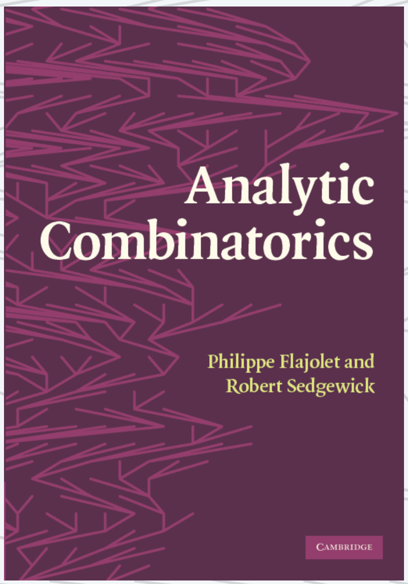
ANALYTIC COMBINATORICS

PART TWO



http://ac.cs.princeton.edu

3. Combinatorial Parameters and MGFs

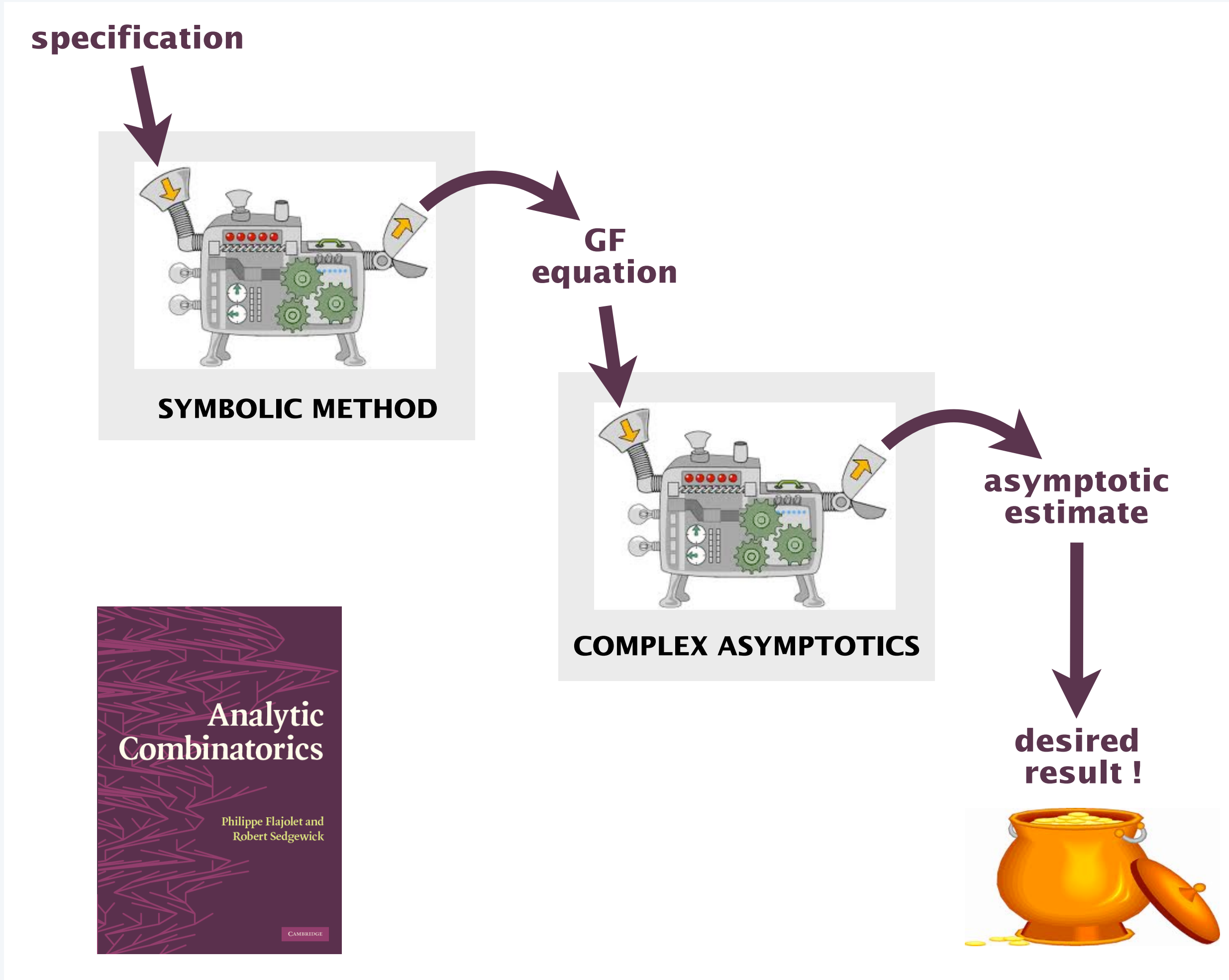
Analytic combinatorics overview

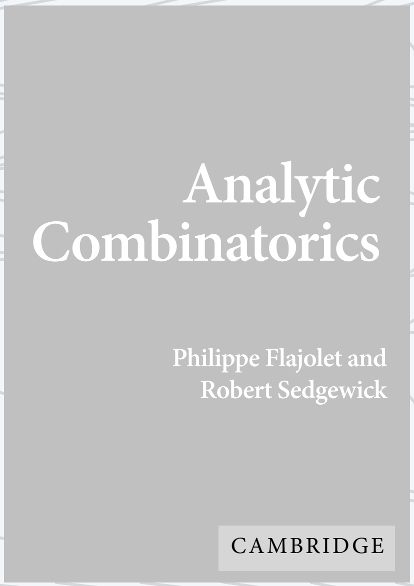
A. SYMBOLIC METHOD

1. OGFs
2. EGFs
3. MGFs

B. COMPLEX ASYMPTOTICS

4. Rational & Meromorphic
5. Applications of R&M
6. Singularity Analysis
7. Applications of SA
8. Saddle point





http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- **Basics**
- Moment calculations
- OBGF examples
- Labelled classes

Natural questions about combinatorial parameters

What is the average number of *subsets* in a random **set partition** ?

What is the average number of *parts* in a random **composition** ?

What is the average number of *parts* in a random **partition** ?

What is the average number of *cycles* in a random **permutation** ?

What is the average *root degree* of a random **tree** ?

What is the average number of *times each letter appears* in a random ***M*-word** ?

What is the average number of *leaves* in a random **tree** ?

Natural questions about combinatorial parameters

Problem: Average-case results are sometime easy to derive but unsatisfying.

Example. *Separate chaining hashing* randomly assigns N keys to M lists.

Q. Average length of a list ?

A. N/M.

A *trivial* result that is not very useful because it says nothing about the length of a particular list.

Ex: All the keys could be on one list.
Avg. length = $(N + 0 + 0 + \ldots + 0)/M = N/M$

Section 3.4

Solution: Find *distribution* (probability parameter value is k for all k)

