/\zeta)^\alpha\big)$$

in the region

$$\{z \in \mathbb{C} : |z| \le (1+\delta)|\zeta|, \ |\arg(z/\zeta - 1)| > \phi\}$$

for a function $A(z)$ that is analytic within this region.

Then the asymptotic formula

$$\frac{a_{h,n}}{a_n} = \exp\left(-An\rho^h\right)\big(1 + o(1)\big) \tag{2}$$

holds as $n \to \infty$ and $h = \log_b n + O(1)$, where $b = \rho^{-1}$ and $A = c/\zeta$. Hence the shifted random variable $X_n - \log_b n$ converges weakly to a limiting distribution if n runs through a subset of the positive integers such that the fractional part $\{\log_b n\}$ of $\log_b n$ converges.

Remark 1 *Naturally, it is sufficient if all conditions hold only for $h \ge h_0$.*

Remark 2 *One can easily extend this theorem to the case when the behaviour at the dominant singularity also contains logarithmic terms.*

Remark 3 *The condition that the fractional part $\{\log_b n\}$ converges (modulo 1, i.e., 0 and 1 are identified) is both necessary and sufficient for weak convergence. Indeed, it follows from* (2) *that*

$$\mathbb{P}(X_n - \log_b n \le a) = \exp\left(-An\rho^{\lfloor \log_b n + a\rfloor}\right)(1 + o(1)) = \exp\left(-A\rho^{a - \{\log_b n + a\}}\right)(1 + o(1))$$

for all $a \in \mathbb{R}$, which shows the equivalence of weak convergence of random variables and convergence of $\{\log_b n\}$ modulo 1.

Under slightly stronger conditions, one can also prove an asymptotic formula for the mean, which shows the typical fluctuating behaviour in its second-order term. Higher order moments can be treated along the same lines, see [18].

Theorem 2 *In the setting of Theorem 1, assume additionally that*

1. *there exists a constant K such that $a_{h,n} = a_n$ for $h > Kn$,*
2. $\sum_{h\ge 0} |C - C_h| < \infty$,
3. *the asymptotic expansions of G_h and G around their singularities are given by*

$$G_h(z) = A_h(z) + C_h(1 - z/\zeta_h)^\alpha + B_h(1 - z/\zeta_h)^{\alpha+1} + o\big((1 - z/\zeta_h)^{\alpha+1}\big),$$

uniformly in h, and

$$G(z) = A(z) + C(1 - z/\zeta)^\alpha + B(1 - z/\zeta)^{\alpha+1} + o\big((1 - z/\zeta)^{\alpha+1}\big),$$

respectively, such that $\lim_{h\to\infty} B_h = B$.

Then the mean of X_n satisfies

$$\mathbb{E}(X_n) = \log_b n + \log_b A + \frac{\gamma}{\log b} + \frac{1}{2} + \psi_b\big(\log_b(An)\big) + o(1),$$

where γ denotes the Euler-Mascheroni constant, $A = c/\zeta$ as before and ψ_b is the 1-periodic function that is defined by the Fourier series

$$\psi_b(x) = -\frac{1}{\log b} \sum_{k \neq 0} \Gamma\Big(-\frac{2k\pi i}{\log b}\Big) e^{2k\pi i x}.$$

Remark 4 *The conditions are quite natural for combinatorial applications, but of course it is possible to modify them in various ways (e.g., by allowing further terms in the asymptotic expansions with exponents between α and $\alpha + 1$, with additional assumptions on the coefficients.)*

Proofs of these two theorems are provided in the following section. As they stand, they are quite general, but it might be tedious to check the conditions. Hence we discuss an important special case in Section 3. Thereafter, we consider a variety of combinatorial examples to which this general asymptotic scheme can be applied.

2 Proof of the main results

In order to prove Theorem 1, all that needs to be done is to invoke the Flajolet-Odlyzko singularity analysis, see for instance Chapter VI in [5]. By the uniformity condition, we have

$$a_{h,n} = \frac{C_h}{\Gamma(-\alpha)} n^{-\alpha-1} \zeta_h^{-n} (1 + o(1))$$

uniformly in h as $n \to \infty$ as well as

$$a_n = \frac{C}{\Gamma(-\alpha)} n^{-\alpha-1} \zeta^{-n} (1 + o(1)).$$

Since in addition $C_h \to C$ and $\zeta_h = \zeta + c\rho^h + o(\rho^h)$, we obtain

$$\begin{aligned} \frac{a_{h,n}}{a_n} &= \Big(\frac{\zeta_h}{\zeta}\Big)^{-n} (1 + o(1)) = \exp\Big(-\frac{cn}{\zeta}\rho^h + o(n\rho^h)\Big)(1 + o(1)) \\ &= \exp\Big(-\frac{cn}{\zeta}\rho^h\Big)(1 + o(1)) \end{aligned}$$

as $n, h \to \infty$ and $h = \log_b n + O(1)$. $\square$

For the proof of Theorem 2, we need to refine the estimate for $\frac{a_{h,n}}{a_n}$ above. Note first that the mean of X_n is given by

$$\mathbb{E}(X_n) = \sum_{h \geq 0} \mathbb{P}(X_n > h) = \sum_{h \geq 0} \Big(1 - \mathbb{P}(X_n \leq h)\Big) = \sum_{h \geq 0} \Big(1 - \frac{a_{h,n}}{a_n}\Big),$$

and by our assumptions, we can reduce this to a finite sum:

$$\mathbb{E}(X_n) = \sum_{0 \le h \le Kn} \Big(1 - \frac{a_{h,n}}{a_n}\Big).$$

Translating the asymptotic expansions of G_h and G around their singularities to asymptotic expansions of their coefficients by means of singularity analysis, we obtain

$$a_{h,n} = \frac{C_h}{\Gamma(-\alpha)} n^{-\alpha-1} \zeta_h^{-n} \Big(1 - \frac{(\alpha+1)B_h}{C_h n} + \frac{\alpha(\alpha+1)}{2n} + o(n^{-1})\Big),$$

uniformly in h, as well as

$$a_n = \frac{C}{\Gamma(-\alpha)} n^{-\alpha-1} \zeta^{-n} \Big(1 - \frac{(\alpha+1)B}{Cn} + \frac{\alpha(\alpha+1)}{2n} + o(n^{-1})\Big).$$

Putting them together, we find

$$\frac{a_{h,n}}{a_n} = \Big(\frac{\zeta_h}{\zeta}\Big)^{-n} \bigg(\frac{C_h}{C} - \frac{(\alpha+1)(B_h C - B C_h)}{C^2 n} + o(n^{-1})\bigg).$$

If now $h_0 = \lfloor (\log_b n)/2 \rfloor$, then

$$\frac{\zeta_{h_0}}{\zeta} = 1 + \frac{c}{\zeta} \rho^{h_0} + o(\rho^{h_0}) \ge 1 + \frac{c}{\zeta \sqrt{n}} + o(n^{-1/2}),$$

which implies that

$$\frac{a_{h_0,n}}{a_n} \ll \exp(-\kappa \sqrt{n}\,)$$

for a positive constant κ, and by monotonicity of $a_{h,n}$ we also have

$$\frac{a_{h,n}}{a_n} \ll \exp(-\kappa \sqrt{n}\,)$$

for $h \le h_0$. Moreover,

$$\exp(-cn/\zeta \cdot \rho^h) \ll \exp(-\kappa \sqrt{n}\,)$$

for $h \le h_0$, which finally means that

$$\sum_{h \le h_0} \Big(1 - \frac{a_{h,n}}{a_n}\Big) = \sum_{h \le h_0} \Big(1 - \exp(-cn/\zeta \cdot \rho^h)\Big) + O(h_0 \exp(-\kappa \sqrt{n}\,)). \tag{3}$$

Hence we can replace $\frac{a_{h,n}}{a_n}$ by $\exp(-cn/\zeta \cdot \rho^h)$ within this region at the expense of an error term that tends faster to zero than any power of n.

