Philippe Flajolet's Early Work in Combinatorics

Helmut Prodinger

*Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa, and
Department of Mathematics and Mathematical Statistics, Umea University, 907 36 Umea, Sweden
Email: hproding@sun.ac.za*

ABSTRACT: Some of Philippe Flajolet's* combinatorial contributions that he wrote between 1976 and 1995, say, are described. In most of Flajolet's papers, asymptotic/analytic considerations play a major role. To be true to the spirit of the journal ECA, the emphasis is on the *combinatorial* part.

Covered are the Register function of binary trees, approximate counting, digital search trees and the Rice formula, probabilistic counting, adding a new slice, Ramanujan's influence, finite differences and harmonic numbers, digits, and the Mellin-Perron formula. This (very personal) choice has been made to keep this paper within the usual length of the historic papers in ECA, and because the subjects are thought to be attractive for combinatorialists, and finally because the author has studied them very closely, being even a coauthor of some.

1. Introduction

Professor Philippe Flajolet

AofA 2006 in Alden Biezen, Belgium

Philippe Flajolet (1948–2011) was, among many other things, a very powerful combinatorialist. In 1998, on the occasion of his 50th birthday, a paper with the title *Philippe Flajolet's research in Combinatorics and Analysis of Algorithms* was published [30]. An attempt was then made to be as complete as possible and say at least of few words about all his works. Here, the approach is different: a few prominent themes will be discussed in *more detail*. They were all very influential and highly cited, and were written in Flajolet's younger years. The author had the privilege to follow the development of these papers when they were written, and they were all studied back in Vienna with interested students.

Flajolet, after a few years in Logic and Linguistics, moved around 1976 to the *Analysis of Algorithms*; if I am not mistaken, Jean Vuillemin brought the problem about the register-function (Horton-Strahler numbers) to him. After that, Flajolet wrote a series of very powerful papers, and they proved how strong he was as a problem solver. Typically, his papers have a combinatorial first part with an asymptotic analysis part to follow. In the spirit of this journal, we concentrate on the combinatorial part. In France, at the period, there was M. P. Schützenberger as a very influential figure, and, following him, it was not uncommon for young french researchers to deal with power series, words, trees, lattice paths, and so on. The use of power series was, however, more of a formal type, and even involved non-commuting variables.

Flajolet told me on several occasions that Comtet's book [3] was one of his early inspirations, and he liked the *example-driven* presentation. His asymptotic methods are not so easy to trace: certainly, some early contact with A. Odlyzko played a role, and what was then often nick-named *à la Odlyzko* later became *singularity*

---

analysis of generating functions [15]. One of his favorite methods was the Mellin transform. He learnt from Rainer Kemp [14] that the *Gamma function method* [28] is based on Mellin transforms. After that, he quickly read parts of Doetsch's book [6] on integral transforms. He also had a strong liking of Ramanujan and loved Hardy's book [25].

Later in life, Flajolet became more of a systematic theory-builder and concentrated mostly on (complex) analysis rather than on combinatorics. Much of his legacy can be found in his book (with Sedgewick) *Analytic Combinatorics* [23].

This historic paper tries to emphasize his earlier years and some of his combinatorial results. From a personal perspective, the years 1980–1995 were the ones where I learned most from him, and that is the period that will be described here in some way.

Philippe Flajolet has published much more material until 1995; a selection had to be made to keep this paper within the usual length of a historic paper in ECA. The selection is also a very personal choice; the author had published on the subjects himself and is a co-author of some.

Some important subjects that had to be left out are continued fractions [8], generation of combinatorial objects [24], a powerful approach to treat the Birthday paradox and related questions [10], several papers on combinatorial constructions, number-theoretic algorithms, Boltzmann samplers, urn models, etc. If this paper gets favorable reviews it would be worthwhile to write a sequel.

# 2. Register function (Horton-Strahler numbers)

Flajolet's research on the register function of binary trees as published in the final article [18] was his first work in the analysis of algorithms. This problem is nontrivial but manageable, and one can learn a lot while working on it. As in most of Flajolet's work, the first part is of a combinatorial nature, while the second deals more with the asymptotic evaluation.

The problem is related to binary trees (counted by Catalan numbers) and the parameter reg, which associates to each binary tree (which is used to code an arithmetic expression, with data in the leaves and operators in the internal nodes) the minimal number of extra registers that are needed to evaluate the tree. The optimal strategy is to evaluate difficult subtrees first and use one register to keep its value, which does not hurt if the other subtree requires fewer registers. If both subtrees are equally difficult, then one more register is used, compared to the requirements of the subtrees.

