Factorisatio Numerorum, Combinatorial Constructions and Gaussian Limit Distributions

Hsien-Kuei Hwang

LIX, École Polytechnique

April 25, 1994

[summary by Xavier Gourdon and Michèle Soria]

Abstract

This talk gives a general overview of some analytic schemas related to counting the number of components in combinatorial constructions over multiplicative structures.

1. Introduction

Each integer n can be uniquely, up to the order of factors, decomposed into a product of prime factors: $n = \prod_{1\le j\le k} p_j^{a_j}$, with $p_1 < \cdots < p_k$, p_j prime, and $j \ge 1, k \ge 1$. From a combinatorial viewpoint, this decomposition is a multiset of prime numbers (components) whose product equals n. Following this line, the problem of integer factorizations can be extended by considering other sets of components, and other ways of counting the factors. For example two classical problems in "Factorisatio Numerorum" are to count the number of factorizations of n into 2,3,4,..., the order of factors being relevant [5] or irrelevant [7].

Hsien-Kuei Hwang's presentation develops a framework for combinatorial constructions on multiplicative structures and their associated Dirichlet series, which shows a perfect similarity with the now classical results in combinatorics for constructions on non labelled additive structures and their associated power series (cf. [9]).

Enumeration problems and probability distributions of factors then rely on an analytic study of Dirichlet series. Along the same line as for additive combinatorial structures (see e.g. [3, 4]), Hwang analyses statistical properties of the number of factors in analytic schemas of *exp-log* form. Precise results of asymptotic normality are given, using Perron's formula and Hankel contours for evaluating integrals. Some extensions to other classes of analytic schemas are also studied. The content of this talk can be found in [4].

2. Combinatorial Constructions

A multiplicative combinatorial structure $\mathcal{C}$ is a set of objects $\alpha \in \mathcal{C}$ with size $|\alpha|$ such that for all n, the number c_n of objects of size n is finite. The enumerating Dirichlet series $C(s)$ of multiplicative class $\mathcal{C}$ is defined by $C(s) = \sum_{\alpha\in\mathcal{C}} |\alpha|^{-s} = \sum_{n\ge 1} c_n n^{-s}$.

Many combinatorial constructions on multiplicative combinatorial structures, related to integer factorizations, translate into simple forms for the associated Dirichlet series. For example the ordinary factorization on $\mathcal{C}$: $n = |\alpha_1|^{a_1}|\alpha_2|^{a_2}\cdots|\alpha_k|^{a_k}$, $\alpha_i \ne \alpha_j$, $\alpha_j \in \mathcal{C}$, $a_j \ge 1$ is associated to the

Construction	Power series $\sum_{n,k} p_{n,k} w^k z^n$	Dirichlet Series $\sum_{n,k} p_{n,k} w^k n^{-s}$
Multiset	$\exp\Big(\sum_{k\geq 1} \frac{w^k}{k} C(z^k)\Big)$	$\exp\Big[\sum_{k\geq 1} \frac{w^k}{k} C(ks)\Big]$
Set	$\exp\Big(\sum_{k\geq 1} \frac{(-1)^{k-1}}{k} w^k C(z^k)\Big]$	$\exp\Big[\sum_{k\geq 1} \frac{(-1)^{k-1}}{k} w^k C(ks)\Big]$
Sequence	$\frac{1}{1-wC(z)}$	$\frac{1}{1-wC(s)}$
Cycle	$\sum_{k\geq 1} \frac{\phi(k)}{k} \log \frac{1}{1-w^k C(z^k)}$	$\sum_{k\geq 1} \frac{\phi(k)}{k} \log \frac{1}{1-w^k C(ks)}$

FIGURE 1. Combinatorial constructions and generating functions
For power series $C(z) = \sum_{n\geq 1} c_n z^n$, and for Dirichlet series $C(s) = \sum_{n\geq 2} c_n n^{-s}$. In the bivariate functions, p_{nk} is the number of structures of size n with k components.

multiset construction $\mathcal{M}[\mathcal{C}]$ of $\mathcal{C}$, symbolically defined by $\mathcal{M}[\mathcal{C}] = \prod_{\alpha\in\mathcal{C}}(\epsilon + \alpha + \alpha^2 + \alpha^3 + \cdots)$. This construction translates easily into Dirichlet series, namely

$$M(s) = \prod_{\alpha\in\mathcal{C}} (1 + |\alpha|^{-s} + |\alpha|^{-2s} + |\alpha|^{-3s} + \cdots) = \exp\left(\sum_{k\geq 1} \frac{C(ks)}{k}\right).$$

For example, taking $\mathcal{C} = \mathcal{P}$ the set of prime numbers leads to Euler's equality

$$\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s} = \prod_{p\in\mathcal{P}} (1 + p^{-s} + p^{-2s} + \cdots) = \exp\left(\sum_{k\geq 1} \frac{P(ks)}{k}\right) \tag{1}$$

since every integer can be uniquely factored into a product of prime numbers.

