ANALYSIS OF AN EXHAUSTIVE SEARCH ALGORITHM IN RANDOM GRAPHS AND THE $n^{c \log n}$-ASYMPTOTICS

CYRIL BANDERIER*, HSIEN-KUEI HWANG†, VLADY RAVELOMANANA‡, AND VYTAS ZACHAROVAS§

Abstract.
We analyze the cost used by a naive exhaustive search algorithm for finding a maximum independent set in random graphs under the usual $\mathscr{G}_{n,p}$-model where each possible edge appears independently with the same probability p. The expected cost turns out to be of the less common asymptotic order $n^{c \log n}$, which we explore from several different perspectives. Also we collect many instances where such an order appears, from algorithmics to analysis, from probability to algebra. The limiting distribution of the cost required by the algorithm under a purely idealized random model is proved to be normal. The approach we develop is of some generality and is amenable for other graph algorithms.

Key words. Asymptotic expansion, random graphs, graph algorithms, generating functions, Laplace transform, saddle-point method, Pantograph equation.

AMS subject classifications. 60C05 05A16 05C69 05C80 68W40 68R10

1. Introduction.

An *independent set* or *stable set* of a graph G is a subset of vertices in G no two of which are adjacent. The *Maximum Independent Set (MIS) Problem* consists in finding an independent set with the largest cardinality; it is among the first known NP-hard problems and has become a fundamental, representative, prototype instance of combinatorial optimization and computational complexity; see [27]. A large number of algorithms (exact or approximate, deterministic or randomized), as well as many applications, have been studied in the literature; see [6, 25, 68] and the references therein for more information.

The fact that there exist several problems that are essentially equivalent (including maximum clique and minimum node cover) adds particularly further dimensions to the algorithmic aspects and structural richness of the problem. One of the simplest ways of computing $\alpha(G)$, the cardinality of an MIS of G (or the stability number), is the following formulation in terms of polynomial optimization (see [1, 31])

$$\alpha(G) = \max_{(x_1,\dots,x_n)\in[0,1]^n} \left(\sum_{1\leqslant i\leqslant n} x_i - \sum_{(i,j)\in E} x_i x_j \right),$$

where E is the set of edges of G. Such an expression leads readily to an easily coded algorithm, but with deterministic exponential complexity $O(2^n)$. The algorithmic, theoretical and practical connections of many other formulations similar to this one have also been widely discussed; see [1].

Another simple means to find an MIS of a graph G is the following exhaustive (or branching or enumerative) algorithm. Start with any node, say v in G. Then either v is in an MIS or

*LIPN, Institut Galilée, Université Paris 13, 93430, Villetaneuse, France
†Institute of Statistical Science, Institute of Information Science, Academia Sinica, Taipei 115, Taiwan. Part of the work of this author was done while he was visiting ISM (Institute of Statistical Mathematics), Tokyo; he thanks ISM for its hospitality and support.
‡LIAFA, UMR CNRS 7089, Université Denis Diderot, 75205, Paris, France
§Dept. Mathematics & Informatics, Vilnius University, Vilnius, Lithuania

it is not. This leads to the recursive decomposition

$$\alpha(G) = \max\Big\{\underbrace{\alpha\left(G \setminus \{v\}\right)}_{v \notin \mathrm{MIS}(G)}, \underbrace{1 + \alpha\left(G \setminus N^*(v)\right)}_{v \in \mathrm{MIS}(G)}\Big\}, \tag{1.1}$$

where $\mathrm{MIS}(G)$ denotes an MIS of G and $N^*(v)$ denotes the union of v and all its neighbors. Such a simple procedure was the origin of many refined algorithms in the literature, including alternative formulations such as backtracking (see [67]) or branch and bound (see [25]).

Tarjan and Trojanowski [63] proposed an improved exhaustive algorithm with worst-case time complexity $O(2^{n/3})$. Their paper was followed and refined by many since then; see [6, 68] and [25] for more information and references. In particular, Chvátal [12] generalized Tarjan and Trojanowski's algorithm and showed *inter alia* that for almost all graphs with n nodes, a special class of algorithms (which he called order-driven) has time bound $O(n^{c_0 \log n+2})$, where $c_0 := 2/\log 2$. He also characterized exponential algorithms and conjectured that a similar bound of the form $O(n^{c \log n})$ holds for a wider class of recursive algorithms for some $c > 0$. Pittel [54] then refined Chvátal's bounds by showing that, under the usual $\mathscr{G}_{n,p}$-model (namely, *each pair of nodes has the same probability* $p \in (0,1)$ *of being connected by an edge, and one independent of the others*), the cost of Chvátal's algorithms (called f-driven, more general than order-driven) is bounded between $n^{(\frac14-\varepsilon)\log_\kappa n}$ and $n^{(\frac12+\varepsilon)\log_\kappa n}$ with high probability, for any $\varepsilon > 0$, where $q := 1-p$ and $\kappa := 1/q$.

The infrequent scale $n^{c\log n} = e^{c(\log n)^2}$ is central to our study here and can be seen through several different angles that will be examined in the following paragraphs. The simplest algorithmic connection to MIS problem is via the following argument. It is well-known that for any random graph G (under the $\mathscr{G}_{n,p}$-model), the value of $\alpha(G)$ is highly concentrated for fixed $p \in (0,1)$, namely, there exists a sequence m_n such that $\alpha(G) = m_n$ or $\alpha(G) = m_n + 1$ with high probability; see [5]. Asymptotically ($\kappa := 1/q$),

$$m_n = 2\log_\kappa n - 2\log_\kappa \log_\kappa n + O(1).$$

For more information on this and related estimates, see [5] and the references therein. Thus a simple randomized (approximate) MIS-finding algorithm consists in examining all possible

$$\binom{n}{m_n} + \binom{n}{m_n+1} = O\left(n^{2\log_\kappa n}\right)$$

subsets and determining if at least one of them is independent; otherwise (which happens with very small probability; see [5]), we resort to exhaustive algorithms such as that discussed in this paper.

From a different algorithmic viewpoint, Jerrum [39] studied the following Metropolis algorithm for maximum clique. Sequentially increase the clique, say K by (*i*) choose a vertex v uniformly at random; (*ii*) if $v \notin K$ and v is connected to every vertex of K, then add v to K; (*iii*) if $v \in K$, then v is subtracted from K with probability Λ^{-1}. He proved that for all $\Lambda \geqslant 1$, there exists an initial state from which the expected time for the Metropolis process to reach a clique of size at least $(1+\varepsilon)\log_\kappa(pn)$ exceeds $n^{\Omega(\log pn)}$. See [13] for an account of more recent developments on the complexity of the MIS problem.

We aim in this paper at a more precise analysis of the cost used by the simple recursive, exhaustive algorithm implied by (1.1). The exact details of the algorithm matter less and the overall cost is dominated by the total number of recursive calls, denoted by X_n, which is a random variable under the same $\mathscr{G}_{n,p}$-model. Then the mean value $\mu_n := \mathbb{E}(X_n)$ satisfies

$$\mu_n = \underbrace{\mu_{n-1}}_{v\notin \mathrm{MIS}(G)} + \underbrace{\sum_{0\leqslant k<n} \pi_{n,k}\mu_k}_{v\in \mathrm{MIS}(G)}, \tag{1.2}$$

for $n \geqslant 2$, with the initial conditions $\mu_0 = 0$ and $\mu_1 = 1$, where

$$\pi_{n,k} := \mathbb{P}(v \text{ has } n-1-k \text{ neighbors}) = \binom{n-1}{k} p^{n-1-k} q^k.$$

How fast does μ_n grow as a function of n? (*i*) If p is close to 1, then the graph is very dense and thus the sum in (1.2) is small (many nodes being removed), so we expect a polynomial time bound by simple iteration; (*ii*) If p is sufficiently small, then the second term is large, and we expect an exponential time bound; (*iii*) What happens for p in between? In this case the asymptotics of μ_n turns out to be nontrivial and we will show that

$$\log \mu_n = \frac{\left(\log \frac{n}{\log_\kappa n}\right)^2}{2\log\kappa} + \left(\tfrac{1}{2} + \tfrac{1}{\log\kappa}\right)\log n - \log\log n + P_0\left(\log_\kappa \tfrac{n}{\log_\kappa n}\right) + o(1), \tag{1.3}$$

where $P_0(t)$ is a bounded, periodic function of period 1. We will give a precise expression for P_0. Note that

$$\frac{\mu_n}{n^{\frac{1}{2}\log_\kappa n}} \asymp \frac{(\log n)^{\frac{1}{2}\log_\kappa \log n - 1 - \frac{\log\log\kappa}{\log\kappa}}}{n^{\log_\kappa \log n - \frac{1}{2} - \frac{1}{\log\kappa} - \frac{\log\log\kappa}{\log\kappa}}} \ll n^{-K} \to 0, \tag{1.4}$$

for any $K > 0$, where the symbol $a_n \asymp b_n$ means that a_n and b_n are asymptotically of the same order. Thus $\mu_n = o\left(n^{\frac{1}{2}\log_\kappa n - K}\right)$. On the other hand, the asymptotic pattern (1.3) is to some extent generic, as we will see below.

An intuitive way to see why we have the asymptotic form (1.3) for $\log \mu_n$ is to look at the simpler functional equation

$$\nu(x) = \nu(x-1) + \nu(qx), \tag{1.5}$$

since the binomial distribution is highly concentrated around its mean value pn, and we expect that $\mu_n \approx \nu(n)$ (under suitable initial conditions). This functional equation and the like (such as $\nu_n = \nu_{n-1} + \nu_{\lfloor qn\rfloor}$) has a rich literature. Most of them are connected to special integer partitions; important pointers are provided in Encyclopedia of Integer Sequences; see for example A000123, A002577, A005704, A005705, and A005706. In particular, it is connected to partitions of integers into powers of $\kappa = 1/q \geqslant 2$ when κ is a positive integer; see [15, 26, 48]. It is known that (under suitable initial conditions)

$$\log \nu(x) = \frac{\left(\log \frac{x}{\log_\kappa x}\right)^2}{2\log\kappa} + \left(\tfrac{1}{2} + \tfrac{1}{\log\kappa}\right)\log x - \log\log x + P_1\left(\log_\kappa \tfrac{x}{\log_\kappa x}\right) + o(1), \tag{1.6}$$

for large x, where $P_1(t)$ is a bounded 1-periodic function; see [15, 20]. Thus

$$|\log \mu_n - \log \nu(n)| = \left|P_0\left(\log_\kappa \tfrac{x}{\log_\kappa x}\right) - P_1\left(\log_\kappa \tfrac{x}{\log_\kappa x}\right)\right| + o(1).$$

We see that approximating the binomial distribution in (1.2) by its mean value

$$\mathbb{E}(\mu_{n-1-\mathrm{Binom}(n-1;p)}) \approx \mu_{n-1-\mathbb{E}(\mathrm{Binom}(n-1;p))} \approx \mu_{\lfloor qn \rfloor}$$

gives a very precise estimate, where $\mathrm{Binom}(n-1; p)$ denotes a binomial distribution with parameters $n-1$ and p.

An even simpler way to see the dominant order $x^{c \log x}$ is to approximate (1.5) by the *delay differential equation* (since $\nu(x) - \nu(x-1) \approx \nu'(x)$ for large x)

$$\omega'(x) = \omega(qx), \tag{1.7}$$

which is a special case of the so-called *"pantograph equations"*

$$\omega'(x) = a\omega(qx) + b\omega(x),$$

originally arising from the study of current collection systems for electric locomotives; see [37, 43, 51]. Since the usual polynomial or exponential functions fail to satisfy (1.7), we try instead a solution of the form $\omega(x) = x^{c \log x}$; then c should be chosen to satisfy the equation

$$x^{1-2c\log\kappa} = 2ce^{c(\log\kappa)^2} \log x.$$

So we should take $c = 1/(2\log\kappa) + O(x^{-1}\log x)$. This gives the dominant term $\frac{(\log x)^2}{2\log\kappa}$ for $\log\omega(x)$. More precise asymptotic solutions are thoroughly discussed in [16, 43]. In particular, all solutions of the equation $\omega'(x) = a\omega(qx)$ with $a > 0$ satisfies

$$\begin{aligned}\log\omega(x) = \frac{\left(\log \frac{x}{\log_\kappa x}\right)^2}{2\log\kappa} + \left(\tfrac{1}{2} + \tfrac{1}{\log\kappa} + \tfrac{\log a}{\log\kappa}\right)\log x - \left(1 + \tfrac{\log a}{\log\kappa}\right)\log\log x \\ + P_2\left(\log_\kappa \tfrac{x}{\log_\kappa x}\right) + o(1),\end{aligned}$$

for large x, where $P_2(t)$ is a bounded 1-periodic function. We see once again the generality of the asymptotic pattern (1.3).