There is a recursive description of this function: $\mathsf{reg}(\square) = 0$, and if tree t has subtrees t_1 and t_2, then

$$\mathsf{reg}(t) = \begin{cases} \max\{\mathsf{reg}(t_1), \mathsf{reg}(t_2)\} & \text{if } \mathsf{reg}(t_1) \neq \mathsf{reg}(t_2), \\ 1 + \mathsf{reg}(t_1) & \text{otherwise.} \end{cases}$$

The register function is also known as Horton-Strahler numbers in the study of the complexity of river networks. The original papers are [26, 34]; since then many papers have been written on the subject. The recursive description attaches numbers to the nodes, starting with 0's at the leaves and then going up; the number appearing at the root is the register function of the tree.
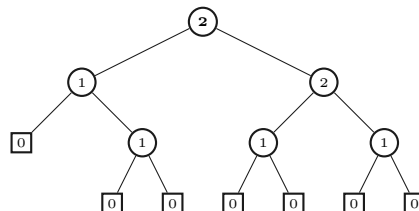
Let $\mathcal{R}_p$ denote the family of trees with register function $= p$, then one gets immediately from the recursive definition:

$$\mathcal{R}_p = \quad \mathcal{R}_{p-1} \quad \mathcal{R}_{p-1} \quad + \quad \mathcal{R}_p \quad \sum_{j<p} \mathcal{R}_j \quad + \quad \sum_{j<p} \mathcal{R}_j \quad \mathcal{R}_p$$

Later in life, Flajolet would write such *symbolic equations* in a linearized form. In terms of generating functions, these equations read as

$$R_p(z) = zR_{p-1}^2(z) + 2zR_p(z)\sum_{j<p} R_j(z) = \frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}} \quad \text{with} \quad z = \frac{u}{(1+u)^2};$$

the variable z is used to mark the size (i. e., the number of internal nodes) of the binary tree. The substitution into the u-world originated in an important paper by de Bruijn, Knuth, and Rice [4].

The explicit formula for $R_p(z)$ can be proved by induction. This was of course not the way it was found. Nowadays, with computer algebra, from a list of the first few values of $R_p(z)$, expressed in the variable u, one would quickly guess the formula.

Explicit forms for $[z^n]R_p(z)$, the number of binary trees of size n with register function $= p$, are available via contour integration (always around a small circle around the origin):

$$[z^n]R_p(z) = \frac{1}{2\pi i}\oint \frac{dz}{z^{n+1}}\frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}} = [u^n](1+u)^{2n+2}\frac{1-u}{(1+u)^3}\frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}}$$

$$= \sum_{k\geq 0}\left[\binom{2n}{n+1-(2k+1)2^p} - 2\binom{2n}{n-(2k+1)2^p} + \binom{2n}{n-1-(2k+1)2^p}\right].$$

From this, the average value of the register function of a random binary tree with n nodes can be computed:

$$[z^n]\sum_{p\geq 1}pR_p(z) = \sum_{k\geq 1}v_2(k)\left[\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k}\right] \tag{1}$$

with $v_2(k)$ being the number of trailing zeroes in the binary representation of k. This is an interesting number theoretic quantity described by the recursive rules $v_2(2k) = 1 + v_2(k)$, $v_2(2k+1) = 0$, $v_2(1) = 0$. Let $S_2(n)$ be the sum-of-digits function of n in the binary number system. Then $S_2(n) - S_2(n-1) = 1 - v_2(n)$, and by telescoping $S_2(n) = n - \sum_{k\leq n}v_2(k)$. A further summation leads to a result by Delange [5], published just before Flajolet's work on the register function was begun: $\sum_{m<n}S_2(m) = \frac{n}{2}\log_2 n + nF(\log_2 n)$, where $F(x)$ is a continuous periodic function of period 1 with explicitly known Fourier coefficients. Using Abel's partial summation twice on $[z^n]R_p(z)$, one gets eventually:

The average number of registers to evaluate a binary tree with n nodes is asymptotically given by

$$\log_4 n + D(\log_4 n) + o(1)$$

with

$$D(x) = \sum_{k\in\mathbb{Z}}d_k e^{2\pi ikx} \quad \text{and} \quad d_0 = \frac{1}{2} - \frac{\gamma}{2\log 2} - \frac{1}{\log 2} + \log_2\pi, \; d_k = \frac{1}{\log 2}\zeta(\chi_k)\Gamma(\tfrac{\chi_k}{2})(\chi_k - 1), \quad k\neq 0,$$

and $\chi_k = \frac{2\pi ik}{\log 2}$. The classical Gamma- and zeta-functions appear here.

At the same time, Rainer Kemp [27] proved equivalent results using different methods.

3. Digital search trees and Tries

Digital search trees are constructed from (sufficiently long) sequences of 0 and 1 (appearing with equal probability in the simplest model). The order in which the items are inserted is important. In the following example, the items are inserted in the order $A, B, C, \ldots$. The most important parameter is the *insertion depth*, i. e., the average distance of a node to the root. In [28], there was an open problem that Flajolet and Sedgewick [21] solved: What is the (average) number of internal endnodes if a random digital search tree, built of n data is considered. In the example, C, F, I, H are such endnodes, and the count is 4. We will describe the combinatorial part of the solution. The asymptotic part is based on Rice's method and is perhaps the method of choice when it comes to analyzing Digital search trees and their closely related cousins, called Tries.

$A : 1001$
$B : 0110$
$C : 0000$
$D : 1111$
$E : 0100$
$F : 0101$
$G : 1101$
$H : 1110$
$I \ : 1100$
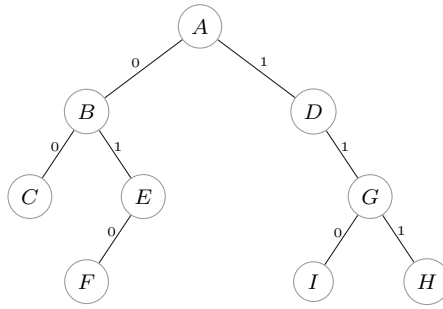
This is a *digital search tree* with 9 nodes. A *trie* is very similar, except that the data is located in the leaves, not the internal nodes. Parameters like *size* and *path-length* have interesting connections to q-analysis.