For $h > h_0$, note first that

$$\frac{(\alpha+1)(B_h C - B C_h)}{C^2} = o(1)$$

by the assumptions on B_h and C_h (uniformly in h), since h_0 tends to infinity as n does. By monotonicity of $a_{h,n}$, we also have $\zeta_0 \geq \zeta_1 \geq \cdots \geq \zeta$, hence $(\zeta_h/\zeta)^{-n} \leq 1$ for all h. It follows that

$$\frac{a_{h,n}}{a_n} = \frac{C_h}{C}\Big(\frac{\zeta_h}{\zeta}\Big)^{-n} + o(n^{-1})$$

for $h > h_0$, uniformly in h. Summation over $h_0 < h \leq Kn$ yields

$$\sum_{h_0<h\leq Kn}\Big(1-\frac{a_{h,n}}{a_n}\Big) = \sum_{h_0<h\leq Kn}\Big(1-\frac{C_h}{C}\Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big) + o(1).$$

Next we replace C_h by C at the expense of another error term:

$$\begin{aligned}\sum_{h_0<h\leq Kn}\Big(1-\frac{C_h}{C}\Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big) &= \sum_{h_0<h\leq Kn}\Big(1-\Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big) + \sum_{h_0<h\leq Kn}\frac{C-C_h}{C}\Big(\frac{\zeta_h}{\zeta}\Big)^{-n} \\ &= \sum_{h_0<h\leq Kn}\Big(1-\Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big) + O\Big(\sum_{h_0<h\leq Kn}|C-C_h|\Big) \\ &= \sum_{h_0<h\leq Kn}\Big(1-\Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big) + o(1), \qquad (4)\end{aligned}$$

since the sum $\sum_{h\geq 0}|C-C_h|$ converges by assumption, implying that $\sum_{h>h_0}|C-C_h|$ tends to 0 as n (and thus h_0) approaches infinity. Next we estimate the sum

$$\sum_{h_0<h\leq Kn}\Big(\exp\Big(-\frac{cn}{\zeta}\cdot\rho^h\Big) - \Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big). \qquad (5)$$

For this purpose, we need the following asymptotic formula, whose proof is a classical application of the Mellin transform (see for instance [4]):

$$\sum_{h\geq 0}\Big(1-\exp\big(-xb^{-h}\big)\Big) = \log_b x + \frac{\gamma}{\log b} + \frac{1}{2} + \psi_b(\log_b x) + O\Big(\frac{1}{x}\Big), \qquad (6)$$

where ψ_b is defined as in the statement of Theorem 2. Now let

$$\epsilon = \epsilon(n) = \sup_{h\geq h_0}\left|\frac{\zeta\rho^{-h}}{c}\log\frac{\zeta_h}{\zeta} - 1\right|$$

and note that $\epsilon \to 0$ as $n \to \infty$ by our assumptions on ζ_h. Then, for $h \geq h_0$,

$$\exp\Big(-\frac{c(1+\epsilon)n}{\zeta}\cdot\rho^h\Big) \leq \Big(\frac{\zeta_h}{\zeta}\Big)^{-n} \leq \exp\Big(-\frac{c(1-\epsilon)n}{\zeta}\cdot\rho^h\Big)$$

and thus

$$\begin{aligned}\sum_{h\ge 0}\Big(\exp\Big(-\frac{cn}{\zeta}\cdot\rho^h\Big)-\exp\Big(-\frac{c(1-\epsilon)n}{\zeta}\cdot\rho^h\Big)\Big)\\ &\le \sum_{h_0\le h\le Kn}\Big(\exp\Big(-\frac{cn}{\zeta}\cdot\rho^h\Big)-\Big(\frac{\zeta_h}{\zeta}\Big)^{-n}\Big)\\ &\le \sum_{h\ge 0}\Big(\exp\Big(-\frac{cn}{\zeta}\cdot\rho^h\Big)-\exp\Big(-\frac{c(1+\epsilon)n}{\zeta}\cdot\rho^h\Big)\Big).\end{aligned}\tag{7}$$

By (6), the left hand side of (7) is

$$\begin{aligned}&\sum_{h\ge 0}\Big(\exp\Big(-\frac{cn}{\zeta}\cdot\rho^h\Big)-\exp\Big(-\frac{c(1-\epsilon)n}{\zeta}\cdot\rho^h\Big)\Big)\\ &\quad=\sum_{h\ge 0}\Big(1-\exp\Big(-\frac{c(1-\epsilon)n}{\zeta}\cdot\rho^h\Big)\Big)-\sum_{h\ge 0}\Big(1-\exp\Big(-\frac{cn}{\zeta}\cdot\rho^h\Big)\Big)\\ &\quad=\log_b\frac{c(1-\epsilon)n}{\zeta}+\psi_b\Big(\log_b\frac{c(1-\epsilon)n}{\zeta}\Big)-\log_b\frac{cn}{\zeta}-\psi_b\Big(\log_b\frac{cn}{\zeta}\Big)+O(n^{-1})\\ &\quad=\log_b(1-\epsilon)+\psi_b\Big(\log_b\frac{cn}{\zeta}+\log_b(1-\epsilon)\Big)-\psi_b\Big(\log_b\frac{cn}{\zeta}\Big)+O(n^{-1})\\ &\quad=O(\epsilon+n^{-1}).\end{aligned}$$

The last step follows from the fact that ψ_b is periodic and infinitely differentiable in view of the exponential decay of the Fourier coefficients, hence Lipschitz continuous. An analogous estimate holds for the right hand side of (7) and thus also for the expression (5). Together with (4), we obtain

$$\sum_{h_0<h\le Kn}\Big(1-\frac{a_{h,n}}{a_n}\Big)=\sum_{h_0<h\le Kn}\Big(1-\exp(-cn/\zeta\cdot\rho^h)\Big)+o(1).$$

Now we combine this estimate with (3) to get

$$\mathbb{E}(X_n)=\sum_{0\le h\le Kn}\Big(1-\frac{a_{h,n}}{a_n}\Big)=\sum_{0\le h\le Kn}\Big(1-\exp(-cn/\zeta\cdot\rho^h)\Big)+o(1).$$