Practical compromises:

- compute average *and* variance
- compute average *extremal* values

Ex: Bound probability that list length deviates significantly from average.

Ex: Compute average length of the *longest* list.

Goals for this lecture: Learn enough about parameters to be able to

- compute full distribution (in principle)
- compute moments and extremal values (in practice)

Natural questions about combinatorial parameters

How many ways to partition a **set** of *N* objects into *k* subsets?

How many **compositions** (sequences of positive integers that sum to N) have *k* parts?

How many **partitions** (sets of positive integers that sum to *N*) have *k* parts?

How many **permutations** of size *N* have *k* cycles?

How many **trees** with *N* nodes have root degree *k* ?

How many letters appear *k* times in an ***M*-word** of length N ?

How many **trees** with *N* nodes have *k* leaves ?

Basic definitions (combinatorial parameters for unlabelled classes)

Def. A *combinatorial class* is a set of combinatorial objects and an associated size function that may have an associated parameter.

Def. The *ordinary bivariate generating function* (OBGF) associated with a class is the formal power series

size function

$$A(z,u) = \sum_{a \in A} z^{|a|} u^{cost(a)}$$

parameter value

object name

class name

Fundamental (elementary) identity

$$A(z) \equiv \sum_{a \in A} z^{|a|} u^{cost(a)} = \sum_{N \geq 0} \sum_{k \geq 0} A_{Nk} z^N u^k$$

Terminology.

The variable *z marks* size
The variable *u marks* the parameter

Q. How many objects of size *N* with value *k*?

A. $A_{Nk} = [z^N][u^k]A(z,u)$

Terminology.