Let
$$F_n(z) = \sum_{k \geq 0}[\text{probability that a random digital search tree of size } n \text{ has } k \text{ endnodes}]z^k.$$

It is easy to see that

$$F_{n+1}(z) = \sum_{k=0}^{n} 2^{-n}\binom{n}{k}F_k(z)F_{n-k}(z), \quad n \geq 1, \quad F_0(z) = 1, \quad F_1(z) = z,$$

since $2^{-n}\binom{n}{k}$ is the probability that k items go to the left and $n-k$ go the right. The desired average is then obtained via $\ell_n = F'_n(1)$, following the recursion

$$\ell_{n+1} = 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}\ell_k, \quad n \geq 1, \quad \ell_0 = 0, \quad \ell_1 = 1.$$

This type of recursions is usually solved using exponential generating functions since it translates into a functional-differential equation:

$$L(z) = \sum_{n \geq 0} \ell_n \frac{z^n}{n!} \qquad \text{and} \qquad L'(z) = 1 + 2L\Big(\frac{z}{2}\Big)e^{z/2}.$$

This becomes better using the *Poisson transformed function*:

$$\widehat{L}(z) = \sum_{n \geq 0} \widehat{\ell}_n \frac{z^n}{n!} = e^{-z}L(z) \qquad \text{and} \qquad \widehat{L}'(z) = e^{-z} + 2\widehat{L}\Big(\frac{z}{2}\Big).$$

Reading off the coefficients of $z^n/n!$ on both sides, we are led to the recursion

$$\widehat{\ell}_{n+1} = (-1)^n - (1 - 2^{1-n})\widehat{\ell}_n, \quad n \geq 0, \quad \widehat{\ell}_0 = 0,$$

which can be solved by iteration. Setting

$$Q_n = \prod_{1 \leq \ell \leq n}\Big(1 - \frac{1}{2^k}\Big), \quad \text{then} \quad \widehat{\ell}_n = (-1)^{n+1}Q_{n-2}\Big(\frac{1}{Q_0} + \cdots + \frac{1}{Q_{n-2}}\Big), \quad n \geq 2, \quad \widehat{\ell}_0 = 0, \widehat{\ell}_1 = 1.$$

Since $L(z) = e^z\widehat{L}(z)$, comparing coefficients eventually leads to the formula

$$l_n = n - \sum_{k \geq 2}\binom{n}{k}(-1)^k R_{k-2}, \quad \text{with} \quad R_n = Q_n\Big(\frac{1}{Q_0} + \cdots + \frac{1}{Q_n}\Big), \quad \text{for } n \geq 2.$$

For the asymptotic evaluation of this with Rice's method, it is essential to have a definition of R_n, where n can also be a complex number; currently, the definition only makes sense for nonnegative integers.

Although not much will be done in this historic paper with Rice's method, at least the key formula should be mentioned:

$$\sum_{k=1}^{n}\binom{n}{k}(-1)^k f(k) = \frac{1}{2\pi i}\int_{\mathcal{C}}\frac{(-1)^{n-1}n!}{z(z-1)\dots(z-n)}dz,$$

where the positively oriented curve $\mathcal{C}$ encircles the poles $1, \dots, n$ and no others. The lower index 1 in the summation and contour could be replaced by any number; in our application, 2 is the number of choice. For

that, the sequence $f(k)$ must be extended to an analytic function $f(z)$. This is often easy, but sometimes not. What one usually does (if it is feasible) is to change the contour and to compute some extra residues.

It is interesting to observe that Q_n really belongs to q-analysis. Using the notation $(z;q)_n = (1-z)(1-qz)\dots(1-q^{n-1}z)$, one could write $Q_n = (\frac{1}{2}; \frac{1}{2})_n$. Using ideas from q-analysis, the finite product can be written as a quotient of two infinite products, and that leads already to the definition of Q_n for complex n:

$$Q_n = \frac{Q_\infty}{Q(2^{-n})}, \quad \text{with} \quad Q(x) := \prod_{k \geq 1}\left(1 - \frac{x}{2^k}\right) \quad \text{and} \quad Q_\infty = Q(1) = 0.288788\dots.$$

For R_n/Q_n, this is not so easy. A first idea is to write the finite sum as a difference of two infinite series; the problem is that $\frac{1}{Q_0} + \frac{1}{Q_1} + \cdots$ does not converge. In order to overcome this, one must pull out something so that the strategy works for the (tiny) rest. Skipping a few details, what works is

$$R_n = n + 1 - \alpha + R_n^*, \quad \text{with} \quad \alpha := \sum_{k \geq 1}\frac{1}{2^k - 1} \quad \text{and} \quad R^*(z) = \sum_{j \geq 0}\frac{(z+1+j-\alpha)2^{-z-1-j}}{(1-2^{-z-1})(1-2^{-z-2})\dots(1-2^{-z-1-j})},$$

then $R_n^* = R^*(n)$. In this series, z can be a complex number; there is however a nicer version:

$$R^*(z) = \frac{1}{Q_\infty} \sum_{j \geq 1}\frac{(-1)^{j-1}2^{-\binom{j}{2}}}{Q_{j-1}}\frac{z+j}{2^{z+j}-1},$$

which can be obtained by partial fraction decomposition and a partition identity of Euler

$$\sum_{n \geq 0}\frac{t^n}{(q;q)_n} = \frac{1}{(1-t)(1-qt)(1-q^2t)\dots} = \frac{1}{(t;q)_\infty}.$$

