A very brief introduction to analytic combinatorics (for and by non-specialist(s))

Mathilde Bouvel
(Loria, CNRS, Univ. Lorraine)

Groupe de travail *Probabilités et Statistiques* de l'IECL
juin 2021

About this talk

My motivations:

- As a preparation for my second talk, on the graphon limit of cographs.
  The proof uses analytic combinatorics strongly, but I won't have time to enter the details in the coming talk.
- If you have never (or rarely) seen it, it is worth some advertisement.
  It is a systematic and powerful approach to the study of discrete structures.
- It can yield results on moments/laws of statistics on discrete objects.
  This is not what I need, but maybe it can be interesting/useful to you.

Contents of the talk:

- The (most classical) example of binary trees
- Some of the main theorems of analytic combinatorics
- One application (which I will use in the second talk)
- Bivariate generating functions and applications in probability theory

Binary trees

Our objects

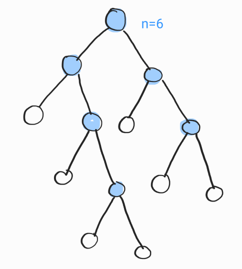
We consider the family $\mathcal{C} = \cup_n \mathcal{C}_n$ of trees which are:

- rooted ($\Rightarrow$ there is a notion of parent/child)
- plane (the children of any node are ordered)
- binary (every node has either 0 or 2 children)

The size, n, of such a tree is the number of internal nodes (=those with 2 children; the nodes with 0 children are leaves).

$\mathcal{C}_n$ = the set of such trees of size n.

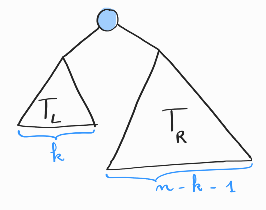
Let $c_n = |\mathcal{C}_n|$.
Which information can we find about c_n?



Decomposition and recurrence

A rooted plane binary tree (of size ≥ 1) consists of

- a root
- a left subtree
- a right subtree



If n is the size of the tree, its left and right subtrees are of sizes k and $n + 1 - k$, for some $0 \leq k \leq n - 1$. Therefore

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k},$$

with initial condition $c_0 = 1$. This recurrence can be solved to yield $c_n = \frac{1}{n+1}\binom{2n}{n}$, the famous sequence of Catalan numbers.

How to solve the recurrence: generating functions

Definition: The ordinary generating function of a sequence (a_n) is the formal power series $\sum_{n\geq 0} a_n z^n$.

Proof of the explicit formula for c_n:

For $C(z) = \sum_{n\geq 0} c_n z^n$, the previous recurrence on c_n translates into an equation satisfied by $C(z)$, namely:

$$C(z) = 1 + z \cdot C(z)^2.$$

Solving this quadratic equation gives $C(z) = \frac{1\pm\sqrt{1-4z}}{2z}$.

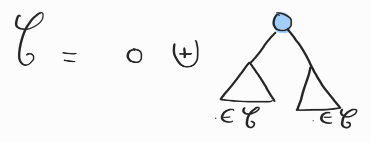
Because $C(0) = c_0 = 1$, we deduce that $C(z) = \frac{1-\sqrt{1-4z}}{2z}$.

We can therefore compute the Taylor expansion of $C(z)$ in $z = 0$, and identifying coefficients we arrive at $c_n = \frac{1}{n+1}\binom{2n}{n}$. □

Symbolic combinatorics: skipping the recurrence

The decomposition of our objects yields the combinatorial specification on the right. From it, we go directly to

$$C(z) = 1 + z \cdot C(z)^2.$$



This is an instance of the dictionary translating operations on combinatorial objects into operations on their generating functions.

Object of size 0	1
Object of size 1 ("atom")	z
Disjoint union	sum of GF
Cartesian product	product of GF
Sequence of elts of $\mathcal{A}$	$\frac{1}{1-A(z)}$
Sets of elts of $\mathcal{A}$	$\exp(A(z))^{(*)}$
...	...

(*): True in the *labeled* context only! I am describing the *unlabeled* context here, where the translation of "sets" is a bit more complicated.