Moreover,

$$\sum_{h>Kn}\Big(1-\exp(-cn/\zeta\cdot\rho^h)\Big)=O\Big(\sum_{h>Kn}n\rho^h\Big)=O\Big(n\rho^{Kn}\Big),$$

and (6) with $b=1/\rho$ and $x=cn/\zeta$ yields

$$\sum_{h\ge 0}\Big(1-\exp(-cn/\zeta\cdot\rho^h)\Big)=\log_b\frac{cn}{\zeta}+\frac{\gamma}{\log b}+\frac{1}{2}+\psi_b\Big(\log_b\frac{cn}{\zeta}\Big)+O(n^{-1}).$$

Putting everything together, we end up with

$$\mathbb{E}(X_n)=\log_b\frac{cn}{\zeta}+\frac{\gamma}{\log b}+\frac{1}{2}+\psi_b\Big(\log_b\frac{cn}{\zeta}\Big)+o(1),$$

which completes the proof. □

3 A special case

An important special case of Theorem 1, which applies to many interesting examples as will be shown in the following section, occurs when the generating functions $G_h(z)$ can be written in terms of z and z^h:

Theorem 3 *Let $G_h(z) = \sum_{n\geq 0} a_{h,n} z^n$ ($h \geq 0$) be a sequence of generating functions with nonnegative coefficients such that $a_{h,n}$ is nondecreasing in h, and assume that G_h can be written as*

$$G_h(z) = A(z, z^h) + B(z, z^h) R(z, z^h)^\alpha$$

for $\alpha \in \mathbb{R} \setminus \mathbb{N}_0$ and bivariate functions $A(z,u)$, $B(z,u)$, $R(z,u)$ with real-valued Taylor coefficients at $(0,0)$ that are analytic for $|z| \leq M_1$ and $|u| \leq M_2$. Furthermore, suppose that

- *$r_0(z) := R(z,0)$ has a simple positive real root $\zeta < \min(M_1, 1)$ and no other roots in the region $\{z : |z| \leq \zeta\}$,*
- *$r_0(z)$ is positive for $0 < z < \zeta$,*
- *and $r_1(z) := \frac{\partial R}{\partial u}(z,0)$ is positive for $z = \zeta$.*

Then the conditions of Theorem 1 and Theorem 2 are satisfied for sufficiently large h with

$$\rho = \zeta, \quad c = -\frac{r_1(\zeta)}{r_0'(\zeta)} = -\frac{\frac{\partial R}{\partial u}(\zeta,0)}{\frac{\partial R}{\partial z}(\zeta,0)} \quad \textit{and} \quad A = -\frac{\frac{\partial R}{\partial u}(\zeta,0)}{\zeta \frac{\partial R}{\partial z}(\zeta,0)},$$

i.e., X_n follows the distribution of Theorem 1 and

$$\mathbb{E}(X_n) = \log_b n + \log_b A + \frac{\gamma}{\log b} + \frac{1}{2} + \psi_b\big(\log_b(An)\big) + o(1),$$

where $b = \rho^{-1} = \zeta^{-1}$.

Remark 5 *This theorem includes the special cases $\alpha = -1$ and $\alpha = 1/2$ treated in [14] and [25] respectively.*

Proof: We follow the same arguments as in [14] and [25]. For suitable $0 < \kappa < \min(1-\zeta, (M_1-\zeta)/2)$, there is no other root of r_0 than ζ inside the disk $\{z : |z| \leq \zeta + 2\kappa\}$, and the inequality

$$|R(z,z^h) - R(z,0)| \ll (\zeta+\kappa)^h < |R(z,0)|$$

holds on the circle $\{z : |z| = \zeta + \kappa\}$ for sufficiently large h. Rouché's theorem shows that $R(z,z^h)$ has exactly one simple root in the disk $\{z : |z| \leq \zeta + \kappa\}$ for sufficiently large h. Moreover, since $r_0(z)$ has a simple zero at ζ and is assumed to be positive for $z < \zeta$, it must be negative for $z > \zeta$ in a sufficiently small interval around ζ. This, together with our condition on $r_1(z)$, yields

$$\operatorname{sgn}(R(\zeta,\zeta^h)) = \operatorname{sgn}\Big(\frac{\partial R}{\partial u}(\zeta,0)\Big) = \operatorname{sgn}(r_1(\zeta)) = 1$$

and

$$\operatorname{sgn}\big(R(\zeta+h^{-1},(\zeta+h^{-1})^h)\big) = \operatorname{sgn}(R(\zeta+h^{-1},0)) = \operatorname{sgn}(r_0(\zeta+h^{-1})) = -1$$

for sufficiently large h, so there must be a real root $\zeta_h = \zeta + \epsilon_h$ of $R(z, z^h)$ with $0 < \epsilon_h < h^{-1}$. It follows that $\zeta_h^h \ll \zeta^h$ as $h \to \infty$. We can now determine the asymptotic behaviour of ζ_h by means of bootstrapping. Expanding $R(z, z^k)$ around $z = \zeta_h$ yields

$$0 = R(\zeta_h, \zeta_h^h) = \epsilon_h \cdot \frac{\partial R}{\partial z}(\zeta, 0) + O(\epsilon_h^2 + \zeta^h),$$

hence $\epsilon_h = O(\zeta^h)$ and $\zeta_h^h = \zeta^h + O(h\zeta^{2h})$, which gives us the more precise expansion

$$0 = R(\zeta_h, \zeta_h^h) = \epsilon_h \cdot \frac{\partial R}{\partial z}(\zeta, 0) + \zeta^h \frac{\partial R}{\partial u}(\zeta, 0) + O(\epsilon_h^2 + h\zeta^{2h}),$$

so $\epsilon_h = -\frac{\frac{\partial R}{\partial u}(\zeta,0)}{\frac{\partial R}{\partial z}(\zeta,0)}\zeta^h + O(h\zeta^{2h}) = c\zeta^h + O(h\zeta^{2h})$ and finally $\zeta_h = \zeta + c\zeta^h + O(h\zeta^{2h})$, which shows that the requirements of Theorem 1 are indeed satisfied with $\rho = \zeta$ and $c = -r_1(\zeta)/r_0'(\zeta)$.