On the other hand, the function

$$\varpi(x) := \exp\left(\frac{\left(\log(x/\sqrt{q})\right)^2}{2\log(1/q)}\right)$$

satisfies the q-difference equation

$$\varpi(x) = x\varpi(qx),$$

and is a fundamental factor in the asymptotic theory of q-difference equations; see the two survey papers [3, 18] and the references therein. This equation will also play an important role in our analysis.

From yet another angle, one easily checks that the series

$$M(x) := \sum_{j\geqslant 0} \frac{q^{\binom{j}{2}}}{j!} x^j$$

satisfies the equation (1.7). The largest term occurs, by simple calculus, at

$$j \approx \log_\kappa x - \log_\kappa \log_\kappa x + \tfrac{1}{2} + o(1),$$

and, by the analytic approach we use in this paper, we can deduce that the logarithm of the series is, up to an error of $O(1)$, of the same asymptotic order as $\log \nu(x)$; see (1.6) and Section 6. The function $M(x)$ arises sporadically in many different contexts and plays an important rôle in the corresponding asymptotic estimates; see Section 6 for a list of some representative references.

A closely related sum arises in the average-case analysis of a simple backtracking algorithm (see [67]), which corresponds to the expected number of independent sets in a random graph (or, equivalently, the expected number of cliques by interchanging q and p)

$$J_n := \sum_{1\leqslant j\leqslant n} \binom{n}{j} q^{j(j-1)/2}, \tag{1.8}$$

see [49, 67]. Wilf [67] showed that $J_n = O(n^{\log n})$ when $p = 1/2$. While such a crude bound is easily obtained, the more precise asymptotics of J_n is nontrivial. First, it is straightforward to check that $J_n \sim M(n)$ for large n. Second, the approach we develop in this paper can be used to show that J_n has an asymptotic expansion similar to (1.3). Indeed, it is readily checked that $J_n + 1$ satisfies the same recurrence relation as μ_n with different initial conditions. So the asymptotics of J_n follows the same pattern (1.3) as that of μ_n; see Section 6 for more details.

Thus examining all independent sets one after another in the backtracking style of Wilf [67] and identifying the one with the maximum cardinality also leads to an expected $n^{c\log n}$-complexity.

The diverse aspects we discussed of algorithms or equations leading to the scale $n^{c\log n}$ are summarized in Figure 1. The bridge connecting the algorithms and the analysis is the binomial recurrence (1.2) as explained above.

This paper is organized as follows. We derive in the next section an asymptotic expansion for μ_n using a purely analytic approach. The interest of deriving such a precise asymptotic approximation is at least fourfold.

- **Asymptotics:** It goes much beyond the crude description $n^{c\log n}$ and provides a more precise description; see particularly (1.4) and its implication mentioned there. Indeed, few papers in the literature address such an aspect; see [15, 16, 20, 43, 53, 57].
- **Numerics:** All scales involved in problems of similar nature here are expressed either in log or in log log, making them more subtle to be identified by numerical simulations. The inherent periodic functions and the slow convergence further add to the complications.
- **Methodology:** Our approach, different from previous ones that rely on explicit generating functions in product forms, is based on the underlying functional equation and is of some generality; it is akin to some extent to Mahler's analysis in [48].
- **Generality:** The asymptotic pattern (1.3) is also of some generality, an aspect already examined in details in several papers; see for example [16, 20, 43]. See also the last section for a list of diverse contexts where the order $n^{c\log n}$ appears.

Alternative approaches leading to different asymptotic expansions are discussed in Section 3. A rough estimate for μ_n was derived in [47] by an elementary approach.

The next curiosity after the expected value is the variance. But due to strong dependence of the subproblems, the variance is quite challenging at this stage. We consider instead an idealized *independent* version of X_n (the total cost of the exhaustive algorithm implied by (1.1)), namely

$$Y_n \overset{d}{=} Y_{n-1} + Y^*_{n-1-\mathrm{Binom}(n-1;p)} \qquad (n \geqslant 2), \tag{1.9}$$

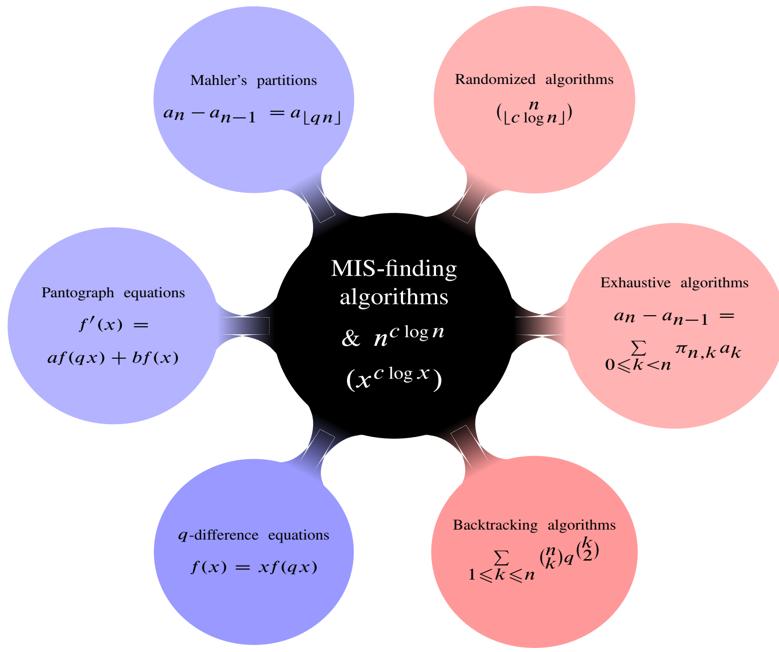

FIG. 1.1. The connection between MIS-finding algorithms and the scale $n^{c \log n}$ (discrete) or $x^{c \log x}$ (continuous). The circles on the right-hand side are more algorithmic in nature, while those on the left-hand side more analytic in nature.

with $Y_1 := 1$ and $Y_0 := 0$, where "$\stackrel{d}{=}$" stands for equality in distribution, Y_n^* is an identical copy of Y_n and the two terms on the right-hand side are *independent*. The original random variable X_n satisfies the same distributional recurrence (see (2.1)) but with the two terms (X_{n-1} and $X^*_{n-1-\text{Binom}(n-1;p)}$) on the right-hand side *dependent*. We expect that Y_n would provide an insight of the possible stochastic behavior of X_n although we were unable to evaluate their difference. We show, by a method of moments, that Y_n *is asymptotically normally distributed* in addition to deriving an asymptotic estimate for the variance. Monte Carlo simulations for n up to a few hundreds show that the limiting distribution of X_n seems likely to be normal, although the ratio between its variance and that of Y_n grows like a concave function. But the sample size n is not large enough to provide more convincing conclusions from simulations.

Once the asymptotic normality of Y_n is clarified, a natural question then is the limit law of the random variables (by changing the underlying binomial to uniform distribution)

$$Z_n \stackrel{d}{=} Z_{n-1} + Z_{\text{Uniform}(0,n-1)} \qquad (n \geqslant 2), \tag{1.10}$$

with $Z_0 = 0$ and $Z_1 = 1$. In this case, we prove that the mean is asymptotic to $cn^{-1/4}e^{2\sqrt{n}}$ and the limit law is *no more normal*. Surprisingly, the mean sequence, which is, up to a factor of $n!$, A005189 in Encyclopedia of Integer Sequences) also occurs in the study of the theory of measurement (and two-sided generalized Fibonacci sequences); see [21, 22].

We conclude this paper with a few remarks and a list of many instances where $n^{c \log n}$ arises, further clarifications and connections being given elsewhere.

Notations. Throughout this paper, $0 < p < 1$, $q = 1 - p$, and $\kappa = 1/q$.

2. Expected cost. We derive asymptotic approximations to μ_n in this section by an analytic approach, which is briefly sketched in Figure 2.1.

2.1. Preliminaries and main result. Recall that X_n denotes the cost used by the exhaustive search algorithm (implied by (1.1)) for finding an MIS in a random graph, and it satisfies the recurrence

$$X_n \stackrel{d}{=} X_{n-1} + X^*_{n-1-\mathrm{Binom}(n-1;p)}, \tag{2.1}$$

with $X_0 = 0$ and $X_1 = 1$, where $X_n^* \stackrel{d}{=} X_n$, and the two terms on the right-hand side are *dependent*.

From (2.1), we see that the expected value μ_n of X_n satisfies the recurrence (1.2). Our analytic approach then proceeds along the line depicted in Figure 2.1. While the approach appears standard (see [24, 38, 62]), the major difference is that instead of Mellin transform, we need Laplace transform since the quantity in question is not polynomially bounded. Another technical novelty is the justification of the analytic de-Poissonization for which we rely strongly on the manipulation of functional equations, differing significantly from previous approaches; [38, 62].

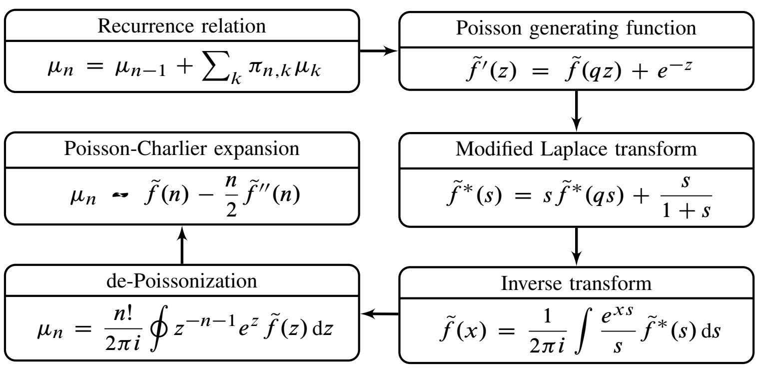

FIG. 2.1. Our analytic approach to the asymptotics of μ_n. Here $\pi_{n,k} := \binom{n-1}{k}q^k p^{n-1-k}$.

Generating functions (GFs).. Let $f(z) := \sum_{n\geqslant 0}\mu_n z^n/n!$ denote the exponential GFs of μ_n. Then f satisfies, by (1.2), the equation

$$f'(z) = 1 + f(z) + e^{pz}f(qz),$$

with $f(0) = 0$, or, equivalently, denoting by $\tilde f(z) := e^{-z}f(z)$ the Poisson GF of μ_n,

$$\tilde f'(z) = \tilde f(qz) + e^{-z}, \tag{2.2}$$

with $\tilde f(0) = 0$.

Closed-form expressions.. Let $\tilde f(z) = \sum_{n\geqslant 0}\tilde\mu_n z^n/n!$. From the q-differential equation (2.2), we derive the recurrence

$$\tilde\mu_{n+1} = q^n\tilde\mu_n + (-1)^n \qquad (n \geqslant 1).$$

By iteration, we then obtain the closed-form expression

$$\tilde\mu_n = \sum_{0\leqslant j<n}(-1)^j q^{(n-1-j)(n+j)/2} \qquad (n \geqslant 1).$$

Since $f(z) = e^z \tilde{f}(z)$, we then have

$$\mu_n = \sum_{1\leqslant k\leqslant n} \binom{n}{k} \sum_{0\leqslant j<k} (-1)^j q^{(k-1-j)(k+j)/2} \qquad (n \geqslant 1). \tag{2.3}$$

This expression is, although exact, less useful for large n; also its asymptotic behavior remains opaque, notably due to the appearance of $(-1)^j$. See also (3.4) for another closed-form expression for μ_n without any power of -1.

Asymptotic approximations.. Our aim in this section is to derive the following asymptotic approximation to μ_n.

THEOREM 2.1. *The expected cost μ_n of the exhaustive search on a random graph satisfies*

$$\mu_n = \frac{G\left(\log_\kappa \frac{n}{\log_\kappa n}\right)}{\sqrt{2\pi}} \cdot \frac{n^{1/\log\kappa+1/2}}{\log_\kappa n} \exp\left(\frac{\left(\log \frac{n}{\log_\kappa n}\right)^2}{2\log\kappa}\right)\left(1 + O\left(\frac{(\log\log n)^2}{\log n}\right)\right), \tag{2.4}$$

as $n \to \infty$, where $G(u)$ is defined by ($\{u\}$ being the fractional part of u)

$$G(u) = q^{(\{u\}^2-\{u\})/2} \sum_{j\in\mathbb{Z}} \frac{q^{j(j+1)/2}}{1+q^{j-\{u\}}} q^{-j\{u\}},$$

(see (2.8)) and is a bounded, 1-periodic function of u. Note that (2.4) implies (1.3) with

$$P_0(u) = -\tfrac{1}{2}\log 2\pi - \log\kappa + \log G(u).$$

Our approach leads indeed to an asymptotic expansion, but we content ourselves with the statement of (2.4); see (2.18), (2.23) and (3.3).