The special values $t = q = \frac{1}{2}$ evaluate the series

$$\sum_{\ell \geq 0}\frac{2^{-\ell}}{Q_\ell} = \frac{1}{Q_\infty},$$

and, after differentiation,

$$\sum_{\ell \geq 0}\frac{\ell 2^{-\ell}}{Q_\ell} = \frac{\alpha}{Q_\infty}.$$

The rest of the method is of a more analytic nature and shall not be discussed here. Eventually, it turns out that the average number of endnodes is (apart from small fluctuations) $\ell_n \approx n \cdot 0.372048\dots$; the constant can be analytically expressed as $\alpha + 1 - R^*(-1)$. Recall that in the term for $j = 1$ in $R^*(-1)$ a limit must be taken.

4. Approximate Counting

In [9] the following scenario is investigated. There is a counter C, initially set to 1. Assume that the counter's value is currently k and a random element arrives, then the counter advances to value $k + 1$ with probability 2^{-k}.

After n random increments, the counter has a value between 1 and n, but is typically around $\log_2 n$. So the counter serves as an approximate count, with significantly fewer bits.
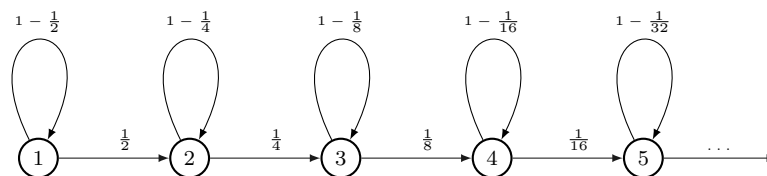
The graph describes this very well, and also allows to find the generating function ($q = \frac{1}{2}$)

$$H_\ell(z) = \sum_{n \geq 0} z^n[\text{probability to reach state } \ell \text{ after } n \text{ random steps}] = \frac{z^{\ell-1}q^{\binom{\ell}{2}}}{\prod_{i=1}^{\ell}\left(1 - (1-q^i)z\right)}$$

easily. Partial fraction decomposition and reading off coefficients leads to:

$$p_{n,\ell} := [z^n] H_\ell(z) = \sum_{i=0}^{\ell-1} \frac{(-1)^i 2^{-\binom{i}{2}}}{Q_i Q_{\ell-1-i}} \left(1 - 2^{-(\ell-i)}\right)^n \quad \text{with} \quad Q_i := \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{4}\right)\dots\left(1 - \frac{1}{2^i}\right);$$

the products Q_i appeared already in the section on digital search trees. Expanding the binomial in $p_{n,\ell}$ according to the binomial theorem, and using Euler's partition identities

$$\prod_{m \geq 0} \left(1 + xq^m\right) = \sum_{j \geq 0} \frac{x^j q^{\binom{j}{2}}}{(q;q)_j}, \qquad \prod_{m \geq 0} \left(1 - xq^m\right)^{-1} = \sum_{j \geq 0} \frac{x^j}{(q;q)_j}$$

one can find the average value C_n of the counter after n random increments to be

$$C_n = 1 - \sum_{k=1}^n (-1)^k \binom{n}{k} 2^{-k} Q_{k-1} = 1 - \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{(-1)^{n-1} n!}{z(z-1)\dots(z-n)} 2^{-z} Q_{z-1} dz.$$

The contour $\mathcal{C}$ encircles the poles $1, 2, \dots, n$ and no others. The appearance of such contour integrals is typical for Rice's method. Eventually, using methods from complex analysis one finds that $C_n \approx \log_2 n$. The study of the lower order (fluctuating) terms is very instructive, but shall not be presented here.

There are other ways to compute the probabilities $p_{n,k}$. In [29], this is done by iteration:

$$\text{set} \quad F(z,u) := \sum_{n \geq 0} \sum_{k \geq 1} p_{n,k} z^n u^k, \quad \text{then one derives } (q = \tfrac{1}{2} \text{ here})$$

$$F(z,u) = \frac{u}{1-z} + \frac{z(u-1)}{1-z} F(z,qu) = \sum_{j \geq 0} \frac{(-1)^j z^j (q;u)_j u q^j}{(1-z)^{j+1}}.$$

5. Probabilistic Counting

Assume that n data are given and, via a *hash-function*, each data is mapped to a non-negative integer, following the geometric distribution with parameter $q = \frac{1}{2}$. We might think about n persons, and to each person, a ball is associated, with a number printed on it, and the probabilities are $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$. Then each person throws her ball into an urn with the corresponding number. We expect that about $\frac{n}{2}$ balls fall into urn 0, $\frac{n}{4}$ balls fall into urn 1, and so on. The goal of the paper [13] is to use these urns to produce a count of the number of data (persons) without actually keeping track of them. Since we only distinguish which urn is empty or which urn is non-empty, it does not matter whether a person appears at the counter more than once, since the number associated with her is always the same.