In order to show that the conditions of Theorem 2 hold as well, note first that

$$[z^n]\left(A(z, z^h) + B(z, z^h)R(z, z^h)^\alpha\right) = [z^n]\left(A(z, 0) + B(z, 0)R(z, 0)^\alpha\right)$$

for $n < h$, so the first of our conditions is indeed satisfied. Now write $R(z, z^h) = (1 - z/\zeta_h)s_h(z)$. Then

$$\begin{aligned}
s_h(\zeta_h) &= -\zeta_h \frac{d}{dz} R(z, z^h)\Big|_{z=\zeta_h} \\
&= -\zeta_h \left(\frac{\partial R}{\partial z}(\zeta_h, \zeta_h^h) + h\zeta_h^{h-1}\frac{\partial R}{\partial u}(\zeta_h, \zeta_h^h)\right) \\
&= -\zeta_h \left(\frac{\partial R}{\partial z}(\zeta, 0)(1 + O(\epsilon_h + \zeta^h)) + h\zeta_h^{h-1}\frac{\partial R}{\partial u}(\zeta, 0)(1 + O(\epsilon_h + \zeta^h))\right) \\
&= -\zeta \frac{\partial R}{\partial z}(\zeta, 0)\left(1 + O(h\zeta^h)\right) \\
&= -\zeta r_0'(\zeta)\left(1 + O(h\zeta^h)\right).
\end{aligned}$$

Analogously, we find that

$$s_h'(\zeta_h) = -\frac{\zeta}{2} r_0''(\zeta)\left(1 + O(h^2\zeta^h)\right),$$

and that the second derivative $s_h''(z)$ is uniformly bounded in a suitable neighbourhood of ζ_h in view of the analyticity of $R(z, u)$. This yields an expansion

$$R(z, z^h) = (1 - z/\zeta_h)\left(\beta_{0h} + \beta_{1h}(1 - z/\zeta_h) + O(1 - z/\zeta_h)^2\right),$$

where $\beta_{ih} \to \beta_i$ at an exponential rate and the error term is uniform in h. The same can be done with $B(z, z^h)$, showing finally that properties (2) and (3) of Theorem 2 are indeed satisfied. □

4 Examples

4.1 Words and digital expansions

As it was mentioned in the introduction, this is perhaps the most classical example of our asymptotic scheme: *repetitions* of a letter in a word (equivalently, digits in a number), which was first discussed in the context of carry propagation by Knuth [17]. A more general treatment can be found in a paper of Guibas and Odlyzko [11]. First of all, it is easy to generalise the case of 0-1-sequences that was mentioned in the introduction: let $\mathcal{A}$ be an alphabet of k letters, and let $a \in \mathcal{A}$ be a specific letter. We are interested in the distribution of the longest run of a's in a random word of length n over this alphabet. Words with the property that no run of more than h consecutive a's occurs are specified by the symbolic description

$$\textsc{Seq}(\textsc{Seq}_{0..h}(a) \times (\mathcal{A} \setminus a)) \times \textsc{Seq}_{0..h}(1),$$

which yields the generating function

$$G_h(z) = \frac{1}{1 - \frac{(k-1)z(1-z^{h+1})}{1-z}} \cdot \frac{1-z^{h+1}}{1-z} = \frac{1-z^{h+1}}{1-kz+(k-1)z^{h+2}}.$$

Theorem 3 applies with $\alpha = -1$ and $R(z,u) = 1 - kz + (k-1)z^2u$. We find $\zeta = 1/k$, $c = (k-1)/k^3$ and $A = (k-1)/k^2$, and so the mean of the longest run of a's is

$$\log_k n + \frac{\gamma + \log(k-1)}{\log k} - \frac{3}{2} + \psi_k(\log_k((k-1)n)) + o(1).$$

Alternatively, one might be interested in the length of the longest run of *any* letter. If $G_h(z)$ is the generating function for nonempty words over the alphabet $\mathcal{A}$ such that no letter occurs more than h times in a row, then we have

$$G_h(z) = \frac{kz(1-z^h)}{1-z} + G_h(z) \cdot \frac{(k-1)z(1-z^h)}{1-z},$$

since such a word is either a single run of one of the k letters or obtained from a shorter word by appending a run of equal letters at the end. It follows that

$$G_h(z) = \frac{kz(1-z^h)}{1-kz+(k-1)z^{h+1}},$$

and Theorem 3, with $\alpha = -1$ and $R(z,u) = 1 - kz + (k-1)zu$, yields that the average length of the longest run of any letter is

$$\log_k n + \frac{\gamma + \log(k-1)}{\log k} - \frac{1}{2} + \psi_k(\log_k((k-1)n)) + o(1),$$

which is just one more than the average length of the longest run of a fixed letter.

This example can be generalised further in many ways. For instance, consider the longest sequence of zeros in the *Zeckendorf expansion* of an integer: recall that any positive integer can be written uniquely as a sum of nonconsecutive Fibonacci numbers, which gives rise to a digital expansion, e.g. 1000101 is the representation of 25 since $21+3+1=25$. Let us consider, for the sake of simplicity, the Zeckendorf expansions of all positive integers less than F_{n+2} (the $(n+2)$-th Fibonacci number), which have length at most n. Equivalently, these are all the 0-1-sequences of length at most n, starting with a 1, such that there are no consecutive 1's. Again, let us only consider those for which every run of 0's has length at most h. The symbolic description is now

$$\textsc{Seq}(1 \times \textsc{Seq}_{1..h}(0)) \times 1 \times \textsc{Seq}_{0..h}(0).$$

We get the generating function

$$G_h(z) = \frac{1}{1-z} \cdot \left(1 - \frac{z^2(1-z^h)}{1-z}\right)^{-1} \cdot z \cdot \frac{1-z^{h+1}}{1-z} = \frac{z(1-z^{h+1})}{(1-z)(1-z-z^2+z^{h+2})},$$

where the first factor $1/(1-z)$ takes the fact into account that we want to count sequences of length *at most* n rather than *precisely* n (which does not make a big difference though). Again it is plain to see that Theorem 3 applies, this time with $\alpha = -1$ and $R(z,u) = 1 - z - z^2 + z^2 u$. Specifically, we find the values $\zeta = (\sqrt{5}-1)/2$, $b = (\sqrt{5}+1)/2$, $c = (3\sqrt{5}-5)/10$ and $A = (5-\sqrt{5})/10$.

Another variant of the same problem is to consider *balanced* 0-1-sequences, i.e., sequences with the same number n of zeros and ones, as it was done in [24]. As we will see, the situation becomes somewhat more complicated in this example. It is advantageous to consider a bivariate generating function first in which z marks ones and w marks zeros. Using the symbolic description (1) that was already given in the introduction, we find the generating function

$$F_h(z,w) = \frac{1}{1 - \frac{1-z^{h+1}}{1-z} \cdot w} \cdot \frac{1-z^{h+1}}{1-z} = \frac{1-z^{h+1}}{1-z-w+wz^{h+1}}$$

for 0-1-sequences without runs of more than h consecutive ones. We need to extract the diagonal of this bivariate function, which can be done by means of a standard trick involving contour integration (see for instance [10]):

$$\begin{aligned} G_h(z) &= \frac{1}{2\pi i} \oint_{\mathcal{C}} F_h\left(t, \frac{z}{t}\right) \frac{dt}{t} \\ &= \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{1-t^{h+1}}{t-t^2-z+zt^{h+1}}\, dt \end{aligned}$$

for a suitable curve $\mathcal{C}$ surrounding 0. It follows that $G_h(z)$ is the residue of the integrand at the point $t = t_h(z)$ implicitly defined by $t - t^2 - z + zt^{h+1} = 0$ that lies closest to the origin (there is a unique solution of this equation that is analytic at 0 and satisfies $t_h(0) = 0$). The dominant singularity of $t_h(z)$ (and thus also $G_h(z)$) can be determined from the solution of the characteristic system that is formed by the equation and its partial derivative (cf. [5, Theorem VII.3]):

$$\begin{aligned} &t - t^2 - z + zt^{h+1} = 0, \\ &\frac{\partial}{\partial t}(t - t^2 - z + zt^{h+1}) = 1 - 2t + (h+1)zt^h = 0. \end{aligned}$$