The function f (and thus $\tilde{f}$) is an entire function. It follows immediately that we have the identity

$$\mu_n = \sum_{j\geqslant 0} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n),$$

(referred to as the Poisson-Charlier expansion in [35]) where the $\tau_j(n)$'s are polynomials of n of degree $\lfloor j/2\rfloor$; see (2.24). See also [38] for different representations. However, the hard part always lies in justifying the *asymptotic nature* of the expansion, namely,

$$\mu_n = \sum_{0\leqslant j<k} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n) + O\left(n^{\lfloor k/2\rfloor} \tilde{f}^{(k)}(n)\right),$$

for $k = 2, 3, \ldots$. In particular, the first-order asymptotic equivalent "$\mu_n \sim \tilde{f}(n)$" is often called the *Poisson heuristic*. Thus the asymptotics of μ_n is reduced to that of $\tilde{f}(x)$ once we justify the asymptotic nature of the expansion. Of special mention is that, unlike almost all papers in the literature, we need only the asymptotic behavior of $\tilde{f}(x)$ for *real values of* x, all analysis involving complex parameters being carefully handled by the corresponding functional equation.

We will derive an asymptotic expansion for $\tilde{f}(x)$ for large real x by Laplace transform techniques and suitable manipulation of the saddle-point method, and then bridge the asymptotics of μ_n and $\tilde{f}(n)$ by a variant of the saddle-point method (or de-Poissonization procedure; see [38]); see Figure 2.1 for a sketch of our proof.

2.2. Asymptotics of $\tilde{f}(x)$. We derive an asymptotic expansion for $\tilde{f}(x)$ in this subsection.

Modified Laplace transform.. For technical convenience, consider the modified Laplace transform

$$\tilde{f}^{\star}(s) := \frac{1}{s}\int_0^{\infty} e^{-x/s}\tilde{f}(x)\,\mathrm{d}x.$$

Note that this use of the Laplace transform differs from the usual one by a factor $1/s$ and by a change of variables $s \mapsto 1/s$. Also the use of the exponential GF coupling with this Laplace transform is equivalent to considering the ordinary GF of μ_n; see Section 3.2 for more information.

Then the functional-differential equation (2.2) translates into the following functional equation for $\tilde{f}^{\star}$

$$\tilde{f}^{\star}(s) = s\tilde{f}^{\star}(qs) + \frac{s}{1+s}, \tag{2.5}$$

for $\Re(s) > 0$.

Iterating the equation (2.5) indefinitely, we get

$$\tilde{f}^{\star}(s) = \sum_{j\geqslant 0} \frac{q^{j(j+1)/2}}{1+q^j s}\, s^{j+1}. \tag{2.6}$$

We will approximate $\tilde{f}^{\star}(s)$ for large s by means of the function

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

because adding terms of the form s^{-j}, $j \geqslant 0$, does not alter the dominant asymptotic order of both functions for large $|s|$.

LEMMA 2.2. *For $x > 1$, we have*

$$F(x) = x^{1/2}\exp\left(\frac{(\log x)^2}{2\log\kappa}\right) G\left(\log_\kappa x\right), \tag{2.7}$$

where

$$G(u) := q^{(\{u\}^2+\{u\})/2} F\left(q^{-\{u\}}\right) \tag{2.8}$$

is a continuous, positive, periodic function with period 1.

Proof. One can easily check that $F(s)$ satisfies a functional equation similar to that of Jacobi's theta functions

$$F(s) = sF(qs) \qquad (s \in \mathbb{C}). \tag{2.9}$$

Iterating N times this functional equation, we obtain

$$F(s) = q^{N(N-1)/2} s^N F\left(q^N s\right) \qquad (s \in \mathbb{C}). \tag{2.10}$$

Assume $x > 1$. Take

$$N = \lfloor \log_\kappa x \rfloor = \log_\kappa x + \eta,$$

where $\eta = -\{\log_\kappa x\}$. Then we have

$$\begin{aligned} F(x) &= \exp\left(\frac{N(N-1)}{2}\log q + N\log x\right) F\left(e^{N\log q + \log x}\right) \\ &= \exp\left(\frac{(\log x)^2}{2\log\kappa} + \frac{\log x}{2} + \frac{\eta(\eta-1)}{2}\log q\right) F\left(e^{\eta\log q}\right) \\ &= q^{(\eta^2-\eta)/2} x^{1/2} \exp\left(\frac{(\log x)^2}{2\log\kappa}\right) F\left(e^{\eta\log q}\right), \end{aligned}$$

which, together with the functional equation $F(1/q) = F(1)/q$ (or $G(u+1) = G(u)$), proves the lemma. □

Asymptotic expansion of $\tilde{f}(x)$: saddle-point method. By the Laplace inversion formula with suitable change of variables, we have

$$\tilde{f}(x) = \frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s}\tilde{f}^\star\left(\frac{1}{s}\right) \mathrm{d}s, \tag{2.11}$$

where $r > 0$ is a small number whose value will be specified later. We now derive a few estimates for $\tilde{f}^\star(s)$.

LEMMA 2.3. (i) *If $r > 0$ and $|t| \geqslant 1$, then*

$$\tilde{f}^\star\left(\frac{1}{r+it}\right) = O\left(\frac{1}{|t|}\right); \tag{2.12}$$

(ii) *if $0 < r \leqslant 1$ and $|t| \leqslant 1$, then*

$$\tilde{f}^\star\left(\frac{1}{r+it}\right) = F\left(\frac{1}{r+it}\right) + O(1); \tag{2.13}$$

(iii) *if $r > 0$ and $c_m r \leqslant |t| \leqslant 1$, where $c_m := \sqrt{q^{-2m}-1}$, $m \geqslant 1$, then*

$$\tilde{f}^\star\left(\frac{1}{r+it}\right) = O\left(r^m q^{\binom{m}{2}} F\left(\frac{1}{r}\right)\right). \tag{2.14}$$

Proof. First, (2.12) follows from (2.6) since $\tilde{f}^\star(s) = O(|s|)$ as $|s| \to 0$. For the estimate (2.13), we observe that

$$\left|\frac{1}{1+sq^j}\right| \leqslant \min\{q^{-j}|s|^{-1}, 1\} \qquad (\Re(s) \geqslant 0).$$

Then

$$\tilde{f}^\star(s) = F(s) + O\left(|s|^{-1}\right),$$

for $\Re(s) \geqslant 0$ and $|s| \geqslant c > 0$. Also for $r > 0$

$$\Re\left(\frac{1}{r+it}\right) = \frac{r}{r^2+t^2} > 0;$$

and, for $|t| \leqslant 1$ and $0 < r \leqslant 1$

$$\frac{1}{|r+it|} \geqslant \frac{1}{\sqrt{2}}.$$

From these two estimates, we then deduce (2.13).

On the other hand if $\Re(s) \geqslant 0$, then

$$|\tilde{f}^{\star}(s)| \leqslant \sum_{j \geqslant 0} q^{j(j+1)/2}|s|^{j+1} \leqslant \vartheta(|s|),$$

where

$$\vartheta(x) := \sum_{-\infty<j<\infty} q^{j(j-1)/2} x^j.$$

It is easily checked that $\vartheta(x)$ satisfies the same functional equation (2.9) as $F(x)$, namely,

$$\vartheta(x) = x\vartheta(qx).$$

Thus, by the same arguments used for proving (2.7), we have, for $x > 1$,

$$\vartheta(x) = x^{1/2} \exp\left(\frac{(\log x)^2}{2\log\kappa}\right) g(\log_\kappa x),$$

where $g(x)$ is a continuous, bounded, periodic function. Comparing this expression with (2.7) for $F(x)$, we conclude that $\vartheta(x) = O(F(x))$ for $x \geqslant 1$.

Let $c_m := \sqrt{q^{-2m}-1}$, $m > 1$. Then, for $0 < r < 1$,

$$\begin{aligned}\max_{c_m r \leqslant |t| \leqslant 1}\left|\tilde{f}^{\star}\left(\frac{1}{r+it}\right)\right| &\leqslant \max_{c_m r \leqslant |t| \leqslant 1}\left|\vartheta\left(\frac{1}{\sqrt{r^2+t^2}}\right)\right| \\ &= \vartheta(q^m/r) \\ &= r^m q^{m(m-1)/2}\vartheta(1/r) \\ &= O\left(r^m q^{\binom{m}{2}} F(1/r)\right).\end{aligned}$$

This proves (2.14) and the lemma. □

By splitting the integral in (2.11) into three ranges $|t| \leqslant c_m r$, $c_m r < |t| \leqslant 1$, and $|t| > 1$, where $t = \Im(s)$, and then applying the estimates (2.12) and (2.14), we deduce that

$$\tilde{f}(x)e^{-xr} = I_r(x) + O\left(r^{m-1}q^{\binom{m}{2}}F(1/r) + 1\right), \tag{2.15}$$

where

$$I_r(x) := \frac{1}{2\pi}\int_{-c_m r}^{c_m r} \frac{e^{ixt}}{r+it} F\left(\frac{1}{r+it}\right) \mathrm{d}t.$$

It remains to evaluate more precisely the integral $I_r(x)$ by the saddle-point method.

We now take

$$N = \lfloor \log_\kappa(1/r) \rfloor = \log_\kappa(1/r) + \eta,$$

where $\eta = -\{\log_\kappa(1/r)\}$. Applying the functional equation (2.10) with $s = 1/(r+it)$, we get

$$I_r(x) = \frac{1}{2\pi}\int_{-c_m r}^{c_m r} \frac{e^{ixt}q^{N(N-1)/2}}{(r+it)^{N+1}} F\left(\frac{rq^\eta}{r+it}\right) \mathrm{d}t.$$

By the relation

$$F(1/r) = q^{N(N-1)/2} r^{-N} F(q^{\eta}),$$

we then have

$$\begin{aligned} I_r(x) &= \frac{F(1/r)}{2\pi r} \int_{-c_m r}^{c_m r} e^{ixt} \left(\frac{r}{r+it}\right)^{N+1} \frac{F(rq^{\eta}/(r+it))}{F(q^{\eta})}\,\mathrm{d}t \\ &= \frac{F(1/r)e^{xr}}{2\pi} \int_{-c_m}^{c_m} e^{irxt} \left(\frac{1}{1+it}\right)^{N+1} \frac{F(q^{\eta}/(1+it))}{F(q^{\eta})}\,\mathrm{d}t \\ &= \frac{F(1/r)e^{xr}}{2\pi} \int_{-c_m}^{c_m} e^{-xrt^2/2} H(t)\,\mathrm{d}t, \end{aligned}$$

where

$$H(t) := e^{xr(it-\log(1+it)+t^2/2)} \frac{F(q^{\eta}/(1+it))}{(1+it)^{1+\eta} F(q^{\eta})}.$$

So far the choice of $r > 0$ is arbitrary. For an optimal choice, take $r = r(x) > 0$ to be the approximate saddle-point such that

$$\frac{1}{r} \log \frac{1}{r} = x \log \kappa. \tag{2.16}$$

Note that r can be expressed in terms of the Lambert-W function (principal solution of the equation $W(x)e^{W(x)} = x$) as

$$r = \frac{W(x \log \kappa)}{x \log \kappa};$$

thus $\log(1/r) = W(x \log \kappa)$. Asymptotically,

$$W(x) = \log x - \log\log x + \frac{\log\log x}{\log x} + \frac{(\log\log x)^2 - 2\log\log x}{2(\log x)^2} + O\left(\frac{(\log\log x)^3}{(\log x)^3}\right), \tag{2.17}$$

as $x \to \infty$; see [14] for more information on W.

Since $m > 1$ is arbitrary and $r \asymp x^{-1} \log x$, the relation (2.15) is an asymptotic approximation, albeit less explicit.

To derive a more explicit expansion, we first observe that

$$e^{xr} F(1/r) = r^{-1/\log\kappa - 1/2} e^{(\log(1/r))^2/(2\log\kappa)} G(\log_{\kappa}(1/r)),$$

by (2.7) and (2.16). Then what remains is standard (see [24]): evaluating the integral in (2.15) by Laplace's method (a change of variable $t \mapsto t/\sqrt{xr}$ followed by an asymptotic expansion of $H(t/\sqrt{xr})$ for large xr and then an integration term by term), and we obtain the following expansion.