Main ingredients of analytic combinatorics

(First) goal: Find information on the enumerative sequence (a_n) of a combinatorial class $\mathcal{A}$

The chosen setting:

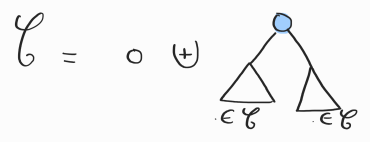
- Consider the generating function $A(z)$ of (a_n)
- Find a specification for $\mathcal{A}$ by combinatorial arguments
- Translate it using the dictionary into a (possibly recursive) equation (or system of equations) satisfied by $A(z)$

Obtaining information from the equation for $A(z)$:

- Consider $A(z)$ as a function of the complex variable z, analytic around $z = 0$
- Study the behavior of $A(z)$ at its dominant singularity(ies), which is (are) the z of smallest modulus where $A(z)$ is not analytic
- Deduce asymptotic estimates of a_n ("transfer")

Back to our example of rooted binary trees

The specification for the class of rooted binary trees shown on the right translates into the equation



$$C(z) = 1 + z \cdot C(z)^2,$$

whose solution is $C(z) = \frac{1-\sqrt{1-4z}}{2z}$.

The dominant singularity of $C(z)$ is $\rho = 1/4$,
and the singular behavior of $C(z)$, as z approaches ρ, is

$$C(z) = 2 - 2\sqrt{1-4z} + o(\sqrt{1-4z}).$$

The transfer theorem then yields $c_n \sim \frac{4^n}{\sqrt{\pi n^3}}$.

Remark: We obtain an asymptotic estimate for c_n, not the closed formula $c_n = \frac{1}{n+1}\binom{2n}{n}$.

Main theorem(s) of analytic combinatorics: the transfer theorem(s)

The two principles of coefficient asymptotics

Let $A(z) = \sum_n a_n z^n$. We aim at showing that $a_n \sim \theta(n) \cdot c^n$.

First principle: The location of the singularities of $A(z)$ determines the exponential growth of a_n (*i.e.*, the constant c).

Second principle: The nature of the singularities of $A(z)$ determines the subexponential factor in the asymptotics of a_n (*i.e.*, the function θ).

On our example of binary trees:

- $C(z) = \frac{1-\sqrt{1-4z}}{2z}$ and $c_n \sim \frac{4^n}{\sqrt{\pi n^3}}$
- The dominant singularity of $C(z)$ is $1/4$,
  determining the exponential growth 4^n of c_n
- It is a squareroot singularity,
  determining the subexponential factor of order $n^{-3/2}$

Starting slow: rational generating functions

Assume $A(z) = \frac{P(z)}{Q(z)}$ for polynomials P and Q.

Compute the partial fraction decomposition: $A(z) = \sum_{i \leq K} \frac{b_i}{(z - z_i)^{m_i}}$

whose expansion is $\sum_{i \leq K} \sum_{n} p_i(n) \cdot z_i^{-n} \cdot z^n$ for some polynomials p_i.

identifying coefficients gives $a_n = \sum_{i \leq K} p_i(n) \cdot z_i^{-n}$.

To obtain an asymptotic estimate of a_n, it is enough to consider only the pole(s) z_i of smallest modulus (*i.e.*, the dominant singularity/ies).

Assuming that there is only one such pole, ρ, we have $a_n \sim p(n) \cdot \rho^{-n}$, and the error is controlled by looking at poles of larger modulus.

Typical application: enumeration of words (avoiding patterns), the Fibonacci sequence

Generalization to meromorphic functions

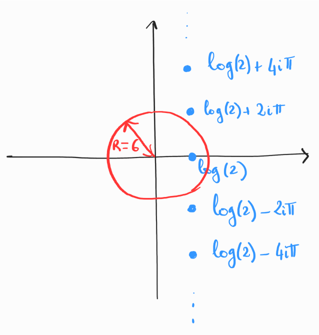
Definition: $A(z)$ is meromorphic if $A(z) = \frac{f(z)}{g(z)}$ for f and g analytic.
Default of analyticity: the poles of $A(z)$ are the $z_1, \dots z_K$ s.t. $g(z_i) = 0$.

Consider $\mathcal{C}_R$, a circle of radius R, such that

- A is analytic on $\mathcal{C}_R$;
- has poles at $z_1, \dots z_m$ inside this circle.

Then, for some polynomials p_i, we have

$$a_n = \sum_{i \leq m} p_i(n) \cdot z_i^{-n} + \mathcal{O}(R^{-n}).$$



Rk.: Exponentially small error term.
The larger the circle, the more precise the estimates.

Proof using residues

Setting: $A(z)$ meromorphic, analytic on $\mathcal{C}_R$, with poles at $z_1, \dots z_m$ inside.