The parameter R_n that is using as a count is the index of the first empty urn, or the longest sequence of non-empty urns (starting with urn 0). Flajolet computed $q_{n,k}$, the probability that $R_n \geq k$, using an inclusion-exclusion argument. This is somewhat surprising since he told me that he does not like inclusion/exclusion too much unless it corresponds to generating functions where a factor $(1-z)^k$ generates the alternating signs automatically.

So we need to work out the probability that urns $0, 1, \dots, k-1$ are non-empty. This is given by 1 minus the probability that a particular urn is empty, plus the probability that two particular urns are empty, and so on. We get

$$q_{n,k} = 1 - \sum_{0 \leq i < k} \left(1 - \frac{1}{2^i}\right)^n + \sum_{0 \leq i < j < k} \left(1 - \frac{1}{2^i} - \frac{1}{2^j}\right)^n - \sum_{0 \leq i < j < \ell < k} \left(1 - \frac{1}{2^i} - \frac{1}{2^j} - \frac{1}{2^\ell}\right)^n + \cdots .$$

Using different indices, we can rewrite this:

$$q_{n,k} = \sum_{t=0}^{k-1} (-1)^t \sum_{0 \leq i_0 < \cdots < i_{t-1} < k} \left(1 - \frac{2^{i_0} + \cdots + 2^{i_{t-1}}}{2^k}\right)^n.$$

We note that $2^{i_0} + \cdots + 2^{i_{t-1}}$ is just a (unique) non-negative integer, written in binary, and the sign depends on the number of bits in the binary representation. Denote by $\nu(k)$ the number of digits 1 in the binary representation of k (it was denoted by $S_2(k)$ in a previous section), then we have the beautiful formula

$$q_{n,k} = \sum_{j=0}^{2^k} (-1)^{\nu(j)} \left(1 - \frac{j}{2^k}\right)^n.$$

The sequence $(-1)^{\nu(j)}$ is known as the Prouhet-Thue-Morse-sequence. The pattern of signs goes like $+ - - +$ $- + + - - + + - + - - + \ldots$. The average value of the parameter R_n is exactly given by $\overline{R_n} = \sum_{k \geq 1} q_{n,k}$. Using methods from real analysis that will not be repeated here, Flajolet approximates:

$$q_{n,k} \approx \psi\left(\frac{n}{k}\right) \quad \text{with} \quad \psi(x) = \prod_{j \geq 0} (1 - e^{-x2^j}) = \sum_{j \geq 0} (-1)^{\nu(j)} e^{-jx}.$$

The last expansion is classical and easier to read when setting $q := e^{-x}$. Continuing his Mellin transform analysis, Flajolet is lead to study an interesting function, which resembles the Riemann ζ-function:

$$N(s) = \sum_{j \geq 1} \frac{(-1)^{\nu(j)}}{j^s};$$

$N(s)$ should be read 'nu-of-s', but there is no special character available. One major question is the analytic continuation of $N(s)$; as it stands, it only converges for $\Re(s) > 1$. The trick is to group four consecutive terms together, according to the pattern $+ - - +$; then there is enough cancellation so that one can go to the left with the variable s:

$$N(s) = -1 - 2^{-s} + 3^{-s} + \sum_{j \geq 0} \frac{(-1)^{\nu(j)}}{(4j)^s} \left[1 - \left(1 + \frac{1}{4j}\right)^{-s} - \left(1 + \frac{2}{4j}\right)^{-s} + \left(1 + \frac{3}{4j}\right)^{-s}\right].$$

Eventually one gets that the average value of the parameter R_n is asymptotically given by

$$\overline{R_n} \approx \log_2(\varphi \cdot n) + \text{tiny fluctuation}, \quad \text{and the constant } \varphi \text{ involves} \prod_{p \geq 1} \left[\frac{(4p+1)(4p+2)}{(4p)(4p+3)}\right]^{(-1)^{\nu(p)}}.$$

Such products are not uncommon in the literature [1], but not too much is known about the Flajolet-Martin constant [7].

6. Adding a new slice

In [16], the sequence $1, 1, 2, 3, 5, 9, 16, 28, \ldots$ (A002572 in [33]) is investigated used the adding-a-new-slice technique. This technique was probably known to some people under no name or a different name, but Flajolet's nickname for it is very descriptive. I vaguely remember that even Cayley was interested in this sequence, but I lost the reference.

The definition of the *level number sequences* is

C1. $n_1 = 1$,   C2. For all j such that $1 < j \leq k$: $n_j \leq 2n_{j-1}$,   C3. $n = n_1 + \cdots + n_k$.