PROPOSITION 2.4. *With r given by (2.16), $\tilde{f}(x)$ satisfies the asymptotic expansion*

$$\tilde{f}(x) \sim \frac{e^{(\log(1/r))^2/(2\log\kappa)} G(\log_{\kappa}(1/r))}{r^{1/\log\kappa + 1/2}\sqrt{2\pi \log_{\kappa}(1/r)}} \left(1 + \sum_{j\geqslant 1} \phi_j(\log_{\kappa}(1/r))(\log_{\kappa}(1/r))^{-j}\right), \tag{2.18}$$

as $x \to \infty$, where G is given in (2.8) and the $\phi_j(u)$'s are bounded, 1-periodic functions of u involving the derivatives of $F\left(q^{-\{u\}}\right)$. In particular,

$$\phi_1(u) = -\left(\frac{1}{12} - \frac{\{u\}(1-\{u\})}{2} + \frac{(1-\{u\})q^{-\{u\}}F'\left(q^{-\{u\}}\right)}{F\left(q^{-\{u\}}\right)} + \frac{q^{-2\{u\}}F''\left(q^{-\{u\}}\right)}{2F\left(q^{-\{u\}}\right)}\right).$$

By using (2.17), the leading term in (2.18) can be expressed completely in terms of $\log x$ as follows.

COROLLARY 2.5. *As $x \to \infty$, $\tilde{f}(x)$ satisfies*

$$\tilde{f}(x) = \frac{G\left(\log_\kappa \frac{x}{\log_\kappa x}\right)}{\sqrt{2\pi}} \cdot \frac{x^{1/\log\kappa + 1/2}}{\log_\kappa x} \exp\left(\frac{\left(\log \frac{x}{\log_\kappa x}\right)^2}{2\log\kappa}\right)\left(1 + O\left(\frac{(\log\log x)^2}{\log x}\right)\right). \tag{2.19}$$

This gives Theorem 2.1 with x here replaced by n.

As another consequence, we see, by (2.2) and (2.19), that

$$\frac{\tilde{f}'(x)}{\tilde{f}(x)} \sim \frac{\tilde{f}(qx)}{\tilde{f}(x)} \sim \frac{\log_\kappa x}{x}.$$

More generally, we have the following asymptotic relations for $\tilde{f}^{(j)}(x)$ and $\tilde{f}(q^j x)$.

COROLLARY 2.6. *For $j \geqslant 1$*

$$\frac{\tilde{f}^{(j)}(x)}{\tilde{f}(x)} \sim \left(\frac{\log_\kappa x}{x}\right)^j \tag{2.20}$$

$$\frac{\tilde{f}(q^j x)}{\tilde{f}(x)} \sim q^{-j(j-1)/2}\left(\frac{\log_\kappa x}{x}\right)^j. \tag{2.21}$$

Note that (2.20) also follows easily from the integral representation

$$\tilde{f}^{(j)}(x) = \frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s^{j-1}} \tilde{f}^\star\left(\frac{1}{s}\right) \mathrm{d}s,$$

and exactly the same arguments used above.

2.3. Asymptotics of μ_n. We first derive a simple lemma for the ratio $f(x+y)/f(x)$ when y is not too large by using (2.20).

LEMMA 2.7. *Assume $x > 1$. If $|y| = o(x/\log x)$, then*

$$\frac{\tilde{f}(x+y)}{\tilde{f}(x)} = 1 + O\left(\frac{|y|\log x}{x}\right). \tag{2.22}$$

Proof. By (2.20), we have

$$\begin{aligned}\log\frac{\tilde{f}(x+y)}{\tilde{f}(x)} &= y\int_0^1 \frac{\tilde{f}'(x+yt)}{\tilde{f}(x+yt)}\,\mathrm{d}t \\ &= yO\left(\int_0^1 \frac{\log|x+yt|}{|x+yt|}\,\mathrm{d}t\right) \\ &= O\left(\frac{|y|\log|x|}{|x|}\right),\end{aligned}$$

from which (2.22) follows. □

THEOREM 2.8. *The expected cost used by the exhaustive search algorithm satisfies the asymptotic expansion*

$$\mu_n \sim \tilde{f}(n) + \sum_{j\geqslant 2} \frac{\tilde{f}^{(j)}(n)}{j!}\,\tau_j(n), \tag{2.23}$$

where $\tau_j(n)$ is a (Charlier) polynomial in n of degree $\lfloor j/2\rfloor$ defined by

$$\tau_j(n) := \sum_{0\leqslant \ell\leqslant j} \binom{j}{\ell}(-1)^{\ell}\frac{n!n^{\ell}}{(n-j+\ell)!} \qquad (j=0,1,\dots). \tag{2.24}$$

In particular, $\tau_0(n)=1$, $\tau_1(n)=0$, $\tau_2(n)=-n$, $\tau_3(n)=2n$, and $\tau_4(n)=3n^2-6n$. Thus, by (2.18) and (2.20),

$$\mu_n = \tilde{f}(n)\left(1+O\left(n^{-1}(\log n)^2\right)\right),$$

which proves Theorem 2.1.

Proof. For simplicity, we prove only the following estimate

$$\mu_n = \tilde{f}(n) - \frac{n}{2}\,\tilde{f}''(n) + O\left(n^{-2}(\log n)^4\,\tilde{f}(n)\right). \tag{2.25}$$

The same method of proof easily extends to the proof of (2.23).

We start with the Taylor expansion of $\tilde{f}(z)$ at $z=n$ to the fourth order

$$\tilde{f}(z) = \tilde{f}(n) + \tilde{f}'(n)(z-n) + \frac{\tilde{f}''(n)}{2!}(z-n)^2 + \frac{\tilde{f}'''(n)}{3!}(z-n)^3 + (z-n)^4R(z), \tag{2.26}$$

where

$$R(z) = \frac{1}{3!}\int_0^1 \tilde{f}^{(4)}\big(n+(z-n)t\big)(1-t)^3\,\mathrm{d}t.$$

By applying successively the equation (2.2), we get

$$\tilde{f}^{(4)}(z) = -e^{-z} + q^3e^{-qz} - q^5e^{-q^2z} + q^6e^{-q^3z} + q^6\tilde{f}(q^4z).$$

It follows that

$$\begin{aligned}\left|R\left(ne^{i\theta}\right)\right| &\leqslant \int_0^1 \left|\tilde{f}^{(4)}\big(n+n(e^{i\theta}-1)t\big)\right|\mathrm{d}t\\ &= O\left(e^{-n\cos\theta} + e^{-q^3n\cos\theta} + \int_0^1\left|\tilde{f}\big(q^4n+q^4n(e^{i\theta}-1)t\big)\right|\mathrm{d}t\right),\end{aligned}$$

for $|\theta|\leqslant\pi$. Replacing first $\tilde{f}(z)$ inside the integral by $e^{-z}f(z)$, using the inequality $|f(z)|\leqslant f(|z|)$ and then substituting back $f(q^4n)$ by $e^{q^4n}\tilde{f}(q^4n)$, we then have

$$\begin{aligned}\left|R\left(ne^{i\theta}\right)\right| &= O\left(e^{-q^3n\cos\theta} + f(q^4n)\int_0^1\left|e^{-q^4n-q^4n(e^{i\theta}-1)t}\right|\mathrm{d}t\right)\\ &= O\left(e^{-q^3n\cos\theta} + \tilde{f}(q^4n)\int_0^1 e^{q^4n(1-\cos\theta)t}\,\mathrm{d}t\right)\\ &= O\left(e^{-q^3n\cos\theta} + \tilde{f}(q^4n)e^{q^4n(1-\cos\theta)}\right),\end{aligned} \tag{2.27}$$

uniformly for $|\theta| \leqslant \pi$. By Cauchy's integral formula and (2.26), we have

$$\begin{aligned}
\mu_n &= \frac{n!}{2\pi i} \oint_{|z|=n} z^{-n-1} e^z \tilde{f}(z)\, \mathrm{d}z \\
&= \frac{n!}{2\pi i} \oint_{|z|=n} z^{-n-1} e^z \left(\tilde{f}(n) + \frac{\tilde{f}'(n)}{1!}(z-n) + \frac{\tilde{f}''(n)}{2!}(z-n)^2 + \frac{\tilde{f}'''(n)}{3!}(z-n)^3 \right) \mathrm{d}z \\
&\quad + R_n \\
&= \tilde{f}(n) - \frac{n}{2}\tilde{f}''(n) + \frac{n}{3}\tilde{f}'''(n) + R_n,
\end{aligned}$$

where

$$R_n := \frac{n!}{2\pi i} \oint_{|z|=n} z^{-n-1} e^z (z-n)^4 R(z)\, \mathrm{d}z.$$

By the estimate (2.27) for $R(z)$, we have

$$\begin{aligned}
R_n &= O\left(n! n^{4-n} \int_{-\pi}^{\pi} \theta^4 e^{n\cos\theta} |R(ne^{i\theta})|\, \mathrm{d}\theta \right) \\
&= O\left(n! n^{4-n} \int_{-\pi}^{\pi} \theta^4 e^{n\cos\theta} \left(e^{-q^3 n\cos\theta} + \tilde{f}(q^4 n) e^{q^4 n(1-\cos\theta)} \right) \mathrm{d}\theta \right) \\
&= O\left(n! n^{4-n} \int_{-\pi}^{\pi} \theta^4 e^{n(1-q^3)\cos\theta}\, \mathrm{d}\theta + n! \tilde{f}(q^4 n) n^{4-n} e^n \int_{-\pi}^{\pi} \theta^4 e^{-(1-q^4)n(1-\cos\theta)}\, \mathrm{d}\theta \right) \\
&= O\left(n! n^{-n+3/2} e^{(1-q^3)n} + n! e^n n^{-n+3/2} \tilde{f}(q^4 n) \right) \\
&= O\left(n^2 e^{-q^3 n} + n^2 \tilde{f}(q^4 n) \right) \\
&= O\left(n^{-2} (\log n)^4 \tilde{f}(n) \right),
\end{aligned}$$

by (2.21). Note that again by (2.20)

$$n\tilde{f}'''(n) = O\left(n^{-2} (\log n)^3 \tilde{f}(n) \right),$$

so this error bound is absorbed in $O(\tilde{f}(n) n^{-2} (\log n)^4)$. This proves (2.25). □

3. Alternative expansions and approaches. For more methodological interest, we discuss in this section other possible approaches to the asymptotic expansions we derived above.

3.1. An alternative expansion for $\tilde{f}(x)$. We begin with an alternative asymptotic expansion for $\tilde{f}(x)$, starting from the integral representation (2.11), which, as showed above, can be approximated by

$$\tilde{f}(x) = \frac{1}{2\pi i} \int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s} F\left(\frac{1}{s}\right) \mathrm{d}s + O(1)$$

For simplicity, we drop the $O(1)$-term and write this as

$$\tilde{f}(x) \simeq \frac{1}{2\pi i} \int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s} F\left(\frac{1}{s}\right) \mathrm{d}s.$$

Now we use the same $N = \lfloor \log_\kappa(1/r) \rfloor = \log_\kappa(1/r) - \eta$ and

$$F\left(\frac{1}{s}\right) = q^{N(N-1)/2} s^{-N} F\left(\frac{q^N}{s}\right),$$

so that

$$\tilde{f}(x) \simeq \frac{q^{\binom{N}{2}}}{2\pi i} \int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s^{N+1}} F\left(\frac{q^N}{s}\right) \mathrm{d}s. \tag{3.1}$$

Now instead of expanding $F(q^N/(r+it))$ at $t=0$, we expand $F(q^N/s)$ at $s=r$, giving

$$F\left(\frac{q^N}{s}\right) = F\left(\frac{q^N}{r} - \frac{q^N}{r}\left(1-\frac{r}{s}\right)\right) = \sum_{m\geqslant 0} \frac{(-1)^m Q^m}{m!} F_m \left(1-\frac{r}{s}\right)^m,$$

where $Q := q^N/r = q^{-\{\log_\kappa(1/r)\}}$ and F_j denotes $F^{(j)}(Q)$. Substituting this expansion into the integral representation (3.1) and then integrating term-by-term, we obtain

$$\begin{aligned}
\tilde{f}(x) q^{-\binom{N}{2}} &\simeq \sum_{m\geqslant 0} \frac{(-1)^m Q^m}{m!} F_m \cdot \frac{1}{2\pi i} \int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s^{N+1}} \left(1-\frac{r}{s}\right)^m \mathrm{d}s \\
&= \frac{x^N}{N!} \sum_{m\geqslant 0} \frac{(-1)^m Q^m}{m!} F_m T_m(N),
\end{aligned} \tag{3.2}$$

where, by the integral representation for Gamma function (see [24]),

$$\begin{aligned}
T_m(N) &:= \frac{1}{2\pi i} \int_{r-i\infty}^{r+i\infty} \frac{e^{xs}}{s^{N+1}} \left(1-\frac{r}{s}\right)^m \mathrm{d}s \\
&= \sum_{0\leqslant j\leqslant m} \binom{m}{j} (-r)^j \frac{N! x^j}{(N+j)!}.
\end{aligned}$$