One can reduce this system to a single equation for t and apply bootstrapping to find that the solution $t = \tau_h$ is asymptotically given by

$$\tau_h = \frac{1}{2} + (h+1)2^{-h-3} + O(h^2 2^{-2h}),$$

which in turn implies that the singularity of $G_h(z)$ (the variable z in the solution of the characteristic system) is located at

$$\zeta_h = \frac{1}{4} + 2^{-h-3} + O(h^2 2^{-2h}).$$

Now one finds that $t(z)$ has an expansion of the form

$$t(z) = \tau_h - a_h\sqrt{\zeta_h - z} + \cdots ,$$

where $a_h = 1 + O(h^2 2^{-h})$, and further

$$G_h(z) = \frac{1 - t(z)^{h+1}}{1 - 2t(z) + (h+1)zt(z)^{h+1}} = \frac{C_h}{\sqrt{1 - z/\zeta_h}} + \cdots ,$$

where $C_h = 1 + O(h^2 2^{-h})$ (further terms can be estimated as well). Now Theorems 1 and 2 apply again ($\alpha = -1/2$, $\rho = 1/2$, $c = 1/8$, $A = 1/2$), meaning that the longest run still follows the same kind of distribution in this case, with mean

$$\log_2 n + \frac{\gamma}{\log 2} - \frac{1}{2} + \psi_2(\log_2 n) + o(1).$$

Of course, there are many further problems of a similar nature; see for example a paper of Bassino, Clément and Nicaud [1], where runs in Lyndon words are considered. Further material on run statistics can also be found in [24].

4.2 Compositions

A *composition* of n is a way to write n as an ordered sum of positive integers. As in the previous section, we are interested in the distribution of the longest run, a problem considered in [9] (see also the recent papers by Wilf [26] and Gafni [6]). Let $G_h(z)$ denote the generating function for compositions whose longest run has length at most h, and let $G_{h,a}(z)$ be the generating function for the subclass of those compositions whose last term is a. Then we have

$$G_{h,a}(z) = (G_h(z) - G_{h,a}(z)) \cdot \frac{z^a(1 - z^{ah})}{1 - z^a},$$

since such compositions can be obtained by appending a run of a's to a composition that does not end in a. Solving for $G_{h,a}(z)$, we find

$$G_{h,a}(z) = \frac{z^a(1 - z^{ah})}{1 - z^{a(h+1)}} G_h(z),$$

and summing over all a yields

$$G_h(z) = 1 + \sum_{a \geq 1} G_{h,a}(z) = 1 + G_h(z) \cdot \sum_{a \geq 1} \frac{z^a(1 - z^{ah})}{1 - z^{a(h+1)}}$$

and finally

$$G_h(z) = \left(1 - \sum_{a \geq 1} \frac{z^a(1 - z^{ah})}{1 - z^{a(h+1)}}\right)^{-1}.$$

The special case $h = 1$ corresponds to so-called *Carlitz compositions*, see [16]. To see why this sequence of generating functions fits our general asymptotic scheme, write

$$\begin{aligned} G_h(z) &= \left(1 - \sum_{a \geq 1} z^a + \sum_{a \geq 1} \frac{z^{a(h+1)}(1 - z^a)}{1 - z^{a(h+1)}}\right)^{-1} \\ &= \left(\frac{1 - 2z}{1 - z} + \sum_{a \geq 1} \frac{z^{a(h+1)}(1 - z^a)}{1 - z^{a(h+1)}}\right)^{-1}. \end{aligned}$$

This is again a family of functions to which Theorem 3 applies: we have $\alpha = -1$ and

$$R(z, u) = \frac{1 - 2z}{1 - z} + \sum_{a \geq 1} \frac{(uz)^a(1 - z^a)}{1 - (uz)^a},$$

which is indeed analytic for z and u inside the unit circle. It is easy to see that $\zeta = 1/2$, and

$$\frac{\partial R}{\partial z}(z, 0) = -\frac{1}{(1 - z)^2}, \qquad \frac{\partial R}{\partial u}(z, 0) = z(1 - z),$$

hence $c = 1/16$, $A = 1/8$, and the mean length of the longest run in a composition is therefore

$$\log_2 n + \frac{\gamma}{\log 2} - \frac{5}{2} + \psi_2(\log_2 n) + o(1).$$

The largest part of a composition [23] can also be studied with our methods: note that the generating function for compositions whose parts are all $\leq h$ is given by

$$\frac{1}{1 - (z + z^2 + \cdots + z^h)} = \frac{1 - z}{1 - 2z + z^{h+1}},$$

and in the same way as in the previous section, we find that the mean of the largest part is

$$\log_2 n + \frac{\gamma}{\log 2} - \frac{3}{2} + \psi_2(\log_2 n) + o(1).$$

The number of distinct parts is closely related, see Hwang and Yeh [15] for details.

4.3 Geometric random variables

This example is essentially an extension of the previous one: the analogy between compositions and sequences of geometrically distributed random variables is well known and was also exploited in the aforementioned paper [9]. Consider sequences of n independent geometrically distributed random variables $X_1, X_2, \ldots, X_n$, where $\mathbb{P}(X_i = j) = pq^{j-1}$ for all $i, j \geq 1$ ($0 < p, q < 1$, $p + q = 1$). Once again, we study the behaviour of the longest run. This time, $G_h(z)$ denotes the generating function for the probability that the length of the longest run is at most h, and $G_{h,a}(z)$ is the generating function for this probability, restricted to the case that the value of the last random variable is a. Then we have, in complete analogy to the previous example,

$$G_{h,a}(z) = \big(G_h(z) - G_{h,a}(z)\big) \cdot \frac{pq^{a-1}z(1-(pq^{a-1}z)^h)}{1-pq^{a-1}z},$$

which yields

$$\begin{aligned} G_h(z) &= \left(1 - \sum_{a\ge 1} \frac{pq^{a-1}z(1-(pq^{a-1}z)^h)}{1-(pq^{a-1}z)^{h+1}}\right)^{-1} \\ &= \left(1 - \sum_{a\ge 1} pq^{a-1}z + \sum_{a\ge 1} \frac{(pq^{a-1}z)^{h+1}(1-pq^{a-1}z)}{1-(pq^{a-1}z)^{h+1}}\right)^{-1} \\ &= \left(1 - z + \sum_{a\ge 1} \frac{(pq^{a-1}z)^{h+1}(1-pq^{a-1}z)}{1-(pq^{a-1}z)^{h+1}}\right)^{-1}. \end{aligned}$$

Theorem 3 does not apply directly (it is not possible to write $G_h(z)$ as a function of z and z^h only that is otherwise independent of h), but the same bootstrapping mechanism can be employed to determine the behaviour of the dominant singularity. We are looking for zeros of the function

$$R_h(z) = 1 - z + \sum_{a\ge 1} \frac{(pq^{a-1}z)^{h+1}(1-pq^{a-1}z)}{1-(pq^{a-1}z)^{h+1}}.$$