The number k is called the height and the number of representations of n created in this way is denoted by H_n. A generating function $F(q, u)$ is formed where q keeps track of the current total value and u of the last entry. One writes $F(q, u) = F_1(q, u) + F_2(q, u) + \cdots$, and the index refers to the height of the sequence. The process to go from F_k to F_{k+1} is called adding a new slice. It is driven by the substitution

$$u^j \longrightarrow uq + (uq)^2 + \cdots + (uq)^{2j} = \frac{uq}{1 - uq}\left(1 - (uq)^{2j}\right), \quad \text{or} \quad F_{k+1}(q, u) = \frac{uq}{1 - uq}\left(F_k(q, 1) - F_k(q, u^2q^2)\right).$$

Summing on k leads to

$$F(q, u) = uq + \frac{uq}{1 - uq}\left(F(q, 1) - F(q, u^2q^2)\right),$$

which can be solved by iteration:

$$F(q, u) = uq + \frac{uq}{1 - uq} F(q, 1) - \frac{uq}{1 - uq} \left[(uq)^2 + \frac{(uq)^2}{1 - (uq)^2} F(q, 1) - \frac{(uq)^2}{1 - (uq)^4} \left[(uq)^4 + \frac{(uq)^4}{1 - (uq)^4} F(q, 1) + \cdots \right. \right.$$

Plugging in $u = 1$ and collecting leads to the explicit representation

$$F(q, 1) = \frac{\sum_{j \geq 1} (-1)^{j+1} \frac{q^{2^{j+1} - j - 2}}{(1-q)(1-q^3)(1-q^7)\dots(1-q^{2^j - 1} - 1)}}{1 - \sum_{j \geq 1} (-1)^{j+1} \frac{q^{2^{j+1} - j - 2}}{(1-q)(1-q^3)(1-q^7)\dots(1-q^{2^j} - 1)}}.$$

Recall that $F(q, 1)$ is the generating function of the numbers H_n, since setting $u = 1$ means that the height k is no longer of relevance. Asymptotics is not difficult from here (but not discussed in detail), since it can be shown that there is a dominant simple pole at 0.5573678719. Eventually, this leads to

$$H_n = 0.25450 \cdot 1.794147^n + O(1.43^n);$$

the exponential growth constant is the reciprocal of the simple pole. The author had a chance to use the adding-a-new-slice technique several times after the event of the level numbers paper.

7. Harmonic numbers and Euler sums

It is known that the quantity $S_n(m) = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k^m}$ for a negative integer m is basically a subset Stirling number ('of the second kind'). Somebody asked Flajolet what happens for *positive* integers m? Here is the answer, expressed in terms of *harmonic numbers* $H_n^{(j)} = \sum_{1 \leq k \leq n} \frac{1}{k^j}$: We start from a (cleverly chosen) quantity T and use partial fraction decomposition:

$$T = \frac{n!}{z(z-1)(z-2)\dots(z-n)} \frac{1}{z^m} = \sum_{k=1}^{n} \binom{n}{k} (-1)^{n-k} \frac{1}{k^m} \frac{1}{z-k} + \frac{\lambda}{z^{m+1}} + \cdots + \frac{\mu}{z}.$$

Now we consider $z \cdot T$ and let $z \to \infty$ with the result

$$\sum_{k=1}^{n} \binom{n}{k} (-1)^{n-k} \frac{1}{k^m} + \mu = 0; \quad \mu = [z^{-1}]T = [z^m] \frac{n!}{(z-1)\dots(z-n)} = (-1)^n [z^m] \frac{1}{(1-z)(1 - \frac{z}{2})\dots(1 - \frac{z}{n})}.$$

The computation continues:

$$\mu = -\sum_{k=1}^{n} \binom{n}{k} (-1)^{n-k} \frac{1}{k^m} = (-1)^n [z^m] \exp \log \frac{1}{(1-z)(1 - \frac{z}{2})\dots(1 - \frac{z}{n})} = (-1)^n [z^m] \exp \left\{ \sum_{k=1}^{n} \sum_{j \geq 1} \frac{1}{jk^j} \right\},$$

or

$$\sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} \frac{1}{k^m} = [z^m] \exp \left\{ \sum_{j \geq 1} \frac{H_n^{(j)}}{j} \right\}$$

$$= [z^m] \left(1 + H_n z + \left(\frac{H_n^2}{2} + \frac{H_n^{(2)}}{2} \right) z^2 + \left(\frac{H_n^3}{6} + \frac{H_n H_n^{(2)}}{2} + \frac{H_n^{(3)}}{3} \right) z^3 + \cdots \right),$$

which provides the first three evaluations:

$$H_n, \quad \frac{H_n^2}{2} + \frac{H_n^{(2)}}{2}, \quad \frac{H_n^3}{6} + \frac{H_n H_n^{(2)}}{2} + \frac{H_n^{(3)}}{3};$$

a general formula can be written using (partial) Bell polynomials. The paper [22] contains more such alternating summations, where rational functions no longer work, and instead of exact evaluations one gets asymptotic evaluations using contour integrations and Rice's integrals.

Harmonic numbers appear again, for instance in [20] when studying *Euler sums*

$$S_{p,q} = \sum_{n=1}^{\infty} \frac{H_n^{(p)}}{n^q}, \quad \text{like} \quad S_{1,2} = 2\zeta(3), \ S_{1,3} = \frac{5}{4}\zeta(4), \ S_{2,4} = \zeta(3)^2 - \frac{1}{3}\zeta(6), \ \&c.$$

The technique of the highly cited paper [20] is *residue calculus*, writing the sum in question in two different ways as a sum of residues using a suitable *kernel* function.