For computational purposes, it is preferable to use the recurrence

$$T_m(N) = T_{m-1}(N) - \frac{rx}{N+1} T_{m-1}(N+1).$$

The value of r is arbitrary up to now. If we take $r = N/x$, then

$$T_m(N) := \sum_{0\leqslant j\leqslant m} \binom{m}{j} (-1)^j \frac{N! N^j}{(N+j)!}.$$

Note that $|T_m(N)| \asymp N^{-\lceil m/2 \rceil}$. In particular,

$$T_0(N) = 1,\ T_1(N) = \frac{1}{N+1},\ T_3(N) = -\frac{N-2}{(N+1)(N+2)},\ \cdots.$$

Since q^N/r remains bounded, we can regroup the terms and get an asymptotic expansion in terms of increasing powers of N^{-1}, the first few terms being given as follows

$$\begin{aligned}
\frac{\tilde{f}(x)}{q^{\binom{N}{2}} x^N/N!} &\simeq F_0 - \frac{Q(2F_1 + F_2 Q)}{2N} + \frac{Q(3F_4 Q^3 + 28 F_3 Q^2 + 60 F_2 Q + 24 F_1)}{24 N^2} \\
&\quad - \frac{Q(F_6 Q^5 + 22 F_5 Q^4 + 152 F_4 Q^3 + 384 F_3 Q^2 + 312 F_2 Q + 48 F_1)}{48 N^3} \\
&\quad + \cdots.
\end{aligned}$$

On the other hand, if we choose $r = (N+1)/x$, then $T_1(N) = 0$ and

$$T_0(N) = 1,\ T_2(N) = -\frac{1}{N+2},\ T_3(N) = -\frac{4}{(N+2)(N+3)},\ \cdots,$$

so that

$$\begin{aligned}\frac{\tilde{f}(x)}{q^{\binom{N}{2}}x^N/N!} \simeq F_0 - \frac{F_2 Q^2}{2(N+2)} + \frac{Q^3(3F_4Q + 16F_3)}{24(N+2)^2}\\ - \frac{Q^3(F_6Q^3 + 16F_5Q^2 + 60F_4Q + 32F_3)}{48(N+2)^3} + \cdots .\end{aligned}$$

While $|T_m(N)| \asymp N^{-\lceil m/2\rceil}$ for $m \geqslant 2$ as in the case of $r = N/x$, this is a better expansion because the first term incorporates more information.

The more transparent expansion (3.2) is *a priori* a *formal* one whose asymptotic nature can be justified by the same local analysis as above. We summarize the analysis in the following theorem.

THEOREM 3.1. *The Poisson generating function of μ_n satisfies the asymptotic expansion*

$$\tilde{f}(x) \sim q^{\binom{N}{2}}\frac{x^N}{N!}\sum_{m\geqslant 0}\frac{(-1)^m Q^m}{m!}F^{(m)}(Q)T_m(N), \tag{3.3}$$

where $N = \lfloor \log_\kappa(1/r)\rfloor = \log_\kappa(1/r) - \eta$, $r := N/x$, $Q := q^{-\log_\kappa(1/r)}$ and $T_m(N)$ is defined by

$$T_m(N) := \sum_{0\leqslant j\leqslant m}\binom{m}{j}(-1)^j\frac{N!(N+1)^j}{(N+j)!}.$$

Straightforward calculations give (when $r = N/x$)

$$\begin{aligned}\log\left(q^{\binom{N}{2}}\frac{x^N}{N!}\right) = \frac{\left(\log\frac{x}{\log_\kappa x}\right)^2}{2\log\kappa} + \left(\frac{1}{\log\kappa} + \frac{1}{2}\right)\log x - \log\log x\\ - \frac{1}{2}\log 2\pi - \frac{\eta^2+\eta}{2} + O\left(\frac{(\log\log x)^2}{\log x}\right),\end{aligned}$$

consistent with what we proved in (2.19) via directly applying the saddle-point method. For similar types of approximation, see [32, 48].

3.2. Exponential GFs vs ordinary GFs. The different forms of the GFs of the sequence μ_n have several interesting features which we now briefly explore.

Instead of $\tilde{f}^\star(s)$, we start with considering the usual Laplace transform of $\tilde{f}(z)$

$$\mathscr{L}(s) = \int_0^\infty e^{-xs}\tilde{f}(x)\,\mathrm{d}x,$$

which, by (2.6), satisfies

$$\mathscr{L}(s) = \sum_{j\geqslant 0}\frac{q^{\binom{j+1}{2}}}{s^{j+1}(s+q^j)}.$$

By inverting this series, we obtain

$$\tilde{f}(z) = \sum_{j\geqslant 0} \frac{q^{\binom{j+1}{2}}}{j!} z^{j+1} \int_0^1 e^{-q^j uz}(1-u)^j \,\mathrm{d}u.$$

From this exact expression, we deduce not only the exact expression (2.3) but also the following one (by multiplying both sides by e^z and then expanding)

$$\mu_n = n \sum_{0\leqslant j<n} \binom{n-1}{j} q^{\binom{j+1}{2}} \sum_{0\leqslant \ell<n-j} \binom{n-1-j}{\ell} \frac{q^{j\ell}(1-q^j)^{n-1-j-\ell}}{j+\ell+1}, \tag{3.4}$$

where all terms are now positive; compare (2.3). But this expression and (2.3) are less useful for numerical purposes for large n.

On the other hand, the consideration of our $\tilde{f}^\star(s)$ bridges essentially exponential GF (EGF) and ordinary GF (OGF) of μ_n. Indeed,

$$\begin{aligned}\tilde{f}^\star(s) &= \frac{1}{s}\int_0^\infty e^{-x-x/s} \sum_{n\geqslant 0} \frac{\mu_n}{n!} x^n \,\mathrm{d}x \\ &= \frac{1}{1+s}\sum_{n\geqslant 0} \mu_n \left(\frac{s}{1+s}\right)^n,\end{aligned}$$

which is essentially the Euler transform of the OGF; see [23].

Our proofs given above rely strongly on the use of EGF, but the use of OGF works equally well for some of them. We consider the general recurrence

$$a_n = a_{n-1} + \sum_{0\leqslant j<n} \pi_{n,j} a_j + b_n \qquad (n \geqslant 1), \tag{3.5}$$

with a_0 given. Then the OGF $A(z) := \sum_{n\geqslant 1} a_n z^n$ satisfies

$$A(z) = zA(z) + \frac{z}{1-pz} A\left(\frac{qz}{1-pz}\right) + B(z),$$

where $B(z) := \sum_{n\geqslant 1} b_n z^n$. Thus $\bar{A}(z) := (1-z)A(z)$ satisfies

$$\bar{A}(z) = B(z) + \frac{z}{1-z}\bar{A}\left(\frac{qz}{1-pz}\right),$$

which after iteration gives

$$\bar{A}(z) = \sum_{j\geqslant 0} q^{j(j-1)/2} \left(\frac{z}{1-z}\right)^j B\left(\frac{q^j z}{1-(1-q^j)z}\right).$$

Thus

$$A(z) = \sum_{j\geqslant 0} \frac{q^{j(j-1)/2} z^j}{(1-z)^{j+1}} B\left(\frac{q^j z}{1-(1-q^j)z}\right). \tag{3.6}$$

Closed-form expressions can be derived from this; we omit the details here.

4. Variance of Y_n. We derive in this section the asymptotics of the variance Y_n (see (1.9)), which can be regarded as a very rough independent approximation to X_n. We use an elementary approach (no complex analysis being needed) here based on the recurrences of the central moments and suitable tools of "asymptotic transfer" for the underlying recurrence. The approach is, up to the technical development of the required asymptotic transfer tools, by now standard; see [34, 36]. The same analysis provided here is also applicable to higher central moments, which will be analyzed in the next section.

4.1. Recurrence. For the variance of Y_n, we start with the recurrence (1.9), which translates into the recurrence satisfied by the moment GF $M_n(y) := \mathbb{E}\left(e^{Y_n y}\right)$

$$M_n(y) = M_{n-1}(y) \sum_{0\leqslant j<n} \pi_{n,j} M_j(y) \qquad (n \geqslant 2),$$

with $M_0(y) = 1$ and $M_1(y) = e^y$, where $\pi_{n,j} := \binom{n-1}{j} q^j p^{n-1-j}$. This implies, with $\bar{M}_n(y) := e^{-\mu_n y} M_n(y) = \mathbb{E}\left(e^{(Y_n-\mu_n)y}\right)$, that

$$\bar{M}_n(y) = \bar{M}_{n-1}(y) \sum_{0\leqslant j<n} \pi_{n,j} \bar{M}_j(y) e^{\Delta_{n,j} y} \qquad (n \geqslant 2), \tag{4.1}$$

with $\bar{M}_n(y) = 1$ for $n < 2$, where

$$\Delta_{n,j} := \mu_j + \mu_{n-1} - \mu_n.$$

Let $M_{n,m} := \mathbb{E}(Y_n - \mu_n)^m = \bar{M}_n^{(m)}(0)$, $m \geqslant 0$. Then from (4.1), we deduce that

$$M_{n,m} = M_{n-1,m} + \sum_{0\leqslant j<n} \pi_{n,j} M_{j,m} + T_{n,m}, \tag{4.2}$$

where, for $m \geqslant 1$,

$$\begin{aligned} T_{n,m} &= \sum_{\substack{k+\ell+h=m\\ 0\leqslant k,\ell<m\\ 0\leqslant h\leqslant m}} \binom{m}{k,\ell,h} M_{n-1,k} \sum_{0\leqslant j<n} \pi_{n,j} M_{j,\ell} \Delta_{n,j}^h \\ &= \sum_{0\leqslant \ell<m} \binom{m}{\ell} \sum_{0\leqslant j<n} \pi_{n,j} M_{j,\ell} \Delta_{n,j}^{m-\ell} \\ &\quad + \sum_{2\leqslant k\leqslant m-2} \binom{m}{k} M_{n-1,k} \sum_{0\leqslant \ell\leqslant m-k} \binom{m-k}{\ell} \sum_{0\leqslant j<n} \pi_{n,j} M_{j,\ell} \Delta_{n,j}^{m-k-\ell}. \end{aligned} \tag{4.3}$$

Note that since $M_{n,1} = 0$ and $\sum_{0\leqslant j<n} \pi_{n,j} \Delta_{n,j} = 0$, the terms with $k = 1$ and $k = m-1$ vanish.

In particular, the variance $\sigma_n^2 = M_{n,2}$ satisfies

$$\sigma_n^2 = \sigma_{n-1}^2 + \sum_{0\leqslant j<n} \pi_{n,j} \sigma_j^2 + T_{n,2},$$

where

$$T_{n,2} = \sum_{0\leqslant j<n} \pi_{n,j} \Delta_{n,j}^2.$$

This will be useful for our asymptotic analysis for σ_n^2.