BGF: bivariate GF.
MGF: multivariate GF

might add arbitrary number of markers

With the symbolic method, we specify the class *and at the same time* characterize the OBGF

Combinatorial enumeration: classic example

Q. How many binary strings with N bits?

0
1

$B_1 = 2$

0	0
0	1
1	0
1	1

$B_2 = 4$

0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

$B_3 = 8$

0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

$B_4 = 16$

A. $B_N = 2^N$

Combinatorial parameters: classic example

Q. How many N-bit binary strings have k 0 bits?

0
1

$B_{10} = 1$
$B_{11} = 1$

0 0
0 1
1 0
1 1

$B_{20} = 1$
$B_{21} = 2$
$B_{22} = 1$

0 0 0
0 0 1
0 1 0
0 1 1
1 0 0
1 0 1
1 1 0
1 1 1

$B_{30} = 1$
$B_{31} = 3$
$B_{32} = 3$
$B_{33} = 1$

0 0 0 0
0 0 0 1
0 0 1 0
0 0 1 1
0 1 0 0
0 1 0 1
0 1 1 0
0 1 1 1
1 0 0 0
1 0 0 1
1 0 1 0
1 0 1 1
1 1 0 0
1 1 0 1
1 1 1 0
1 1 1 1

$B_{40} = 1$
$B_{41} = 4$
$B_{42} = 6$
$B_{43} = 4$
$B_{43} = 1$

A. $B_{Nk} = \binom{N}{k}$

OBGF of binomial coefficients

$$\sum_{N\geq 0}\sum_{k\geq 0}\binom{N}{k}u^k z^N$$

$$= \sum_{N\geq 0}(1+u)^N z^N \qquad (\textit{horizontal}\ \text{OGF})$$

$$= \sum_{k\geq 0}\frac{z^k}{(1-z)^{k+1}}u^k \qquad (\textit{vertical}\ \text{OGF})$$

$$= \frac{1}{1-z(1+u)} \qquad \text{(OBGF)}$$

$N \searrow$ / $k \rightarrow$	0	1	2	3	4	5	6	7	8	9
0	1									
1	1	1								
2	1	2	1							
3	1	3	3	1						
4	1	4	6	4	1					
5	1	5	10	10	5	1				
6	1	6	15	20	15	6	1			
7	1	7	21	35	35	21	7	1		
8	1	8	28	56	70	56	28	8	1	
9	1	9	36	84	126	126	84	36	9	1

vertical OGF coefficients

horizontal OGF coefficients

$[u^5][z^7] = \binom{7}{5}$

$[u^k](1+u)^7 = \binom{7}{k}$

$[z^N]\dfrac{z^5}{(1-z)^6} = \binom{N}{6}$

The symbolic method for OBGFs (basic constructs)

Suppose that A and B are classes of unlabelled objects with *OBGFs* $A(z,u)$ and $B(z,u)$ where z marks size and u marks a parameter value. Then

operation	notation	semantics	OGF
disjoint union	$A + B$	disjoint copies of objects from A and B	$A(z,u) + B(z,u)$
Cartesian product	$A \times B$	ordered pairs of copies of objects, one from A and one from B	$A(z,u)B(z,u)$
sequence	$SEQ(A)$	sequences of objects from A	$\dfrac{1}{1 - A(z,u)}$

Construction immediately gives OBGF equation, as for enumeration.

Extends immediately to mark multiple parameters simultaneously with MGFs.

Proofs of correspondences

$A + B$

$$\sum_{c\in A+B} z^{|c|}u^{cost(c)} = \sum_{a\in A} z^{|a|}u^{cost(a)} + \sum_{b\in B} z^{|b|}u^{cost(b)} = A(z,u) + B(z,u)$$

$A \times B$

$$\sum_{c\in a\times b} z^{|c|}u^{cost(c)} = \sum_{a\in A}\sum_{b\in B} z^{|a|+|b|}u^{cost(a)+cost(b)} = \Big(\sum_{a\in A} z^{|a|}u^{cost(a)}\Big)\Big(\sum_{b\in B} z^{|b|}u^{cost(b)}\Big)$$

$$= A(z,u)B(z,u)$$

$SEQ(A)$

construction	OGF
$SEQ_k(A) \equiv A^k$	$A(z,u)^k$
$SEQ_T(A) \equiv A^{t_1} + A^{t_2} + A^{t_3} + \ldots$ where $T \equiv t_1, t_2, t_3, \ldots$ is a subset of the integers	$A(z,u)^{t_1} + A(z,u)^{t_2} + A(z,u)^{t_3} + \ldots$
$SEQ(A) \equiv \epsilon + A + A^2 + A^3 + \ldots$	$1 + A(z,u) + A(z,u)^2 + \ldots = \dfrac{1}{1 - A(z,u)}$