The sum can be estimated as follows: if $|z| \le p^{-1/2}$, then

$$\begin{aligned} \left|\sum_{a\ge 1} \frac{(pq^{a-1}z)^{h+1}(1-pq^{a-1}z)}{1-(pq^{a-1}z)^{h+1}}\right| &\le \sum_{a\ge 1} \frac{(pq^{a-1}z)^{h+1}(1+\sqrt{p}\,)}{1-\sqrt{p}} \\ &= \frac{1+\sqrt{p}}{1-\sqrt{p}} \cdot \frac{(pz)^{h+1}}{1-q^{1+h}} \le \frac{1+\sqrt{p}}{1-\sqrt{p}} \cdot p^{(h-1)/2}, \end{aligned}$$

which is less than $p^{-1/2} - 1$ for sufficiently large h, hence Rouché's theorem (where we compare $R_h(z)$ with $1 - z$) shows that there is precisely one zero inside the circle $\{z \,:\, |z| \le p^{-1/2}\}$, and

$$R_h(1) > 0, \qquad R_h(p^{-1/2}) \le 1 - p^{-1/2} + \frac{1+\sqrt{p}}{1-\sqrt{p}} \cdot p^{(h-1)/2} < 0$$

for sufficiently large h, hence the zero ζ_h is real, positive and lies between 1 and $p^{-1/2}$. Moreover, the estimate above shows that $\zeta_h = 1 + O(p^{h/2})$, which in turn implies $\zeta_h^h = 1 + O(hp^{h/2})$. Thus we have

$$\begin{aligned} 0 &= 1 - \zeta_h + \sum_{a\geq 1} \frac{(pq^{a-1}\zeta_h)^{h+1}(1 - pq^{a-1}\zeta_h)}{1 - (pq^{a-1}\zeta_h)^{h+1}} \\ &= 1 - \zeta_h + \sum_{a\geq 1} (pq^{a-1})^{h+1}(1 - pq^{a-1}) \left(1 + O(hp^{h/2})\right) \\ &= 1 - \zeta_h + qp^{h+1} \left(1 + O(hp^{h/2} + q^h)\right), \end{aligned}$$

which finally yields $\zeta_h = 1 + qp^{h+1} + O(hp^{3h/2} + (pq)^h)$. From here, one can easily argue as in the proof of Theorem 3 to show that the conditions of Theorems 1 and 2 are satisfied with $\zeta = 1$, $\rho = p$ and $c = A = pq$. Hence the average of the longest run is

$$\log_b n - \frac{\gamma + \log q}{\log p} - \frac{1}{2} + \psi_b(\log_b n + \log_b q) + o(1)$$

with $b = p^{-1}$.

4.4 Paths

Just as in the other examples, the longest repetition of a certain pattern typically follows a double-exponential distribution—lattice paths such as Dyck paths or Motzkin paths are no exception here. Perhaps the simplest example of this type is the *longest horizontal segment* (i.e., the longest repetition of level steps) in a Motzkin path, which was considered in [25]. Any Motzkin path can be obtained from a Dyck path by inserting a sequence of level steps (possibly of length 0) after every up- or down-step, and possibly also adding a sequence of level steps at the beginning. This yields the generating function

$$\begin{aligned} G_h(z) &= \frac{1 - z^{h+1}}{1 - z} \cdot \frac{1 - \sqrt{1 - 4(z(1 - z^{h+1})/(1 - z))^2}}{2(z(1 - z^{h+1})/(1 - z))^2} \\ &= \frac{1 - z - \sqrt{1 - 2z - 3z^2 + 8z^{h+3} - 4z^{2h+4}}}{2z^2(1 - z^{h+1})}, \end{aligned}$$

to which Theorem 3 can be applied with $\alpha = 1/2$ and $R(z, u) = 1 - 2z - 3z^2 + 8z^3u - 4z^4u^2$. We find $\zeta = 1/3$, $c = 2/27$ and $A = 2/9$, so that the mean in this case is

$$\log_3 n + \frac{\gamma + \log 2}{\log 3} - \frac{3}{2} + \psi_3(\log_3 n + \log_3 2) + o(1).$$

The *longest plateau* (i.e., the longest horizontal segment in a Motzkin path that is preceded by an ascent and followed by a descent) is a very similar example. If one introduces a second variable v to the generating function for Dyck paths (cf. [25]) that marks peaks (by peak, we mean an ascent followed by a descent), one obtains

$$\frac{1 - (v - 1)x^2 - \sqrt{(1 - (v - 1)x^2)^2 - 4x^2}}{2x^2}.$$

We replace x by $z/(1-z)$, v by $1-z^{h+1}$ and multiply by another factor $1/(1-z)$ (for initial level steps) to obtain the generating function for Motzkin paths whose plateaus are no longer than h:

$$G_h(z) = \frac{(1-z)^2 + z^{h+3} - \sqrt{(1-z^2+z^{h+3})(1-4z+3z^2+z^{h+3})}}{2z^2(1-z)}.$$

We can apply Theorem 3 again, now with $\alpha = 1/2$ and $R(z,u) = (1-z^2+z^3u)(1-4z+3z^2+z^3u)$. Then $\zeta = 1/3$ again, $c = 1/54$, $A = 1/18$. Hence the length of the longest plateau in a Motzkin path follows the distribution of Theorem 1 with mean

$$\log_3 n + \frac{\gamma - \log 2}{\log 3} - \frac{3}{2} + \psi_3(\log_3 n - \log_3 2) + o(1).$$

In Dyck paths, there are no horizontal segments, but of course it is still possible to study repeated patterns: consider, for instance, the *longest zigzag sequence* in a Dyck path, i.e., the longest sequence of the form $udu\dots du$ (u standing for up-steps, d for down-steps), a problem that was raised in [20]; see also the recent paper [19].

To this end, let us first determine the generating function for Dyck paths that do not contain the pattern udu. This family can be symbolically described by

$$\mathcal{P} = \epsilon + ud + u(\mathcal{P} - \epsilon)d\mathcal{P},$$

where ϵ stands for the empty sequence. It follows easily that the generating function is given by

$$\frac{1 + x - \sqrt{1-2x-3x^2}}{2x},$$

which yields the Motzkin numbers. A Dyck path for which the longest zigzag sequence contains at most h steps up can then be obtained from a udu-free path by replacing every u by a sequence of the form $udu\dots du$ with at most h u's. This amounts to replacing x by $z(1-z^h)/(1-z)$ in the generating function, which yields

$$G_h(z) = \frac{1 - z^{h+1} - \sqrt{(1-z^{h+1})(1-4z+3z^{h+1})}}{2z(1-z^h)}.$$

Once again, Theorem 3 applies, this time with $\alpha = 1/2$ and $R(z,u) = (1-uz)(1-4z+3uz)$. We obtain $\zeta = 1/4$, $c = 3/16$ and $A = 3/4$. Therefore, the mean length (i.e., number of up-steps) of the longest zigzag sequence in a Dyck path of length $2n$ is

$$\log_4 n + \frac{\log 3 + \gamma}{\log 4} - \frac{1}{2} + \psi_4(\log_4 n + \log_4 3) + o(1).$$