Table 1 presents the Dirichlet series associated with ordinary factorization (Multiset), square-free factorization (Set), ordered factorization (Sequence) and cyclic factorization (Cycle). They show a perfect analogy with the corresponding power series for unlabelled structures. This table actually presents bivariate series $P(w, z) = \sum_{n,k} p_{n,k} w^k z^n$ and $P(w, s) = \sum_{n,k} p_{n,k} w^k n^{-s}$ where variable w, considered additively, marks the number of $\mathcal{C}$-components.

3. Analytic schemas

Multiplicative compositions lead to analytic schemas with different types of singularities, according to the construction and the Dirichlet series of the class of components.

If the components are $2, 3, 4, \ldots$ then $C(s) = \sum_{n\geq 2} n^{-s}$ is equal to $\zeta(s) - 1$, which is known to have a simple pole at $s = 1$. On the other hand, if the components are prime numbers, Möbius inversion applied to Euler's equality (1) gives $P(s) = \sum_{p\in\mathcal{P}} p^{-s} = \sum_{k\geq 1} \frac{\mu(k)}{k} \log\zeta(ks)$, hence a logarithmic singularity at $s = 1$ for $P(s)$. We are thus faced with Dirichlet series having a polar or a logarithmic singularity. For multiset and set constructions, this leads naturally to analytic schemas of the form

$$P(w, s) = e^{wW(s)}\Psi(w, s), \qquad W(s) = \frac{K}{s-\rho} + H(s) \quad \text{or} \quad W(s) = K \log \frac{1}{s-\rho} + H(s), \tag{2}$$

where $\Psi(w,s)$ and $H(s)$ satisfy some regularity conditions, whereas for sequence and cycle constructions, the schemas become

$$P(w,s) = \frac{K(w)}{s-\rho(w)} + H(w,s) \qquad \text{and} \qquad P(w,s) = \log\frac{1}{s-\rho(w)} + H(w,s). \tag{3}$$

3.1. Enumeration of factorizations. Given $F(s) = \sum f_n n^{-s}$ the Dirichlet series associated with some factorization problem (e.g. in "Factorisatio Numerorum", $F(s) = (2-\zeta(s))^{-1}$ or $F(s) = \exp(\sum(\zeta(ks)-1)/k)$ counts the number of ordered or unordered factorizations of n into integers ≥ 2), the question is to find the asymptotic behaviour of the summatory function $A(x) = \sum_{1\leq n\leq x} f_n$.

When $F(s)$ has a polar singularity at $s=\rho$, with some further conditions, Ikehara showed [5] that $A(x) \sim Kx^\rho$, and Delange [1] extended this result for $F(s)$ with logarithmic singularity. These results, whereas with different techniques of proof, can be brought to Singularity Analysis [2] for evaluating the coefficients of a power series with algebraic and logarithmic singularities.

On the other hand, for exponential singularities, as in $F(s) = \exp(\frac{1}{s-1})$, the saddle-point method applies (as well as in the case of power series) to estimate the summatory function $A(x)$ [6, 7].

3.2. Statistical properties of the number of factors. Considering the bivariate schemas $P(w,s)$ arising from multiplicative compositions where factors are marked by w, we want to asymptotically characterize the distribution of the number of factors : mean value, variance and other moments; central and local limit theorems, probabilities of large deviations;...

More formally, from a Dirichlet series $P(w,s) = \sum_{n\geq 1} P_n(w) n^{-s}$, where $P_n(w)$ are polynomial in w with nonnegative coefficients, we shall study statistical properties of the random variable ξ_n counting the average number of objects of size $\leq n$ with a given number of component, precisely defined by

$$\Pr(\xi_n = m) = \frac{\sum_{1\leq k\leq n} [w^m] P_k(w)}{\sum_{1\leq k\leq n} P_k(1)}$$

where $[w^m]P_k(w)$ denotes the coefficient of w^m in $P_k(w)$.

Each analytic schema described before leads to a Gaussian limit distribution [4]. In the following, we concentrate on the exp-log schema.