Expansion at z_i: $A(z) = \sum_{n \geq -M_i} \alpha_{i,n}(z - z_i)^n$ for some $M_i \geq 1$
The coefficient $\alpha_{i,-1} = \text{Res}[A(z), z = z_i]$ is the residue of $A(z)$ at $z = z_i$.

Cauchy's residue theorem: $\frac{1}{2\boldsymbol{i}\pi} \oint_\lambda A(z)dz = \sum_i \text{Res}[A(z), z = z_i]$,
where the sum is taken on all i such that z_i is enclosed by λ.

$\Rightarrow$ **Cauchy's coefficient formula:** $a_n = \frac{1}{2\boldsymbol{i}\pi} \oint_\lambda \frac{A(z)}{z^{n+1}} dz$
for λ a simple loop around 0 enclosing no other singularity of A.

Taking $\lambda = \mathcal{C}_R$, we obtain

$$\mathcal{O}(R^{-n}) = \frac{1}{2\boldsymbol{i}\pi} \oint_\lambda \frac{A(z)}{z^{n+1}} dz = a_n + \sum_{i \leq m} \text{Res}[A(z)z^{-n-1}, z = z_i]$$

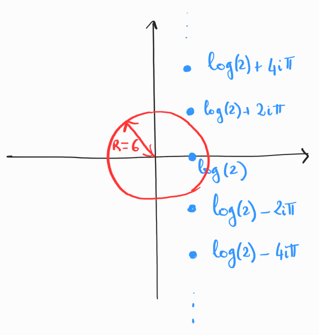
which gives $a_n = \sum_{i \leq m} p_i(n) \cdot z_i^{-n} + \mathcal{O}(R^{-n})$.

Example in the meromorphic case

Typical application: Enumeration of surjections.

For such labeled objects, it is more convenient to use exponential generating functions, defined by $A(z) = \sum_n \frac{a_n}{n!} z^n$.

- EGF of surjections is $A(z) = \frac{1}{2 - \exp(z)}$
- Its poles are $\log(2) + 2ik\pi$.
- Choose $R = 6$.
- Only $\log(2)$ is inside $\mathcal{C}_R$.



This gives

$$\frac{a_n}{n!} = \frac{1}{2\log(2)} \log(2)^{-n} + \mathcal{O}(6^{-n})$$

Implicit functions:

- Earlier cases: $A(z)$ was explicit
- But the combinatorial specification is translated into an equation for $A(z)$, whose solution may be not explicit.
- The equation itself can be enough to estimate a_n.

Typical application: Trees and tree-like structures

$$A(z) = z \cdot \phi(A(z))$$

where $\phi(u)$ can be $\exp(u)$ (Cayley trees), $1 + u^2$ (binary plane trees), ...

Inverse functions:

Defining $\psi(u) := \frac{u}{\phi(u)}$, the above equation rewrites $z = \psi(A(z))$.
Our problem amounts to finding (the behavior of) the inverse of ψ.

Inversion and analyticity

Inversion lemma: An analytic function ψ admits locally an analytic inverse if and only if its first derivative is non-zero.

Intuition (rather than a proof):
Let $\psi(u)$ be analytic at u_0 with $\psi(u_0) = z_0$.

- If $\psi'(u_0) \neq 0$, then $\psi(u) \approx \psi(u_0) + \psi'(u_0)(u - u_0)$,
  hence $u \approx u_0 + \frac{z - z_0}{\psi'(u_0)}$ for $z = \psi(u)$.
- If $\psi'(u_0) = 0$ and $\psi''(u_0) \neq 0$, then $\psi(u) \approx \psi(u_0) + \frac{\psi''(u_0)}{2}(u - u_0)^2$, and this quadratic equation for u has two solutions,
  $u = u_0 \pm \sqrt{\frac{2}{\psi''(u_0)}}\sqrt{z - z_0}$.

  The definition of the squareroot forbids that u is an analytic function of z on a neighborhood of z_0 (we must consider slit neighborhoods instead).

Theorem: Enumeration of trees counted by vertices

Let ϕ be a function analytic at 0, with $\phi(u) = \sum_{n\geq 0} f_n u^n$.
Assume that $f_0 \neq 0$; $\forall n, f_n \geq 0$; and $\exists n \geq 2, f_n \neq 0$.
Let R_ϕ be its radius of convergence.

If $\lim_{u\to R_\phi^-} \frac{u\phi'(u)}{\phi(u)} > 1$, $\frac{\tau\phi'(\tau)}{\phi(\tau)} = 1$ has a unique solution $\tau \in (0, R_\phi)$.