In a similar vein, one can study the *longest ascent*, i.e., the longest sequence of the form u^h. The enumeration of Dyck paths avoiding a pattern of the form u^h (amongst other patterns) was recently studied in [22], albeit from a different point of view. Let $G_h(z)$ denote the generating function for Dyck paths whose longest ascent is of length $\leq h$, and let $G_{h,k}(z)$ denote the generating function for those Dyck paths among this class for which the initial ascent has length k. If $\mathcal{D}_h$ and $\mathcal{D}_{h,k}$ denote the respective families, then we have

$$\mathcal{D}_{h,k} = u\mathcal{D}_{h,k-1}d\mathcal{D}_h$$

for $k \leq h$, which translates into

$$G_{h,k}(z) = zG_{h,k-1}(z)G_h(z)$$

with initial value $G_{h,0}(z) = 1$ and $G_{h,k}(z) = 0$ for $k > h$. Hence we obtain

$$G_{h,k}(z) = z^k G_h(z)^k$$

and by summing over all $k \leq h$

$$G_h(z) = \sum_{k=0}^{h} G_{h,k}(z) = \frac{1-(zG_h(z))^{h+1}}{1-zG_h(z)}.$$

This example is different from the previous ones in this subsection in that it does not fit the scheme of Theorem 3. It is slightly easier to work with $F_h(z) = zG_h(z)$, which satisfies the functional equation

$$F_h(z) = z \cdot \frac{1-F_h(z)^{h+1}}{1-F_h(z)}.$$

This functional equation belongs to the general scheme $y(z) = z\phi(y(z))$, corresponding to simply generated families of trees, see for instance [5, Theorem VII.3]. It is well known that the dominant singularity is of square root type, and that it can be determined as the solution $(z, f) = (\zeta_h, \xi_h)$ to the characteristic system (as in the last example of Section 4.1):

$$\begin{aligned} f &= z \cdot \frac{1-f^{h+1}}{1-f}, \\ 1 &= z \cdot \frac{1-(h+1)f^h + hf^{h+1}}{(1-f)^2}, \end{aligned}$$

which yields

$$1 - 2f = (h-1)f^{h+2} - hf^{h+1},$$

and by bootstrapping $\xi_h = 1/2 + (h+1)2^{-h-3} + O(h^2 2^{-2h})$ as well as $\zeta_h = 1/4 + 2^{-h-3} + O(h^2 2^{-2h})$. Now one can verify the conditions of Theorem 1 and Theorem 2 in the same way as in the proof of Theorem 3 (or as it was shown earlier in Section 4.1 in the example on balanced 0-1-sequences) to obtain that the mean of the longest ascent in a Dyck path of length $2n$ is asymptotically

$$\log_2 n + \frac{\gamma}{\log 2} - \frac{3}{2} + \psi_2(\log_2 n) + o(1).$$

4.5 Trees

Yet another simple example that leads to the same type of behaviour is the *maximum outdegree* in a planted plane tree (i.e., a rooted tree in which the children of each vertex are ordered), cf. [3,21]. The family $\mathcal{T}_h$ of planted plane trees whose maximum outdegree is at most h is given by the recursive symbolic description

$$\mathcal{T}_h = \bullet \times \mathrm{SEQ}_{0..h}(\mathcal{T}_h),$$

from which we obtain

$$G_h(z) = z \cdot \frac{1 - G_h(z)^{h+1}}{1 - G_h(z)}$$

for the associated generating function $G_h(z)$. Note that this is essentially the same equation as in Section 4.4 for the longest sequence of ascents in Dyck paths. It follows that the two examples lead to the same distribution. A similar limit law also holds for the maximum degree of planar maps (logarithmic order, but slightly different distribution), see [7].

Another tree parameter that leads to the same type of distribution is the *longest chain of unary nodes* in a unary-binary tree, which was studied in [8, 25]. Suppose we want to count unary-binary trees with the property that there are no chains of more than h unary nodes. Such a tree consists of a chain of at most h unary nodes (possibly 0), starting at the root, followed by a binary node with two branches (or nothing). Hence the generating function $G_h(z)$ satisfies

$$G_h(z) = \frac{z(1 - z^{h+1})}{1 - z}(1 + G_h(z)^2),$$

from which one obtains

$$G_h(z) = \frac{1 - z + \sqrt{1 - 2z - 3z^2 + 8z^{3+h} - 4z^{2h+4}}}{2z(1 - z^{h+1})}.$$

This is essentially the same as the generating function for Motzkin paths with bounded sequences of level steps, as described in Section 4.4. One can also modify this example to other families of trees to obtain more interesting results.

Let us consider the longest chain of unary nodes in rooted labelled trees, and related to this problem, the longest path without branching (i.e., all internal nodes have degree 2) in a labelled (unrooted) tree. We first determine the (now exponential) generating function F_h for the family $\mathcal{R}_h$ of rooted labelled trees in which no chain of more than $h - 1$ successive unary nodes occurs. In the same way as before, we obtain

$$F_h(z) = \frac{z(1 - z^h)}{1 - z}\left(e^{F_h(z)} - F_h(z)\right)$$

or, after some simple manipulations,

$$F_h(z) = \frac{z(1 - z^h)}{1 - z^{h+1}} e^{F_h(z)}.$$

A rooted tree without a branchless path of length $> h$ belongs to one of the following classes:

- The root has degree 1, and its unique branch belongs to $\mathcal{R}_h$. The generating function for this case is clearly $zF_h(z)$.
- The root has degree 2: then it belongs to a path of length $2 \le \ell \le h$ whose internal nodes have degree 2, while the ends are either leaves or nodes of degree ≥ 3. The generating function is

$$\frac{1}{2}\sum_{\ell=1}^{h}(\ell-1)z^{\ell+1}\left(e^{F_h(z)}-F_h(z)\right)^2.$$

  Here, $\ell - 1$ gives the number of positions of the root within the path, and the factor $1/2$ takes symmetry into account.
- The tree only consists of the root, or the root has degree 3 or more, and the branches belong to $\mathcal{R}_h$: the generating function for this case is

$$z\left(e^{F_h(z)}-F_h(z)-\frac{F_h(z)^2}{2}\right).$$

Similarly, we consider edge-rooted labelled trees without branchless paths of length $> h$. The generating function can be obtained as in the second case above: it is given by

$$\frac{1}{2}\sum_{\ell=1}^{h}\ell z^{\ell+1}\left(e^{F_h(z)}-F_h(z)\right)^2,$$

since any edge has to belong to a branchless path of length $1 \le \ell \le h$. If we subtract the generating function for edge-rooted trees from the generating function for rooted trees, then every labelled tree (without root) is counted exactly once. The result is the generating function

$$\begin{aligned}G_h(z) = zF_h(z) &+ \frac{1}{2}\sum_{\ell=1}^{h}(\ell-1)z^{\ell+1}\left(e^{F_h(z)}-F_h(z)\right)^2 + z\left(e^{F_h(z)}-F_h(z)-\frac{F_h(z)^2}{2}\right)\\ &-\frac{1}{2}\sum_{\ell=1}^{h}\ell z^{\ell+1}\left(e^{F_h(z)}-F_h(z)\right)^2,\end{aligned}$$