4.2. Asymptotics of $T_{n,2}$. To proceed further, we first consider the asymptotics of $\Delta_{n,j}$ for $j = qn + O(n^{2/3})$. By Taylor expansion and (2.2), we have

$$\begin{aligned}\tilde f(n) - \tilde f(n-1) &= \tilde f'(n) - \frac{\tilde f''(n)}{2} + \frac{\tilde f'''(n)}{3!} + O\left(\int_0^1 (1-t)^4 \tilde f^{(4)}(n-t)\,\mathrm{d}t\right)\\ &= \tilde f'(n) - \frac{\tilde f''(n)}{2} + \frac{\tilde f'''(n)}{3!} + O\left(\tilde f\left(q^4 n\right)\right),\end{aligned}$$

and

$$\tilde f''(n) - \tilde f''(n-1) = \tilde f'''(n) + O\left(\tilde f\left(q^4 n\right)\right).$$

These and (2.25) yield

$$\begin{aligned}\mu_n - \mu_{n-1} &= \tilde f'(n) - \frac{\tilde f''(n)}{2} + O\left(n^2 \tilde f\left(q^4 n\right)\right)\\ &= \tilde f(qn) + O\left(n^2 \tilde f\left(q^4 n\right)\right),\end{aligned}$$

since $\tilde f\left(q^2 n\right) = O\left(n^2 (\log n)^{-2} \tilde f\left(q^4 n\right)\right)$. Then, for $j = qn + x\sqrt{pqn}$, $|x| \leqslant n^{1/6}$,

$$\begin{aligned}\Delta_{n,j} &= \mu_j - (\mu_n - \mu_{n-1})\\ &= \tilde f(qn + x\sqrt{pqn}) - \tilde f(qn) + O\left(n^2 \tilde f\left(q^4 n\right)\right)\\ &= \tilde f'(qn) x\sqrt{pqn} + O\left(n^2 (1+x^2) \tilde f\left(q^4 n\right)\right). \qquad (4.4)\end{aligned}$$

Thus, by (2.20) and (2.21),

$$\begin{aligned}T_{n,2} &= \sum_{|x|\leqslant n^{1/6}} \pi_{n,j} \left|\tilde f'(qn) x\sqrt{pqn} + O\big(n^2 \tilde f(q^4 n)\big)\right|^2 + O\left(\mu_n^2 \sum_{|x|>n^{1/6}} \pi_{n,j}\right)\\ &= pqn \tilde f'(qn)^2 \sum_{|x|\leqslant n^{1/6}} \pi_{n,j} |x|^2 + O\left(n^{9/2} \tilde f^2\left(q^4 n\right)\right)\\ &= pqn \tilde f'(qn)^2 + O\left(n^{9/2} \tilde f^2\left(q^4 n\right)\right)\\ &\sim q^{-1} p n^{-3} (\log_\kappa n)^4 \tilde f(n)^2. \qquad (4.5)\end{aligned}$$

The next step then is to "transfer" this estimate to the asymptotics of the variance.

4.3. Asymptotic transfer. We now develop an asymptotic transfer result, which will be used to compute the asymptotics of higher central moments of Y_n (in particular the variance).

More generally, we consider a sequence $\{a_n\}_{n\geqslant 0}$ satisfying the recurrence relation (3.5), where a_0 is finite (whose value is immaterial) and $\{b_n\}_{n\geqslant 1}$ is a given sequence.

LEMMA 4.1. *If $b_n \sim n^\beta (\log n)^\xi \tilde f(n)^\alpha$, where $\alpha > 0$, $\beta, \xi \in \mathbb{R}$. Then*

$$\sum_{j\leqslant n} b_j \sim \frac{n}{\alpha \log_\kappa n}\, b_n.$$

Proof. Define $\varphi(t) := t^{\beta}(\log t)^{\xi}\tilde{f}(t)^{\alpha}$. By assumption, $b_n \sim \varphi(n)$. Since $\tilde{f}'(t)/\tilde{f}(t) \sim t^{-1}\log_\kappa t$ (by (2.20)), we see that $\varphi'(t) > 0$ for t sufficiently large, say $t \geqslant t_0 > 0$. Thus $\varphi(t)$ is monotonically increasing for $t \geqslant t_0$. Then

$$\sum_{j\leqslant n} b_j \sim \sum_{2\leqslant j\leqslant n}\varphi(j) = \int_2^n \varphi(t)\,\mathrm{d}t + O(\varphi(n)).$$

By the asymptotic relation (2.20), we have

$$\begin{aligned}\int_1^n \varphi(t)\,\mathrm{d}t &= \int_1^n t^{\beta}(\log t)^{\xi}\tilde{f}(t)^{\alpha}\,\mathrm{d}t\\ &\sim (\log\kappa)\int_1^n t^{\beta+1}(\log t)^{\xi-1}\tilde{f}(t)^{\alpha-1}\tilde{f}'(t)\,\mathrm{d}t\\ &\sim \frac{\log\kappa}{\alpha}\int_1^n t^{\beta+1}(\log t)^{\xi-1}\,\mathrm{d}\tilde{f}(t)^{\alpha}\\ &= \frac{n\varphi(n)}{\alpha\log_\kappa n} + O\left(\int_1^n \frac{\varphi(t)}{t}\,\mathrm{d}t\right),\end{aligned}$$

by an integration by parts. The integral on the right-hand side is easily estimated as follows.

$$\begin{aligned}\int_1^n \frac{\varphi(t)}{t}\,\mathrm{d}t &= O\left(\varphi(qn)\int_1^{qn} t^{-1}\,\mathrm{d}t + \varphi(n)\int_{qn}^n t^{-1}\,\mathrm{d}t\right)\\ &= O(\varphi(n)).\end{aligned}$$

This proves the lemma. □

PROPOSITION 4.2. *If $b_n \sim n^{\beta}(\log n)^{\xi}\tilde{f}(n)^{\alpha}$, where $\alpha > 1$, $\beta, \xi \in \mathbb{R}$, then*

$$a_n = \left(1 + O\left(n^{1-\alpha}(\log n)^{\alpha-1}\right)\right)\sum_{0\leqslant j\leqslant n} b_j \sim \frac{n}{\alpha\log_\kappa n} b_n. \tag{4.6}$$

Proof. We start with obtaining upper and lower bounds for a_n. Since $b_n > 0$ for sufficiently large n, say $n \geqslant n_0$. We may, without loss of generality, assume that $b_n \geqslant 0$ for $n \geqslant n_0$ (for, otherwise, we consider $b_n' := b_n + \max_{j\leqslant n_0}|b_j|$ and then show the difference between the corresponding a_n' and a_n is of order $\tilde{f}(n)$). Then $a_n \geqslant 0$ and, by (3.5), we have the lower bound

$$a_n \geqslant a_{n-1} + b_n \geqslant \sum_{0\leqslant j\leqslant n} b_j.$$

Now consider the sequence

$$C_n := \frac{a_n}{\sum_{0\leqslant j\leqslant n} b_j} \geqslant 1 \qquad (n \geqslant 1),$$

and the increasing sequence

$$C_n^* := \max_{1\leqslant j\leqslant n}\{C_j\} \geqslant 1.$$

Then we have the upper bound

$$a_k \leqslant C_n^* \sum_{0\leqslant j\leqslant k} b_j,$$

for all $k \leqslant n$.

In view of the recurrence relation (3.5), we have

$$a_n \leqslant C_{n-1}^* \sum_{0\leqslant j\leqslant n} b_j + C_{n-1}^* \sum_{0\leqslant j<n} \pi_{n,j} \sum_{0\leqslant \ell\leqslant j} b_\ell .$$

By Lemma 4.1 and Corollary 2.6, we see that there exist an absolute constant $K > 0$ such that

$$\sum_{0\leqslant j<n} \pi_{n,j} \sum_{0\leqslant \ell\leqslant j} b_\ell \leqslant Kn^{-\alpha}(\log n)^\alpha \sum_{0\leqslant j\leqslant n} b_j = O\left(n^{1-\alpha}(\log n)^{\alpha-1} b_n\right). \tag{4.7}$$

It follows that

$$a_n \leqslant C_{n-1}^* \left(1 + Kn^{-\alpha}(\log n)^\alpha\right) \sum_{0\leqslant j\leqslant n} b_j .$$

By our definition of C_n, we then have

$$C_n \leqslant C_{n-1}^* \left(1 + Kn^{-\alpha}(\log n)^\alpha\right),$$

and

$$C_n^* = \max\{C_{n-1}^*, C_n\} \leqslant C_{n-1}^* \left(1 + Kn^{-\alpha}(\log n)^\alpha\right).$$

Consequently,

$$C_n^* \leqslant C_2^* \prod_{2\leqslant j\leqslant n} \left(1 + Kj^{-\alpha}(\log j)^\alpha\right).$$

Since the finite product on the right-hand side is convergent, we conclude that the sequence C_n^* is bounded, or more precisely,

$$C_n^* \leqslant C_2^* \prod_{j\geqslant 2} \left(1 + Kj^{-\alpha}(\log j)^\alpha\right).$$

Thus we obtain the upper bound

$$a_n \leqslant C \sum_{0\leqslant j\leqslant n} b_j ,$$

where $C > 0$ is an absolute constant depending only on p, α, β and ξ.

With this bound and defining $\tilde{a}_n := \sum_{0\leqslant j<n} \pi_{n,j} a_j$, we can rewrite the recurrence relation (3.5) as

$$\begin{aligned} a_n &= a_{n-1} + \tilde{a}_n + b_n \\ &= \sum_{0\leqslant j\leqslant n} b_j + \sum_{0\leqslant k\leqslant n} \tilde{a}_k . \end{aligned} \tag{4.8}$$

Now by the estimate (4.7), we see that

$$\begin{aligned} \sum_{0\leqslant j\leqslant n} \tilde{a}_j &= O\left(1 + \sum_{2\leqslant j\leqslant n} j^{1-\alpha}(\log j)^{\alpha-1} b_j\right) \\ &= O\left(1 + \varphi(qn) \sum_{2\leqslant j\leqslant qn} j^{1-\alpha}(\log j)^{\alpha-1} + n^{1-\alpha}(\log n)^{\alpha-1} \sum_{qn<j\leqslant n} b_j\right), \end{aligned}$$

where $\varphi(t) := t^\beta (\log t)^\xi \tilde{f}(t)^\alpha$. Observe that

$$\varphi(qn) \sim n^{-\alpha}(\log n)^\alpha b_n \sim n^{-\alpha-1}(\log n)^{\alpha+1} \sum_{0\leqslant j\leqslant n} b_j.$$

Thus

$$\sum_{0\leqslant j\leqslant n} \tilde{a}_j = O\left(n^{1-\alpha}(\log n)^{\alpha-1} \sum_{0\leqslant j\leqslant n} b_j\right).$$

The proof of the Proposition is complete by substituting this estimate into (4.8). □

Denote by $[z^n]A(z)$ for the coefficient of z^n in the Taylor expansion of $A(z)$. Then, in terms of ordinary GFs, the asymptotic transfer (4.6) can be stated alternatively as

$$[z^n]A(z) \sim [z^n]\frac{B(z)}{1-z},$$

(when b_n satisfies the assumption of Proposition 4.2), which means that the contribution from terms in the sum in (3.6) with $j \geqslant 1$ is asymptotically negligible. Roughly, since

$$b_{n,j} := [z^n]B\left(\frac{q^j z}{1-(1-q^j)z}\right) = n^{-1} \sum_{1\leqslant \ell\leqslant n} \binom{n}{\ell} q^{j\ell}(1-q^j)^{n-\ell}\ell b_\ell,$$

we see that $b_{n,j} = O(q^j b_{\lfloor q^j n\rfloor})$. We can then give an alternative proof of (4.6) by using (3.6).

By (4.5) and a direct application of Proposition 4.2, we obtain an asymptotic approximation to the variance.

THEOREM 4.3. *The variance of Y_n satisfies*

$$\sigma_n^2 \sim C_\sigma n^{-2}(\log_\kappa n)^3 \tilde{f}(n)^2, \tag{4.9}$$

where $C_\sigma := p/(2q)$. Thus we have

$$\frac{\mathbb{V}(Y_n)}{(\mathbb{E}(Y_n))^2} \sim C_\sigma n^{-2}(\log n)^3.$$

Asymptotics of $\mathbb{V}(X_n)$ remains open. Monte Carlo simulations (with n a few hundred) suggested that the ratio $\mathbb{V}(X_n)/\mathbb{V}(Y_n)$ grows concavely, so that one would expect an order of the form $n^\beta(\log n)^\xi$ for $\mathbb{V}(X_n)$ for some $0 < \beta < 1$. But due to the complexity of the problem, we could not run simulations of larger samples to draw more convincing conclusions.

5. Asymptotic normality. We prove in this section that Y_n is asymptotically normally distributed by the method of moments. Our approach is to start from the recurrence (4.2) for the central moments and the asymptotic estimate (4.9) and then to apply inductively the asymptotic transfer result (Proposition 4.2), similar to that used in our previous papers [34, 36].

THEOREM 5.1. *The distribution of Y_n is asymptotically normal, namely,*

$$\frac{Y_n - \mu_n}{\sigma_n} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\xrightarrow{d}$ denotes convergence in distribution. We will indeed prove convergence of all moments.