8. Flajolet and Ramanujan

A famous unproven assertion by Ramanujan is:

$$\frac{1}{2}e^n = 1 + \frac{n}{1!} + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\theta \quad \text{with} \quad \theta = \frac{1}{3} + \frac{4}{135(n+k)}, \quad \frac{8}{45} < k < \frac{2}{21}.$$

This was proved in [12]: There are auxiliary functions

$$Q(n) = 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots \quad \text{and} \quad R(n) = 1 + \frac{n}{n+1} + \frac{n^2}{(n+1)(n+2)} + \cdots;$$

it is not difficult to prove that $Q(n) + R(n) = n!e^n/n^n$; $Q(n)$ is called Ramanujan's Q-function. It turns out that Ramanujan's assertion can be rephrased as

$$R(n) - Q(n) = \frac{2}{3} + \frac{8}{135(n+k)}, \quad \text{with} \quad \frac{8}{45} < k < \frac{2}{21}.$$

Instead of real analysis which was used in earlier attempts, complex analysis is used, namely a famous implicit function defined by $y = ze^y$; it either goes by the name Lambert's W function or tree function. The generating function has the explicit expansion

$$y = \sum_{n \geq 1} n^{n-1}\frac{z^n}{n!}; \quad \text{recall that } n^{n-1} \text{ is the number of rooted labelled trees with } n \text{ nodes.}$$

One finds a generating function

$$\sum_{n \geq 1} Q(n)n^{n-1}\frac{z^n}{n!} \quad \text{and from it the asymptotic expansion} \quad Q(n) = \sqrt{\frac{\pi n}{2}} - \frac{1}{3} + \cdots.$$

However, for Ramanujan's assertion, *explicit* error bounds are requested. This is done in the above-mentioned paper using contour integration and careful bounds and computer algebra.

Flajolet was also asked about his personal hit-parade of Ramanujan's results [2]. I am not sure what he answered, but I know for sure that he was very fond of *Ramanujan's master theorem* [31], which can be stated as

$$\int_0^\infty x^{s-1}\big(\lambda(0) - x\lambda(1) + x^2\lambda(2) - \cdots\big)dx = \frac{\pi}{\sin(\pi s)}\lambda(-s) \quad \text{for} \quad 0 < \Re(s) < 1,$$

under some technical conditions on λ. This, among other things, has inspired Flajolet to come up with his Poisson-Newton-Mellin-Rice cycle [19].

9. Digits

The existence of equation (1) motivated Flajolet (later joined by coauthor Ramshaw) [17] to look for something similar in a problem solved by Sedgewick around the same time [32]. Without describing the odd-even merge algorithm here, it turns out that

$$\sum_{k \geq 1} \theta(k)\binom{2n}{n-k} \quad \text{with} \quad \theta(k) = \begin{cases} 1 & \text{if } k = 2^i(4j+1), \\ -1 & \text{if } k = 2^i(4j+3) \end{cases}$$

is the quantity to be studied. Note that each positive integer k has a unique representation in this form. The quantity $\theta(k)$ plays the role of $v_2(k)$. Double summation does not lead to the binary representation, but to something similar, namely the Gray code representation of integers. This representation might be described via

$$(n)_{\mathrm{GR}} = \ldots a_2a_1a_0 \quad \text{with} \quad a_k = \left\lfloor \frac{n}{2^{k+2}} + \frac{3}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac{1}{4} \right\rfloor.$$

The Gray code representation has the pleasant property that from $n-1$ to n exactly one bit changes. This means that for the sum-of-digits function $S_{\mathrm{GR}}(n) - S_{\mathrm{GR}}(n-1) = \pm 1$, and, more specifically, $S_{\mathrm{GR}}(n) - S_{\mathrm{GR}}(n-1) = \theta(n)$.

The telescoping summation leads to $S_{\text{GR}}(n) = \sum_{m \leq n} \theta(m)$. A further summation leads to a beautiful explicit result of the Delange type [17].

In this way, Sedgewick's Mellin-type analysis was replaced by a more elementary approach. The typical periodic fluctuations appear, as well as the Dirichlet series $\zeta(s, \frac{3}{4}) - \zeta(s, \frac{1}{4})$.

Later, as part of his series about applications of the Mellin transformation in Combinatorics, Flajolet (plus four coauthors) [11] proposes an elegant high level approach to such questions related to all kinds of problems involving various digits representations:

The Mellin-Perron formula: Let $c > 0$ lie in the half-plane of absolute convergence of the Dirichlet series $\sum_k \lambda_k^{-s}$. Then for any $m \geq 1$, we have

$$\frac{1}{m!} \sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^m = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\sum_{k \geq 1} \frac{\lambda_k}{k^s}\right) n^s \frac{ds}{s(s+1)\ldots(s+m)}.$$

For $m = 0$,

$$\sum_{1 \leq k < n} \lambda_k + \frac{\lambda_n}{2} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\sum_{k \geq 1} \frac{\lambda_k}{k^s}\right) n^s \frac{ds}{s}.$$

The instance $m = 0$ is usually called Perron's summation formula. For problems with the sum-of-digits function, we usually need $m = 1$, which is fortunate, since there are fewer issues about convergence. In most cases, the sum in question can be evaluated by residues. The result is either asymptotic or even exact, as in Delange's case, or in the instance of the Gray code representation. Without going into details, the formula

$$0 = \int_{-\frac{1}{4}-i\infty}^{-\frac{1}{4}+i\infty} \zeta(s) n^s \frac{ds}{s(s+1)}$$

is responsible for the exact formulae. The interested reader is referred to [11].