4. Exp-log class

The exp-log schema corresponds to bivariate generating functions of the form

$$P(w,s) = \sum_{n\geq 1} P_n(w) n^{-s} = e^{wW(s)}\Psi(w,s).$$

Here, $W(s)$ is a Dirichlet series with non negative coefficients with an abscissa of convergence $\rho > 0$ and can be written

$$W(s) = K\log\frac{1}{s-\rho} + H(s), \qquad K > 0$$

where $H(s)$ is analytic in

$$\Delta(\rho, c) = \{s \mid s = \sigma + it, \sigma \geq \rho - c/V(t)\}$$

with $V(t) = \log(|t|+3)$ and $c > 0$. Moreover $\Psi(w,s)$ is holomorphic for $\Re(s) \geq \rho - \delta$ with a $\delta > 0$ and for $|w| \leq \eta$. Some growth conditions on $P(w,s)$ are needed in the domain $\Delta(\rho,c) \setminus [\rho - c, \rho]$. Roughly speaking, the Dirichlet series $s \mapsto P(w,s)$ has its dominant singularity at $s = \rho$, near which it behaves like $\exp(Kw\log\frac{1}{1-s})$.

From a Perron-like formula, these assumptions lead to an asymptotic expansion of the summatory function $A(x,w) = \sum_{1 \le k \le x} P_k(w)$ as $x \to +\infty$

$$A(x,w) = x^\rho (\log x)^{Kw-1} \left[\sum_{j=0}^{\nu} \frac{\Upsilon_j(w)}{\Gamma(Kw-j)(\log x)^j} + O((\log x)^{-\nu-1}) \right]$$

uniformly for $|w| \le \eta$, where $\Upsilon_j(w)$ is the j-th coefficient of the expansion of $e^{wH(s)}\Psi(w,s)/s$ at $s = \rho$. Thanks to Selberg's method [8], this expansion translates to coefficients, giving

$$\sum_{1 \le k \le x} [w^m] P_k(w) \approx \frac{x^\rho}{\log x} \left(\sum_{j \ge 0} \frac{\kappa_{m,j}(K \log\log x)}{(\log x)^j} \right),$$

where the $\kappa_{m,j}$ are polynomials of degree $m-1$ defined from the $\Upsilon_j(w)$.

From this last expression and thanks to general theorems by Hwang ([4], and see also *Limit Theorems for Combinatorial Structures* in these proceedings), it is possible to derive a full asymptotic expansion of the mean, variance and other moments, to show that the limit distribution is Gaussian and to give its full asymptotic expansion (for the central and local limit theorem), and study large deviations.

Application. Studying the random variable Ω_n counting the number of distinct divisors of a random integer $k \in [1,n]$, that is $\Pr(\Omega_n = m) = \frac{1}{n}|\{k : 1 \le k \le n, \omega(k) = m\}|$ where $\omega(k)$ denotes the number of distinct prime factors of k, leads to a bivariate generating function $P(w,s)$ of exp-log type. The general results of the corresponding schema gives, among others,

$$E(\Omega_n) \sim \log\log n, \quad \mathrm{Var}(\Omega_n) \sim \log\log n$$

and a Gaussian limit distribution as first stated by Erdös-Kac (1940).

Bibliography

[1] Delange (H.). – Généralisation du théorème de Ikehara. *Annales Scientifiques de l'École Normale Supérieure*, vol. 71, 1954, pp. 213–242.

[2] Flajolet (Philippe) and Odlyzko (Andrew M.). – Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, vol. 3, n° 2, 1990, pp. 216–240.

[3] Flajolet (Philippe) and Soria (Michèle). – Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory, Series A*, vol. 53, 1990, pp. 165–182.

[4] Hwang (Hsien-Kuei). – *Théorèmes limites pour les constructions combinatoires et les fonctions arithmétiques.* – Thèse de Doctorat, École Polytechnique, 1994.

[5] Ikehara (S.). – On Kalmar's problem in "factorisatio numerorum". *Journal of the Physico-Mathematical Society of Japan*, vol. 23, 1941, pp. 767–774.

[6] Oppenheim (A.). – On an arithmetic function (II). *Journal of the London Mathematical Society*, vol. 2, 1927, pp. 123–130.

[7] Szekeres (G.) and Turán (P.). – Über das zweite Hauptproblem der "factorisatio numerorum". *Acta Litterarum ac Scientiarum, Szeged*, vol. 6, 1932, pp. 143–154.

[8] Tenenbaum (Gérald). – *Introduction à la Théorie Analytique et Probabiliste des Nombres.* – Institut Élie Cartan, Université de Nancy I, 1990.

[9] Vitter (Jeffrey Scott) and Flajolet (Philippe). – Analysis of algorithms and data structures. In van Leeuwen (J.) (editor), *Handbook of Theoretical Computer Science*, Chapter 9, pp. 431–524. – North Holland, 1990.