Combinatorial parameter example: 0 bits in bitstrings

Class	B, the class of all binary strings
Size	$\lvert b \rvert$, the number of bits in b
Parameter	$zeros(b)$, the number of 0 bits in b
OBGF	$B(z,u) = \sum_{b \in B} z^{\lvert b \rvert} u^{zeros(b)} = \sum_{N \ge 0} \sum_{k \ge 0} B_{Nk} z^N u^k$

variable u "marks" the parameter

Construction

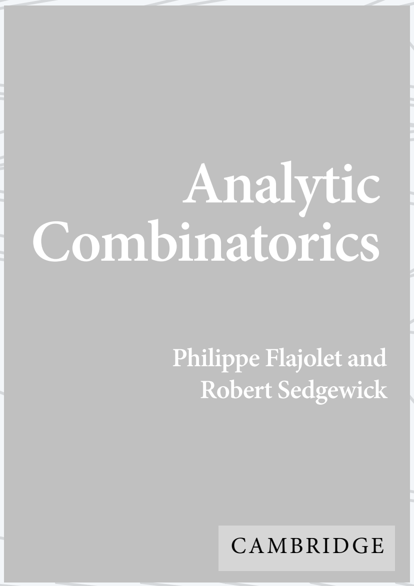
$$B = SEQ\,(uZ_0 + Z_1)$$

OBGF equation

$$B(z,u) = \frac{1}{1 - z(1+u)}$$

Expansion

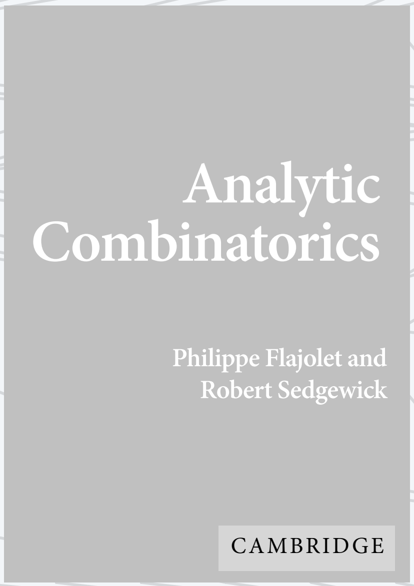
$$B_{Nk} \equiv [u^k][z^N]B(z,u) = [u^k](1+u)^N = [z^N]\frac{z^k}{(1-z)^{k+1}} = \binom{N}{k} \quad ✓$$



http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- **Basics**
- Moment calculations
- OBGF examples
- Labelled classes



http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- **Moment calculations**
- OBGF examples
- Labelled classes

OBGF moment calculations

OBGF $\quad P(z,u) = \sum_{p \in P} z^{|p|} u^{cost(p)}$

size function

parameter value

object name

class name

$$P(z,u) = \sum_{N \geq 0} \sum_{k \geq 0} p_{Nk} u^k z^N$$

Enumeration

$P_N \equiv [z^N]P(z,1) \qquad P(z,1) = \sum_{p \in \mathcal{P}} z^{|p|}$

number of objects of size N

$$P_N \equiv \sum_{k \geq 0} p_{Nk}$$

$$P(z,1) = \sum_{N \geq 0} P_N z^N = \sum_{N \geq 0} \sum_{k \geq 0} p_{Nk} z^N$$

Cumulated cost

$Q_N \equiv [z^N]P_u(z,1) \qquad P_u(z,1) = \sum_{p \in \mathcal{P}} \text{cost}(p) z^{|p|}$

$$\left.\frac{\partial P(z,u)}{\partial u}\right|_{u=1}$$

total cost in objects of size N

$$Q_N = \sum_{k \geq 0} k p_{Nk}$$

$$P_u(z,1) = \sum_{N \geq 0} Q_N z^N = \sum_{N \geq 0} \sum_{k \geq 0} k p_{Nk} z^N$$