Proof. By standard moment convergence theorem, it suffices to show that

$$M_{n,m} = \mathbb{E}(Y_n - \mu_n)^m \begin{cases} \sim \dfrac{(m)!}{(m/2)!2^{m/2}}\, \sigma_n^m, & \text{if } m \text{ is even,} \\ = o(\sigma_n^m), & \text{if } m \text{ is odd,} \end{cases} \tag{5.1}$$

for $m \geqslant 0$.

The cases when $m \leqslant 2$ having been proved above, we assume $m \geqslant 3$. By induction hypothesis, we have

$$M_{n,k} = O\left(\sigma_n^k\right) = O\left(n^{-k} (\log n)^{3k/2} \tilde{f}^k(n)\right),$$

for $k < m$. Then, by (4.4),

$$\begin{aligned} \sum_{0 \leqslant j < n} \pi_{n,j} M_{j,\ell} \Delta_{n,j}^h &= O\left(M_{\lfloor qn \rfloor,\ell} n^{h/2} \tilde{f}(q^2 n)^h\right) \\ &= O\left(n^{-\ell} (\log n)^{3\ell/2} \tilde{f}(qn)^\ell n^{h/2} \tilde{f}(q^2 n)^h\right) \\ &= O\left(n^{-2\ell - 3h/2} (\log n)^{5\ell/2 + 2h} \tilde{f}(n)^{\ell + h}\right). \end{aligned}$$

It follows (see (4.3)) that, for $0 \leqslant \ell < m$,

$$\sum_{0 \leqslant j < n} \pi_{n,j} M_{j,\ell} \Delta_{n,j}^{m-\ell} = O\left(n^{-\ell/2 - 3m/2} (\log n)^{\ell/2 + 2m} \tilde{f}(n)^m\right);$$

and, for $2 \leqslant k \leqslant m-2$ and $0 \leqslant \ell \leqslant m-k$,

$$M_{n-1,k} \sum_{0 \leqslant j < n} \pi_{n,j} M_{j,\ell} \Delta_{n,j}^{m-k-\ell} = O\left(n^{-\ell/2 + k/2 - 3m/2} (\log n)^{\ell/2 - k/2 + 2m} \tilde{f}(n)^m\right).$$

Thus the main contribution to the asymptotics of $T_{n,m}$ will come from the terms in the second group of sums in (4.3) with $k = m-2$ and $\ell = 0$. More precisely

$$T_{n,m} = \binom{m}{2} M_{n-1,m-2} T_{n,2} + O\left(n^{-3/2 - m} (\log n)^{3(m+1)/2} \tilde{f}(n)^m\right).$$

Note that $T_{n,2} \sim 2n(\log_\kappa n)^{-1} \sigma_n^2$; see (4.5).

Thus if m is even, then, by (4.5) and induction hypothesis,

$$\begin{aligned} T_{n,m} &\sim \frac{2m!}{((m-2)/2)!2^{m/2}}\, n^{-1} (\log_\kappa n) \sigma_n^m \\ &\sim \frac{2m!}{((m-2)/2)!2^{m/2}}\, C_\sigma^{m/2} n^{-m-1} (\log_\kappa n)^{(3m/2+1)} \tilde{f}(n)^m. \end{aligned}$$

Applying the asymptotic transfer result (Proposition 4.2) with $\alpha = m$, we obtain

$$\begin{aligned} M_{n,m} &\sim \frac{m!}{(m/2)!2^{m/2}}\, C_\sigma^{m/2} n^{-m} (\log n)^{3m/2} \tilde{f}(n)^m \\ &\sim \frac{m!}{(m/2)!2^{m/2}}\, \sigma_n^m. \end{aligned}$$

In a similar manner, we can prove that if m is odd, then

$$M_{n,m} = o(\sigma_n^m).$$

This concludes the proof of (5.1) and the asymptotic normality of Y_n. □

6. The random variables Z_n. We briefly consider the random variables defined recursively in (1.10). The major interest is in understanding the robustness of the asymptotic normality when changing the underlying probability distribution from binomial to uniform.

THEOREM 6.1. *The mean value of Z_n satisfies*

$$\mathbb{E}(Z_n) = Cn^{-1/4}e^{2\sqrt{n}}\left(1 + \frac{9}{16\sqrt{n}} + \frac{11}{1536n} + O\left(n^{-3/2}\right)\right), \tag{6.1}$$

where

$$C := \frac{1}{2}\sqrt{\frac{e}{\pi}}\left(e^{-1} - \int_1^{\infty} \frac{e^{-v}}{v}\right) \mathrm{d}v \approx 0.06906\,46192\ldots$$

The limit law of the normalized random variables $Z_n/\mathbb{E}(Z_n)$ is not normal

$$\frac{Z_n}{\mathbb{E}(Z_n)} \xrightarrow{d} Z,$$

where the distribution of Z is uniquely characterized by its moment sequence and the GF $\zeta(y) := \sum_{m\geqslant 1} \mathbb{E}(Z^m)y^m/(m\cdot m!)$ satisfies the nonlinear differential equation

$$y^2\zeta''(y) + y\zeta'(y) - \zeta(y) = y\zeta(y)\zeta'(y), \tag{6.2}$$

with $\zeta(0) = \zeta'(0) = 1$.

Proof. (Sketch) The proof of the theorem is simpler and we sketch only the major steps.

Mean value.. First, $\nu_n := \mathbb{E}(Z_n)$ satisfies the recurrence

$$\nu_n = \nu_{n-1} + \frac{1}{n}\sum_{0\leqslant j<n} \nu_j \qquad (n \geqslant 2),$$

with $\nu_0 = 0$, and $\nu_1 = 1$. The GF $f(z)$ of $\mathbb{E}(Z_n)$ satisfies the differential equation

$$f'(z) = \frac{2-z}{(1-z)^2}\,f(z) + \frac{1}{1-z},$$

with the initial condition $f(0) = 0$. The first-order differential equation is easily solved and we obtain the closed-form expression

$$f(z) = -\frac{1}{1-z} + \frac{e^{1/(1-z)}}{1-z}\left(e^{-1} - \int_1^{\infty} \frac{e^{-v} - e^{-v/(1-z)}}{v}\,\mathrm{d}v\right).$$

From this, the asymptotic approximation (6.1) results from a direct application of the saddle-point method (see Flajolet and Sedgewick's book [24, Ch. VIII]); see also [22].

Asymptotic transfer.. For higher moments and the limit law, we are led to consider the following recurrence.

$$a_n = a_{n-1} + \frac{1}{n}\sum_{0\leqslant j<n} a_j + b_n \qquad (n \geqslant 2), \tag{6.3}$$

with a_0 and a_1 given. For simplicity, we assume $a_0 = b_0 = 0$.

PROPOSITION 6.2. *Assume a_n satisfies (6.3). If $b_n \sim cn^{\beta}\nu_n^{\alpha}$, where $\alpha > 1$ and $\beta \in \mathbb{R}$, then*

$$a_n \sim \frac{c}{\alpha - \alpha^{-1}}\,n^{\beta+1/2}\nu_n^{\alpha}. \tag{6.4}$$

The proof is similar to that for Proposition 4.2 and is omitted.

Recurrence and induction.. By Proposition 6.2 and the following recurrence relation for the moment GF $Q(y) := \mathbb{E}(e^{Z_n y})$

$$Q_n(y) = \frac{Q_{n-1}(y)}{n} \sum_{0 \leqslant j < n} Q_j(y) \qquad (n \geqslant 2),$$

with $Q_0(y) = 1$ and $Q_1(y) = e^y$, we deduce, by induction using (6.4), that

$$\mathbb{E}(Z_n^m) \sim \zeta_m \nu_n^m \qquad (m \geqslant 1),$$

where

$$\zeta_m = \frac{1}{m - m^{-1}} \sum_{1 \leqslant j < m} \binom{m}{j} \frac{\zeta_j}{j} \zeta_{m-j} \qquad (m \geqslant 2), \tag{6.5}$$

with $\zeta_0 = \zeta_1 = 1$. It follows that the function $\zeta(y) := \sum_{m \geqslant 1} \zeta_m y^m/(m \cdot m!)$ satisfies the differential equation (6.2).

Unique determination of the distribution.. First, by a simple induction we can show, by (6.5), that $\zeta_m \leqslant cm!K^m$ for a sufficiently large $K > 0$. This is enough for justifying the unique determination. Instead of giving the details, it is more interesting to note that the nonlinear differential equation (6.2) represents another typical case for which the asymptotic behavior of its coefficients ($\mathbb{E}(Z^m)$ for large m) necessitates the use of the psi-series method recently developed in [10]. We can show, by the approach used there, that

$$\mathbb{E}(Z^m) = m \cdot m! \rho^{-m} \left(2 + \frac{2}{3m^2} + O\left(m^{-3}\right)\right),$$

where $\rho > 0$ is an effectively computable constant. Note that there is no term of the form m^{-1} in the expansion, a typical situation when psi-series method applies; see [10]. □

Concluding remarks. The approach we used in this paper is of some generality and is amenable to other quantities. We conclude this paper with a few examples and a list of some concrete applications where the scale $n^{c \log n}$ also appears.

First, the expected number of independent sets in a random graph (under the $\mathscr{G}_{n,p}$ model), as given in (1.8), satisfies the recurrence ($\bar{J}_n := J_n + 1$)

$$\bar{J}_n = \bar{J}_{n-1} + \sum_{0 \leqslant k < n} \binom{n-1}{k} q^k p^{n-1-k} \bar{J}_k \qquad (n \geqslant 1),$$

with $\bar{J}_0 = 1$. Thus the Poisson GF $\tilde{f}(z) := e^{-z} \sum_{n \geqslant 0} \bar{J}_n z^n/n!$ satisfies the equation

$$\tilde{f}'(z) = \tilde{f}(qz),$$

with $\tilde{f}(0) = 1$. The modified Laplace transform then satisfies the functional equation

$$\tilde{f}^\star(s) = 1 + s\tilde{f}^\star(qs),$$

which, by iteration, leads to the closed-form expression

$$\tilde{f}^\star(s) = \sum_{j \geqslant 0} q^{j(j-1)/2} s^j.$$

Thus all analysis as in Section 2 applies with F and G there replaced by

$$F(s) := \sum_{j\in\mathbb{Z}} q^{j(j-1)/2} s^j, \quad G(u) := q^{(\{u\}^2+\{u\})/2} F\left(q^{-\{u\}}\right).$$

We obtain for example

$$J_n = \frac{G\left(\log_\kappa \frac{n}{\log_\kappa n}\right)}{\sqrt{2\pi}} \cdot \frac{n^{1/\log\kappa + 1/2}}{\log_\kappa n} \exp\left(\frac{\left(\log \frac{n}{\log_\kappa n}\right)^2}{2\log\kappa}\right)\left(1 + O\left(\frac{(\log\log n)^2}{\log n}\right)\right).$$

The same approach also applies to the pantograph equation

$$\Phi'(z) = a\Phi(qz) + \Psi(z) \qquad (a > 0),$$

with $\Phi(0)$ and $\Psi(z)$ given, for $\Psi(z)$ satisfying properties that can be easily imposed.

Other extensions will be discussed elsewhere. We conclude with some other algorithmic, combinatorial and analytic contexts where $n^{c\log n}$ appears.

- Algorithmics: isomorphism testing (see [4, 29, 33, 50, 59]), autocorrelations of strings (see [30, 58]), information theory (see [2]), random digital search trees (see [19]), population recovery (see [66]), and asymptotics of recurrences (see [44, 52]);
- Combinatorics: partitions into powers (see [15, 48]; see also [26] for a brief historical account and more references), palindromic compositions (see [40]), combinatorial number theory (see [7, 46]), and universal tree of minimum complexity (see [11, 28]);
- Probability: log-normal distribution (see [41]), renewal theory (see [64, 65]), and total positivity (see [42]);
- Algebra: commutative ring theory (see [8]), and semigroups (see [45, 56, 60]);
- Analysis: pantograph equations (see [37, 43]), eigenfunctions of operators (see [61]), geometric partial differential equations (see [17]), and q-difference equations (see [3, 9, 18, 55, 69, 70]).

This list shows to some extent the generality of the seemingly uncommon scale $n^{c\log n}$; also it suggests the possibly nontrivial connections between instances in various areas, whose clarification in turn may lead to further development of more useful tools such as those in this paper.

REFERENCES

[1] James Abello, Sergiy Butenko, Panos M. Pardalos, and Mauricio G. C. Resende, *Finding independent sets in a graph using continuous multivariable polynomial formulations*, J. Global Optim., 21 (2001), pp. 111–137.

[2] Yaser S. Abu-Mostafa, *The complexity of information extraction*, IEEE Trans. Inform. Theory, 32 (1986), pp. 513–525.