References

[1] J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge University Press, Cambridge, 2003.

[2] G. E. Andrews and B. C. Berndt, *Your hit parade: the top ten most fascinating formulas in Ramanujan's lost notebook*, Notices Amer. Math. Soc. 55:1 (2008), 18–30.

[3] L. Comtet, *Advanced combinatorics*, D. Reidel Publishing Co., Dordrecht, enlarged edition, 1974. The art of finite and infinite expansions.

[4] N. G. de Bruijn, D. E. Knuth, and S. O. Rice, *The average height of planted plane trees*, In Graph theory and computing, 15–22, 1972.

[5] H. Delange, *Sur la fonction sommatoire de la fonction "somme des chiffres"*, Enseign. Math. (2) 21:1 (1975), 31–47.

[6] G. Doetsch, *Handbuch der Laplace-Transformation*, Birkhäuser Verlag, Basel-Stuttgart, 1972. Band II: Anwendungen der Laplace-Transformation. 1. Abteilung, Verbesserter Nachdruck der ersten Auflage 1955, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften, Mathematische Reihe, Band 15.

[7] S. R. Finch, *Mathematical constants*, Volume 94 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2003.

[8] P. Flajolet, *Combinatorial aspects of continued fractions*, Discrete Math. 32:2 (1980), 125–161.

[9] P. Flajolet, *Approximate counting: a detailed analysis*, BIT 25:1 (1985), 113–134.

[10] P. Flajolet, D. Gardy, and L. Thimonier, *Birthday paradox, coupon collectors, caching algorithms and self-organizing search*, Discrete Appl. Math., 39:3 (1992), 207–229.

[11] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, *Mellin transforms and asymptotics: digital sums*, Theoret. Comput. Sci. 123:2 (1994), 291–314.

[12] P. Flajolet, P. Grabner, P. Kirschenhofer, and H. Prodinger, *On Ramanujan's Q-function*, J. Comput. Appl. Math. 58:1 (1995), 103–116.

[13] P. Flajolet and G. N. Martin, *Probabilistic counting algorithms for data base applications*, J. Comput. System Sci. 31:2 (1985), 182–209. Special issue: Twenty-fourth annual symposium on the foundations of computer science (Tucson, Ariz., 1983).

[14] P. Flajolet, M. Nebel, and H. Prodinger, *The scientific works of Rainer Kemp (1949–2004)*, Theoret. Comput. Sci. 355:3 (2006), 371–381.

[15] P. Flajolet and A. Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math. 3:2 (1990), 216–240.

[16] P. Flajolet and H. Prodinger, *Level number sequences for trees*, Discrete Math. 65:2 (1987), 149–156.

[17] P. Flajolet and L. Ramshaw, *A note on Gray code and odd-even merge*, SIAM J. Comput. 9:1 (1980), 142–158.

[18] P. Flajolet, J.-C. Raoult, and J. Vuillemin, *The number of registers required for evaluating arithmetic expressions*, Theoret. Comput. Sci. 9:1 (1979), 99–125.

[19] P. Flajolet, M. Régnier, and R. Sedgewick, *Some uses of the Mellin integral transform in the analysis of algorithms*, NATO ASI Series F12:1 (1985), 241–254.

[20] P. Flajolet and B. Salvy, *Euler sums and contour integral representations*, Experiment. Math. 7:1 (1998), 15–35.

[21] P. Flajolet and R. Sedgewick, *Digital search trees revisited* SIAM J. Comput. 15:3 (1986), 748–767.

[22] P. Flajolet and R. Sedgewick, *Mellin transforms and asymptotics: finite differences and Rice's integrals*, Theoret. Comput. Sci. 144:1-2 (1995), 101–124. Special volume on mathematical analysis of algorithms.

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

[24] P. Flajolet, P. Zimmermann, and B. V. Cutsem, *A calculus of random generation*, In Algorithms—ESA '93 (Bad Honnef, 1993), Volume 726 of Lecture Notes in Comput. Sci., 169–180, Springer, Berlin, 1993.

[25] G. H. Hardy, *Ramanujan. Twelve lectures on subjects suggested by his life and work*, Cambridge University Press, Cambridge, England; Macmillan Company, New York, 1940.

[26] R. E. Horton, *Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology*, Geological Society of America Bulletin 56:3 (1945), 275–370.

[27] R. Kemp, *The average number of registers needed to evaluate a binary tree optimally*, Acta Inform. 11:4 (1978/79), 363–372.

[28] D. E. Knuth, *The art of computer programming*, Volume 3, Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition.

[29] G. Louchard and H. Prodinger, *Generalized approximate counting revisited*, Theoret. Comput. Sci. 391:1-2 (2008), 109–125.

[30] H. Prodinger and W. Szpankowski, *Philippe Flajolet's research in analysis of algorithms and combinatorics*, Algorithmica, 22:4 (1998), 366–387.

[31] S. Ramanujan, *Master theorem*, `https://en.wikipedia.org/wiki`.

[32] R. Sedgewick, *Data movement in odd-even merging*, SIAM J. Comput. 7:3 (1978), 239–272.

[33] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, Notices Amer. Math. Soc. 65:9 (2018), 1062–1074.

[34] A. N. Strahler, *Dynamic basis of geomorphology*, Geological Society of America Bulletin 63 (1952), 923–938.