Then, $T(z) = z \cdot \phi(T(z))$ has a unique solution $T(z)$,
which is analytic at 0, and has radius of convergence $\rho = \frac{\tau}{\phi(\tau)} = \frac{1}{\phi'(\tau)}$.

The singular expansion of $T(z)$ near ρ (in an appropriate Δ-domain) is

$$T(z) = \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\left(1 - \frac{z}{\rho}\right)^{1/2} + \mathcal{O}\left(1 - \frac{z}{\rho}\right).$$

Moreover, under aperiodicity conditions,

$$[z^n]T(z) \sim \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}} \cdot \frac{\rho^{-n}}{n^{3/2}}.$$

Notation: $[z^n]T(z)$ is the coefficient t_n of z^n in the expansion $T(z) = \sum_n t_n z^n$

Variant: enumeration of trees counted by leaves

Let Λ be a function analytic at 0, with $\Lambda(u) = \sum_{n\geq 2} \lambda_n u^n$ and $\lambda_n \geq 0$.
Let R_Λ be its radius of convergence.

If $\lim_{u\to R_\Lambda^-} \Lambda'(u) > 1$, $\Lambda'(\tau) = 1$ has a unique solution $\tau \in (0, R_\Lambda)$.

Then, $T(z) = z + \Lambda(T(z))$ has a unique solution $T(z)$,
which is analytic at 0, and has radius of convergence $\rho = \tau - \Lambda(\tau)$.

The singular expansion of $T(z)$ near ρ (in an appropriate Δ-domain) is

$$T(z) = \tau - \sqrt{\frac{2\rho}{\Lambda''(\tau)}}\left(1 - \frac{z}{\rho}\right)^{1/2} + \mathcal{O}\left(1 - \frac{z}{\rho}\right).$$

Moreover, under aperiodicity conditions,

$$[z^n]T(z) \sim \sqrt{\frac{\rho}{2\pi\Lambda''(\tau)}} \cdot \frac{\rho^{-n}}{n^{3/2}}.$$

Examples of application

Objects	$\phi(u)$	Asymptotic behavior
Plane binary trees	$1+u^2$	$\frac{4^n}{\sqrt{\pi n^3}}$
Unary-binary trees	$1+u+u^2$	$\frac{3^{n+1/2}}{2\sqrt{\pi n^3}}$
General plane trees	$(1-u)^{-1}$	$\frac{4^{n-1}}{\sqrt{\pi n^3}}$
Cayley trees	$\exp(u)$	$\frac{e^n}{\sqrt{2\pi n^3}}$
Canonical cotrees	(*)	$\frac{(2\log(2)-1)^{-n+1/2}}{\sqrt{\pi n^3}}$

(*): The studied family of trees satisfies
$T(z) = z + \Lambda(T(z))$ for $\Lambda(u) = \exp(u) - 1 - u$

Generalization: Drmota-Lalley-Woods

Setting: $A(z) := A_0(z)$ is solution of a system of equations, say

$$A_i(z) = P_i(z, A_0(z), \dots, A_N(z)) \text{ for } 1 \leq i \leq N,$$

where the P_i are polynomials.

The nice case: The dependency graph of the system is strongly connected, meaning that the equation defining each series A_i "uses" (possibly unraveling the definitions) each A_j.

In this case, we observe the same squareroot-type singular behavior, with asymptotics of coefficients in $c \cdot \rho^{-n} n^{-3/2}$.

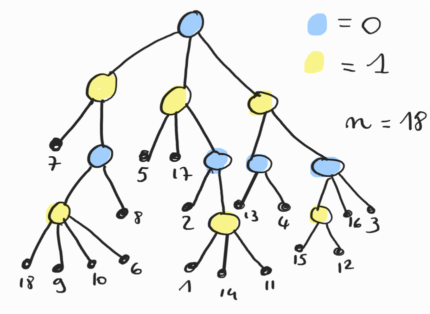
Beyond: Other exponents can occur.
Namely, the coefficients behave asymptotically in $c \cdot \rho^{-n} n^{\beta}$
for $\beta = -1 - 2^{-k}$ for $k \geq 1$ or $\beta = -1 + m2^{-k}$ for $m \geq 1$ and $k \geq 0$.

One application: canonical cotrees
(preparation for next talk)

Canonical cotrees

Definition: A labeled canonical cotree of size n is

- a rooted tree
- which is non-plane,
- having n leaves,
- labeled from 1 to n,
- where internal nodes have at least two children
- and carry decorations 0 or 1,
- in such a way that decorations along each path from the root to a leaf alternate.