which simplifies, after a couple of manipulations, to

$$G_h(z) = ze^{F_h(z)}\left(1-\frac{F_h(z)}{2}\right).$$

Let $T(z)$, defined implicitly by $T(z) = ze^{T(z)}$, be the exponential generating function for rooted labelled trees, which is well known to have a square-root type singularity at $1/e$. Then

$$F_h(z) = T\left(\frac{z(1-z^h)}{1-z^{h+1}}\right),$$

which means that the dominant singularity of G_h is given by the equation

$$\frac{\zeta_h(1-\zeta_h^h)}{1-\zeta_h^{h+1}} = \frac{1}{e},$$

and our usual bootstrapping procedure yields $\zeta_h = 1/e + (e-1)e^{-h-2} + O(he^{-2h})$. Note also that $T(z)$ has the following expansion around $1/e$:

$$T(z) = 1 - \sqrt{2(1-ez)} + \frac{2}{3}(1-ez) + \cdots,$$

and

$$e^{T(z)}\left(1 - \frac{T(z)}{2}\right) = \frac{e}{2} - \frac{e}{2}(1-ez) + \frac{2\sqrt{2}e}{3}(1-ez)^{3/2} + \cdots.$$

Now one can argue as in the proof of Theorem 3 that the conditions of Theorems 1 and 2 are satisfied (this time with $\alpha = 3/2$, $\rho = 1/e$, $c = (e-1)/e^2$ and $A = (e-1)/e$). Hence the distribution of the longest branchless path in a random labelled tree is again of the same type as in all our other examples, and the mean is

$$\log n + \gamma + \log(e-1) - \frac{1}{2} + \psi_e(\log n + \log(e-1)) + o(1).$$

The same analysis can be carried out with unlabelled trees, although the details are slightly more involved. It is also worth mentioning that the case $h = 1$ corresponds to *homeomorphically irreducible trees* (no vertices of degree 2), whose enumeration was already studied in [13] (see also [12]).

5 Conclusion

As we have seen, it is a very typical situation that an extremal parameter in a combinatorial structure follows a discrete double-exponential distribution, and that fluctuations in the average occur. The aim of this paper was to provide a unified approach to problems of this type and to illustrate this approach by means of a variety of examples, both old and new. There are certainly many more natural problems where the same kind of behaviour can be observed, and we hope that our results will be useful for future applications.

References

[1] F. Bassino, J. Clément, and C. Nicaud. The standard factorization of Lyndon words: an average point of view. *Discrete Math.*, 290(1):1–25, 2005.

[2] E. A. Bender and Z. Gao. Part sizes of smooth supercritical compositional structures. *Combin. Probab. Comput.*, 23(5):686–716, 2014.

[3] R. Carr, W. M. Y. Goh, and E. Schmutz. The maximum degree in a random tree and related problems. In *Proceedings of the Fifth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science (Poznań, 1991)*, volume 5, pages 13–24, 1994.

[4] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.*, 144(1-2):3–58, 1995.

[5] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[6] A. Gafni. Longest run of equal parts in a random integer composition. *Discrete Math.*, 338(2):236–247, 2015.

[7] Z. Gao and N. C. Wormald. The distribution of the maximum vertex degree in random planar maps. *J. Combin. Theory Ser. A*, 89(2):201–230, 2000.

[8] X. Gourdon. Largest component in random combinatorial structures. *Discrete Math.*, 180(1-3):185–209, 1998. Proceedings of the 7th Conference on Formal Power Series and Algebraic Combinatorics (Noisy-le-Grand, 1995).

[9] P. J. Grabner, A. Knopfmacher, and H. Prodinger. Combinatorics of geometrically distributed random variables: run statistics. *Theoret. Comput. Sci.*, 297(1-3):261–270, 2003. Latin American theoretical informatics (Punta del Este, 2000).

[10] D. H. Greene and D. E. Knuth. *Mathematics for the analysis of algorithms*, volume 1 of *Progress in Computer Science and Applied Logic*. Birkhäuser Boston Inc., Boston, MA, third edition, 1990.

[11] L. J. Guibas and A. M. Odlyzko. Long repetitive patterns in random sequences. *Z. Wahrsch. Verw. Gebiete*, 53(3):241–262, 1980.

[12] F. Harary and E. M. Palmer. *Graphical enumeration*. Academic Press, New York, 1973.

[13] F. Harary and G. Prins. The number of homeomorphically irreducible trees, and other species. *Acta Math.*, 101:141–162, 1959.

[14] C. Heuberger and H. Prodinger. Carry propagation in signed digit representations. *European J. Combin.*, 24(3):293–320, 2003.

[15] H.-K. Hwang and Y.-N. Yeh. Measures of distinctness for random partitions and compositions of an integer. *Adv. in Appl. Math.*, 19(3):378–414, 1997.

[16] A. Knopfmacher and H. Prodinger. On Carlitz compositions. *European J. Combin.*, 19(5):579–589, 1998.

[17] D. E. Knuth. The average time for carry propagation. *Nederl. Akad. Wetensch. Indag. Math.*, 40(2):238–242, 1978.

[18] G. Louchard and H. Prodinger. Asymptotics of the moments of extreme-value related distribution functions. *Algorithmica*, 46:431–467, 2006.

[19] K. Manes, A. Sapounakis, I. Tasoulas, and P. Tsikouras. Nonleft peaks in Dyck paths: a combinatorial approach. *Discrete Math.*, 337:97–105, 2014.

[20] T. Mansour. Statistics on Dyck paths. *Journal of Integer Sequences*, 9:Article 06.1.5, 2006.

[21] A. Meir and J. W. Moon. On the maximum out-degree in random trees. *Australas. J. Combin.*, 2:147–156, 1990. Combinatorial mathematics and combinatorial computing, Vol. 2 (Brisbane, 1989).

[22] H. Niederhausen and S. Sullivan. Pattern avoiding ballot paths and finite operator calculus. *J. Statist. Plann. Inference*, 140(8):2312–2320, 2010.

[23] A. Odlyzko and B. Richmond. On the compositions of an integer. In *Combinatorial mathematics, VII (Proc. Seventh Australian Conf., Univ. Newcastle, Newcastle, 1979)*, volume 829 of *Lecture Notes in Math.*, pages 199–210. Springer, Berlin-New York, 1980.

[24] H. Prodinger. Über längste 1-Folgen in $0,1$-Folgen. *Lecture Notes in Mathematics*, 1262:124–133, 1987.

[25] H. Prodinger and S. Wagner. Minimal and maximal plateau lengths in Motzkin paths. In *Proceedings of the 2007 Conference on Analysis of Algorithms, AofA 07*, Discrete Math. Theor. Comput. Sci. Proc., AH, pages 353–364. 2007.

[26] H. S. Wilf. The distribution of run lengths in integer compositions. *Electron. J. Combin.*, 18(2):Research Paper 23, 5pp., 2011.