Mean cost of objects of size *N* $\quad \mu_N = \frac{[z^N]P_u(z,1)}{[z^N]P(z,1)} = \frac{Q_N}{P_N}$

$$\mu_N = \sum_{k \geq 0} \frac{p_{Nk}}{P_N} k$$

Variance $\quad \sigma_N^2 = \frac{[z^N]P_{uu}(z,1)}{[z^N]P(z,1)} + \mu_N - \mu_N^2$

$$\sigma_N^2 = \sum_{k \geq 0} \frac{p_{Nk}}{P_N} (k - \mu_N)^2$$

Moments for 0 bits in bitstrings with OBGFs

Class	*B*, the class of all binary strings
Size	$\lvert b \rvert$, the number of bits in *b*
Parameter	*zeros*(*b*), the number of 0 bits in *b*

Example	1 0 1 1 1 0 1 0 0 0 1 0 0 0
OBGF	$B(z,u) = \sum_{b \in B} z^{\lvert b \rvert} u^{zeros(b)}$

Construction $\quad B = SEQ\,(uZ_0 + Z_1)$

OBGF equation $$B(z,u) = \frac{1}{1 - z(1+u)}$$

$$B_u(z,u) = \frac{z}{(1 - z - zu)^2}$$

Enumeration $$[z^N]B(z,1) = [z^N]\frac{1}{1-2z} = 2^N$$

Cumulated cost $$[z^N]B_u(z,1) = [z^N]\frac{z}{(1-2z)^2} = N2^{N-1}$$

Mean cost of objects of size *N* $$\mu_N = \frac{[z^N]B_u(z,1)}{[z^N]B(z,1)} = N/2 \quad \checkmark$$

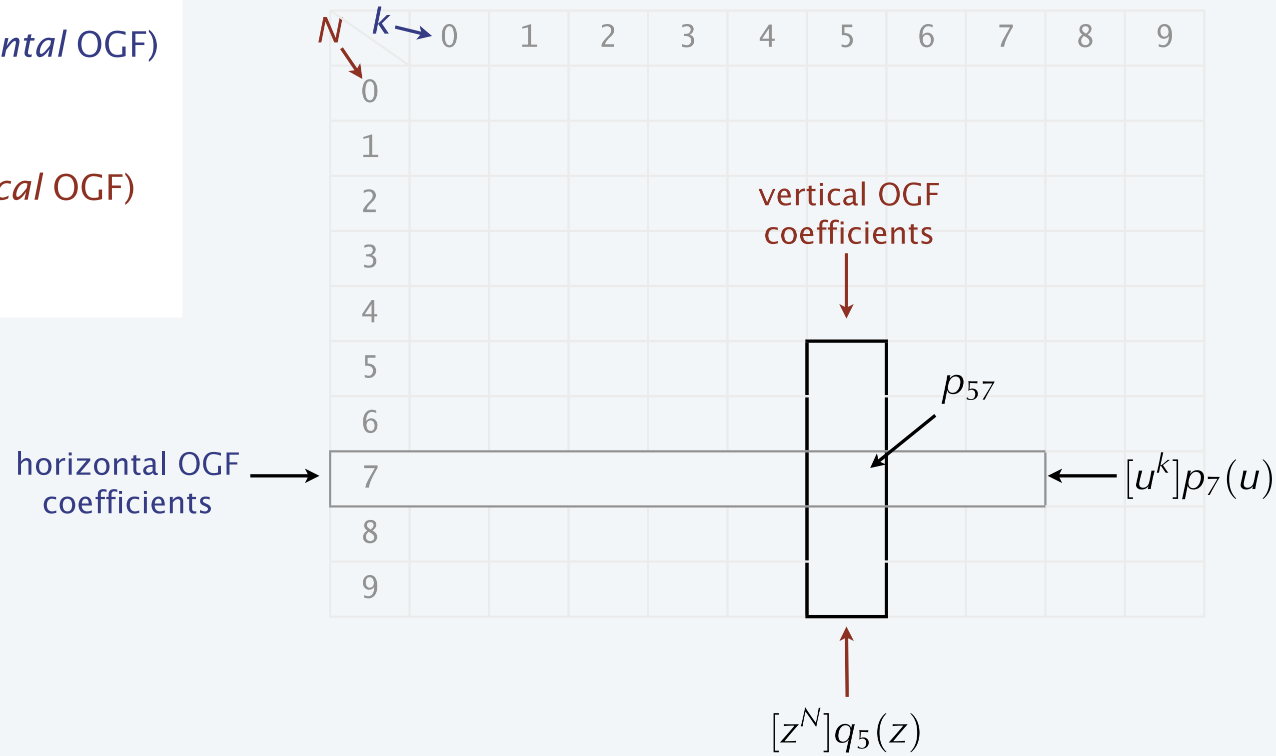
Variance (easier with horizontal GFs: stay tuned)