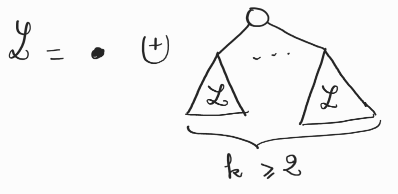
Remark:
Canonical cotrees bijectively encode cographs, which we will study next week.

A simpler class: forget the decorations

Let $\mathcal{L}$ be the family of non-plane rooted trees, labeled on their leaves, where internal nodes have at least two children.

Let $L(z) = \sum_n \frac{\ell_n}{n!} z^n$ be the exponential generating function of $\mathcal{L}$.

A specification for $\mathcal{L}$ is



Applying the dictionary (for labeled objects to EGF) gives

$$L(z) = z + \Lambda(L(z)) \quad \text{for} \quad \Lambda(u) = \exp(u) - 1 - u.$$

Singular behavior of $L(z)$

Since $\Lambda(u) = \exp(u) - 1 - u = \sum_{n\geq 2} \frac{u^n}{n!}$ is analytic at 0, with radius of convergence $+\infty$, we can apply the theorem for enumeration of trees.

- $\Lambda'(\tau) = 1$ defines $\tau = \log(2)$.
- The equation $L(z) = z + \Lambda(L(z))$ has a unique solution $L(z)$ which is analytic at 0, with radius of convergence $\rho = \tau - \Lambda(\tau) = 2\log(2) - 1$.
- Near ρ, we have

$$L(z) = \log(2) - \sqrt{\rho}\sqrt{1 - z/\rho} + \mathcal{O}(1 - z/\rho).$$

- Consequently,

$$[z^n]L(z) = \frac{\ell_n}{n!} \underset{n\to+\infty}{\sim} \frac{1}{2}\sqrt{\frac{\rho}{\pi}} \cdot \frac{\rho^{-n}}{n^{3/2}}.$$

Back to canonical cotrees

Reminder:
Canonical cotrees are trees of $\mathcal{L}$ with additional decorations 0 and 1 on internal nodes, which alternate on each path from the root to a leaf.

Fact:
Because of alternation, all decorations are determined by that of the root.

Consequences:
The exponential generating series $M(z)$ of canonical cotrees satisfies $M(z) = 2(L(z) - z) + z = 2L(z) - z$. Therefore,

- $M(z)$ is analytic at 0 with radius of convergence $\rho = 2\log(2) - 1$.
- Near ρ, we have $M(z) = 1 - 2\sqrt{\rho}\sqrt{1 - z/\rho} + \mathcal{O}(1 - z/\rho)$.
- And $[z^n]M(z) \underset{n\to+\infty}{\sim} \sqrt{\frac{\rho}{\pi}} \cdot \frac{\rho^{-n}}{n^{3/2}}$.

Application:
Used in the proof of the graphon convergence of cographs, next week.

Bivariate generating series and distribution of statistics in random discrete objects

Statistics on objects and bivariate GF

Reminder: The (ordinary) generating function of a combinatorial class $\mathcal{A}$ is $A(z) = \sum_{n \geq 0} a_n z^n$.

Definition: Consider a statistic χ on the objects of $\mathcal{A}$.
The bivariate (ordinary) generating function of $\mathcal{A}$ for the statistic χ is

$$A(z, u) = \sum_{n \geq 0} \sum_{k} a_{n,k} u^k z^n$$

where $a_{n,k}$ is the number of objects α of size n in $\mathcal{A}$ such that $\chi(\alpha) = k$.

Fact: $A(z, 1) = A(z)$, or equivalently $a_n = \sum_k a_{n,k}$ for all n.

Example: For $\mathcal{A} = \{\text{words on } \{a, b\}\}$ and $\chi =$ number of a's, we have

$$A(z, u) = \sum_{n \geq 0} \sum_{k=0}^{n} \binom{n}{k} u^k z^n = \frac{1}{1 - z(u+1)}.$$

Extension of the dictionary

In many cases, the statistics can be tracked in the translation of the dictionary "operation on objects" $\rightarrow$ "operation on GF".