[3] C. R. Adams, *Linear q-difference equations*, Bull. Amer. Math. Soc., 37 (1931), pp. 361–400.

[4] László Babai and Youming Qiao, *Polynomial-time isomorphism test for groups with abelian Sylow towers abelian Sylow towers*, in STACS, Christoph Dürr and Thomas Wilke, eds., vol. 14 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012, pp. 453–464.

[5] Béla Bollobás, *Random Graphs*, vol. 73, Cambridge University Press, Cambridge, second ed., 2001.

[6] Immanuel M. Bomze, Marco Budinich, Panos M. Pardalos, and Marcello Pelillo, *The maximum clique problem*, in Handbook of Combinatorial Optimization, Supplement Vol. A, Kluwer Acad. Publ., Dordrecht, 1999, pp. 1–74.

[7] P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties*, in Number Theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79.

[8] H. E. A. CAMPBELL, J. C. HARRIS, AND D. L. WEHLAU, *On rings of invariants of non-modular abelian groups*, Bull. Austral. Math. Soc., 60 (1999), pp. 509–520.
[9] R. D. CARMICHAEL, *The General Theory of Linear q-Difference Equations*, Amer. J. Math., 34 (1912), pp. 147–168.
[10] H.-H. CHERN, M.-I. FERNÁNDEZ-CAMACHO, H.-K. HWANG, AND C. MARTÍNEZ, *Psi-series method for equality of random trees and quadratic convolution recurrences*, Random Structures and Algorithms, to appear (2012).
[11] F. R. K. CHUNG, R. L. GRAHAM, AND D. COPPERSMITH, *On trees containing all small trees*, in The Theory and Applications of Graphs (Kalamazoo, Mich., 1980), Wiley, New York, 1981, pp. 265–272.
[12] V. CHVÁTAL, *Determining the stability number of a graph*, SIAM J. Comput., 6 (1977), pp. 643–662.
[13] AMIN COJA-OGHLAN AND CHARILAOS EFTHYMIOU, *On independent sets in random graphs*, in Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, PA, 2011, SIAM, pp. 136–144.
[14] R. M. CORLESS, G. H. GONNET, D. E. G. HARE, D. J. JEFFREY, AND D. E. KNUTH, *On the Lambert W function*, Adv. Comput. Math., 5 (1996), pp. 329–359.
[15] N. G. DE BRUIJN, *On Mahler's partition problem*, Nederl. Akad. Wetensch., Proc., 51 (1948), pp. 659–669 = Indagationes Math. 10, 210–220 (1948).
[16] ———, *The difference-differential equation $F'(x) = e^{\alpha x+\beta} F(x-1)$. I, II*, Nederl. Akad. Wetensch. Proc. Ser. A. **56** = Indagationes Math., 15 (1953), pp. 449–458, 459–464.
[17] FRANCESCA DE MARCHIS, *Generic multiplicity for a scalar field equation on compact surfaces*, J. Funct. Anal., 259 (2010), pp. 2165–2192.
[18] L. DI VIZIO, J.-P. RAMIS, J. SAULOY, AND C. ZHANG, *Équations aux q-différences*, Gaz. Math., (2003), pp. 20–49.
[19] MICHAEL DRMOTA, *Random Trees. An Interplay Between Combinatorics and Probability*, SpringerWien-NewYork, Vienna, 2009.
[20] PHILIPPE DUMAS AND PHILIPPE FLAJOLET, *Asymptotique des récurrences mahlériennes: le cas cyclotomique*, J. Théor. Nombres Bordeaux, 8 (1996), pp. 1–30.
[21] PETER C. FISHBURN, HELEN M. MARCUS-ROBERTS, AND FRED S. ROBERTS, *Unique finite difference measurement*, SIAM J. Discrete Math., 1 (1988), pp. 334–354.
[22] PETER C. FISHBURN, ANDREW M. ODLYZKO, AND FRED S. ROBERTS, *Two-sided generalized Fibonacci sequences*, Fibonacci Quart., 27 (1989), pp. 352–361.
[23] PHILIPPE FLAJOLET AND BRUCE RICHMOND, *Generalized digital trees and their difference-differential equations*, Random Structures Algorithms, 3 (1992), pp. 305–320.
[24] PHILIPPE FLAJOLET AND ROBERT SEDGEWICK, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.
[25] F.V. FOMIN AND D. KRATSCH, *Exact Exponential Algorithms*, Springer, 2010.
[26] MICHAEL L. FREDMAN AND DONALD E. KNUTH, *Recurrence relations based on minimization*, J. Math. Anal. Appl., 48 (1974), pp. 534–559.
[27] MICHAEL R. GAREY AND DAVID S. JOHNSON, *Computers and Intractability. A Guide to the Theory of NP-Completeness.*, W. H. Freeman and Co., San Francisco, Calif., 1979.
[28] M. K. GOL'DBERG AND È. M. LIVŠIC, *Minimal universal trees*, Mat. Zametki, 4 (1968), pp. 371–379.
[29] OTOKAR GROŠEK AND MAREK SÝS, *Isotopy of Latin squares in cryptography*, Tatra Mt. Math. Publ., 45 (2010), pp. 27–36.
[30] LEO J. GUIBAS AND ANDREW M. ODLYZKO, *Periods in strings*, J. Combin. Theory Ser. A, 30 (1981), pp. 19–42.
[31] JOCHEN HARANT, *Some news about the independence number of a graph*, Discuss. Math. Graph Theory, 20 (2000), pp. 71–79.
[32] SIDNEY HELLER, *An asymptotic solution of the difference equation $a_{n+1} - a_n = a_{[n/2]}$*, J. Math. Anal. Appl., 34 (1971), pp. 464–469.
[33] MICHAEL HUBER, *Computational complexity of reconstruction and isomorphism testing for designs and line graphs*, J. Combin. Theory Ser. A, 118 (2011), pp. 341–349.
[34] HSIEN-KUEI HWANG, *Second phase changes in random m-ary search trees and generalized quicksort: convergence rates*, Ann. Probab., 31 (2003), pp. 609–629.
[35] HSIEN-KUEI HWANG, MICHAEL FUCHS, AND VYTAS ZACHAROVAS, *Asymptotic variance of random symmetric digital search trees*, Discrete Math. Theor. Comput. Sci., 12 (2010), pp. 103–165.
[36] HSIEN-KUEI HWANG AND RALPH NEININGER, *Phase change of limit laws in the quicksort recurrence under varying toll functions*, SIAM J. Comput., 31 (2002), pp. 1687–1722.
[37] A. ISERLES, *On the generalized pantograph functional-differential equation*, European J. Appl. Math., 4 (1993), pp. 1–38.
[38] PHILIPPE JACQUET AND WOJCIECH SZPANKOWSKI, *Analytical de-Poissonization and its applications*, Theoret. Comput. Sci., 201 (1998), pp. 1–62.
[39] MARK JERRUM, *Large cliques elude the Metropolis process*, Random Structures Algorithms, 3 (1992),

pp. 347–359.
[40] Kathy Q. Ji and Herbert S. Wilf, *Extreme palindromes*, Amer. Math. Monthly, 115 (2008), pp. 447–451.
[41] Norman L. Johnson, Samuel Kotz, and N. Balakrishnan, *Continuous Univariate Distributions. Vol. 1*, John Wiley & Sons Inc., New York, second ed., 1994.
[42] Samuel Karlin and Zvi Ziegler, *Some inequalities of total positivity in pure and applied mathematics*, in Total Positivity and Its Applications (Jaca, 1994), vol. 359 of Math. Appl., Kluwer Acad. Publ., Dordrecht, 1996, pp. 247–261.
[43] Tosio Kato and J. B. McLeod, *The functional-differential equation* $y'(x) = ay(\lambda x) + by(x)$, Bull. Amer. Math. Soc., 77 (1971), pp. 891–937.
[44] Donald E. Knuth, *An almost linear recurrence*, Fibonacci Quart., 4 (1966), pp. 117–128.
[45] Yu. V. Kuzmin, *On the growth function of direct decompositions associated with homology of free abelianized extensions*, J. Pure Appl. Algebra, 86 (1993), pp. 223–229.
[46] Vsevolod F. Lev, Tomasz Łuczak, and Tomasz Schoen, *Sum-free sets in abelian groups*, Israel J. Math., 125 (2001), pp. 347–367.
[47] A. Yu. Levin and V. V. Vasilchikov, *On the average amount of work needed to find the maximal complete subgraph in a random graph (in Russian)*, Heuristic Algorithms of Optimization, (1981), pp. 61–64.
[48] Kurt Mahler, *On a special functional equation*, J. London Math. Soc., 15 (1940), pp. 115–123.
[49] David W. Matula, *On the complete subgraphs of a random graph*, in Proc. Second Chapel Hill Conf. on Combinatorial Mathematics and its Applications (Univ. North Carolina, Chapel Hill, N.C., 1970), Univ. North Carolina, Chapel Hill, N.C., 1970, pp. 356–369.
[50] Gary L. Miller, *On the* $n^{\log n}$ *isomorphism technique: A preliminary report*, in STOC, Richard J. Lipton, Walter A. Burkhard, Walter J. Savitch, Emily P. Friedman, and Alfred V. Aho, eds., ACM, 1978, pp. 51–58.
[51] J. R. Ockendon and A. B. Tayler, *The dynamics of a current collection system for an electric locomotive*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 322 (1971), pp. 447–468.
[52] Edwin O'Shea, M*-partitions: optimal partitions of weight for one scale pan*, Discrete Math., 289 (2004), pp. 81–93.
[53] W. B. Pennington, *On Mahler's partition problem*, Ann. of Math. (2), 57 (1953), pp. 531–546.
[54] B. Pittel, *On the probable behaviour of some algorithms for finding the stability number of a graph*, Math. Proc. Cambridge Philos. Soc., 92 (1982), pp. 511–526.
[55] Jean-Pierre Ramis, *About the growth of entire functions solutions of linear algebraic* q*-difference equations*, Ann. Fac. Sci. Toulouse Math. (6), 1 (1992), pp. 53–94.
[56] Illya I. Reznykov and Vitaliy I. Sushchansky, *On the 3-state Mealy automata over an* m*-symbol alphabet of growth order* $[n^{\log n/2\log m}]$, J. Algebra, 304 (2006), pp. 712–754.
[57] Bruce Richmond, *Mahler's partition problem*, Ars Combinatoria, 2 (1976), pp. 169–189.
[58] Eric Rivals and Sven Rahmann, *Combinatorics of periods in strings*, J. Combin. Theory Ser. A, 104 (2003), pp. 95–113.
[59] D. Rosenbaum, *Breaking the* $n^{\log n}$ *barrier for nilpotent-group isomorphism*, Arxiv Preprint arXiv:1205.0642, (2012).
[60] L.M. Shneerson, *Relatively free semigroups of intermediate growth*, Journal of Algebra, 235 (2001), pp. 484–546.
[61] V. Spiridonov, *Universal superpositions of coherent states and self-similar potentials*, Physical Review A, 52 (1995), p. 1909.
[62] Wojciech Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley-Interscience, New York, 2001.
[63] Robert Endre Tarjan and Anthony E. Trojanowski, *Finding a maximum independent set*, SIAM J. Comput., 6 (1977), pp. 537–546.
[64] Paul van Beek and Jac Braat, *The limits of sequences of iterated overshoot distribution functions*, Stochastic Processes Appl., 1 (1973), pp. 307–316.
[65] Y. Vardi, L. A. Shepp, and B. F. Logan, *Distribution functions invariant under residual-lifetime and length-biased sampling*, Z. Wahrsch. Verw. Gebiete, 56 (1981), pp. 415–426.
[66] A. Wigderson and A. Yehudayoff, *Population recovery and partial identification*. Preprint (available at www.math.ias.edu/avi/publications), 2012.
[67] Herbert S. Wilf, *Algorithms and Complexity*, A K Peters Ltd., Natick, MA, second ed., 2002.
[68] Gerhard J. Woeginger, *Exact algorithms for NP-hard problems: a survey*, in Combinatorial Optimization—Eureka, you shrink!, vol. 2570 of Lecture Notes in Comput. Sci., Springer, Berlin, 2003, pp. 185–207.
[69] Changgui Zhang, *Développements asymptotiques* q*-Gevrey et séries* Gq*-sommables*, Ann. Inst. Fourier (Grenoble), 49 (1999), pp. vi–vii, x, 227–261.
[70] C. Zhang, *Analytic continuation of solutions of the pantograph equation by means of* θ*-modular formula*, ArXiv e-prints, (2012).