Examples:

- Number of components in a sequence: $\frac{1}{1-uA(z)}$
- ...or in a set: $\exp(uA(z))$
- ...or in any operation
- Number of leaves in general plane trees: $G(z,u) = zu + \frac{zG(z,u)}{1-G(z,u)}$
- Pathlength in trees: $A(z,u) = z\phi(A(zu,u))$
- and many more

The pathlength is the sum of all the distances from a node to the root in a tree.

Moments from bivariate generating functions

Setting: Class $\mathcal{A}$, parameter χ, with bivariate GF $\sum_{n\geq 0}\sum_k a_{n,k} u^k z^n$.
Consider the uniform distribution over $\mathcal{A}_n = \{\text{objects of size } n \text{ in } \mathcal{A}\}$.

Expectation: Let X_n be the r.v. which records the value the parameter χ.

$$\mathbb{E}[X_n] = \sum_k k \cdot \mathbb{P}(\chi = k) = \sum_k k \cdot \frac{a_{n,k}}{a_n} = \frac{[z^n]\frac{\partial A(z,u)}{\partial u}\,|_{u=1}}{[z^n]A(z,1)}$$

Higher moments: Classical moments are obtained from factorial moments, and

$$\mathbb{E}[X_n(X_n-1)\dots(X_n-d+1)] = \frac{[z^n]\frac{\partial^d A(z,u)}{\partial u^d}\,|_{u=1}}{[z^n]A(z,1)}$$

To remember: Derivative in $u=1$ gives access to moments of χ.

Letting $n \to \infty$

Goal: Describe the limiting behavior of X_n as $n \to \infty$.

The easy cases: If $A(z, u)$ is explicit, and the moments of X_n can be explicitly computed, then we may be able to

- prove concentration of χ,
- find the distributional limit of X_n provided it is determined by its moments.

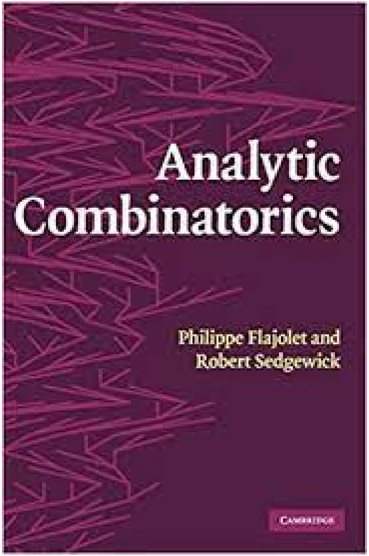
Going further:

- Apply analytic combinatorics to $A(z, u)$ (for the variable z, as usual, u being seen as a parameter).
- Study the dependency in u, close to $u = 1$, of the singular behavior so obtained to derive the limit law of χ.
- We may also obtain information on the speed of convergence and tail estimates.

Example of limit laws encountered

The parameter ...	in ...	has limit law ...
Number of initial a's	binary words	geometric
Number of initial a's	Dyck words	Negative binomial distribution
Cycles of length 1 (or m)	permutations	Poisson
Root degree	general trees	shifted Poisson
Number of a's	binary words	Gaussian
Number of cycles	permutations	Gaussian
Image cardinality	Surjections	Gaussian
Perimeter	Parallelogram polyominoes	Gaussian
Number of leaves	general trees	Gaussian
Position of first max.	random walk	arcsin law
Path length	general trees	Airy distribution
Area under the path	Dyck path	Airy distribution
Contacts with x-axis	bridges	Rayleigh law

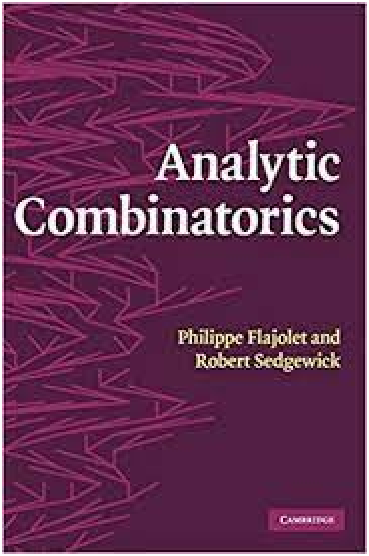
If you want to know more: the purple book



The reference on the topic.

- over 800 pages
- many examples
- “black-box” theorems, but also guidelines to study examples which do not satisfy the hypothesis of these “black-box” theorems

If you want to know more: the purple book



The reference on the topic.

- over 800 pages
- many examples
- "black-box" theorems, but also guidelines to study examples which do not satisfy the hypothesis of these "black-box" theorems

Merci, et à la semaine prochaine pour parler de cographes !