"Horizontal" and "vertical" OGFs

$$\sum_{N\geq 0}\sum_{k\geq 0} p_{Nk} z^N u^k$$

$$= \sum_{N\geq 0} p_N(u) z^N \qquad (\textit{horizontal}\ \text{OGF})$$

$$= \sum_{k\geq 0} q_k(z) u^k \qquad (\textit{vertical}\ \text{OGF})$$



Moment calculations ("horizontal" OGF)

size function

OBGF. $$P(z,u) = \sum_{p \in P} z^{|p|} u^{cost(p)}$$

parameter value

object name

class name

$$P(z,u) = \sum_{N \geq 0} \sum_{k \geq 0} p_{Nk} u^k z^N$$

GF for costs of objects of size N

"Horizontal" OGF $$[z^N]P(u,z) \equiv p_N(u) = \sum_{p \in P \text{ and } size(p)=N} u^{cost(p)}$$

$$p_N(u) = \sum_{k \geq 0} p_{Nk} u^k$$

Enumeration $$p_N(1) = \sum_{p \in \mathcal{P}_N} 1 = P_N$$

$$p_N(1) = \sum_{k \geq 0} p_{Nk} = P_N$$

Cumulated cost $$p'_N(1) = \sum_{p \in \mathcal{P}_N} \mathrm{cost}(p) = Q_N$$

$$p'_N(1) = \sum_{k \geq 0} k p_{Nk} = Q_N$$

Mean cost of objects of size *N* $$\mu_N = \frac{p'_N(1)}{p_N(1)} = \frac{Q_N}{P_N}$$

$$\mu_N = \sum_{k \geq 0} \frac{p_{Nk}}{P_N} k$$

Variance $$\sigma_N^2 = \frac{p''_N(1)}{p_N(1)} + \mu_N - {\mu_N}^2$$

$$\sigma_N^2 = \sum_{k \geq 0} \frac{p_{Nk}}{P_N} (k - \mu_N)^2$$

0 bits in bitstrings with a "horizontal" OGF

OBGF	$B(z,u) = \dfrac{1}{1 - z(1+u)}$
"Horizontal" OGF	$b_N(u) \equiv [z^N]B(z,u) = (1+u)^N$
Enumeration	$b_N(1) = 2^N$
Cumulated cost	$b'_N(1) = N2^{N-1}$
Average # 1-bits in a random *N*-bit string	$b'_N(1)/b_N(1) = N2^{N-1}/2^N = N/2$ ✓
Variance	$b''_N(1)/b_N(1) + N/2 - (N/2)^2 = N/4$

concentrated: $\sigma_N = \sqrt{N}/2$ (stay tuned)

Moment calculations ("vertical" OGF)

size function

OBGF. $$P(z,u) = \sum_{p \in P} z^{|p|} u^{cost(p)}$$

parameter value

object name

class name

GF for costs of objects of cost k

"Vertical" OGF $$[u^k]P(u,z) \equiv q_k(z) = \sum_{p \in P \text{ and } cost(p)=k} z^{|p|}$$

Enumeration $$P_N \equiv [z^N]P(z,1)$$

Cumulated cost. $$[z^N]\sum_k kq_k(z) = Q_N$$

Mean cost of objects of size *N* $$\mu_N = \frac{Q_N}{P_N}$$

Variance (omitted)

$$P(z,u) = \sum_{N \geq 0} \sum_{k \geq 0} p_{Nk} u^k z^N$$

$$q_k(z) = \sum_{N \geq 0} p_{Nk} z^N$$

$$\sum_k kq_k(z) = \sum_k \sum_{N \geq 0} kp_{Nk} z^N$$

$$= \sum_{N \geq 0} \Big(\sum_k kp_{Nk} \Big) z^N$$

Q_N

$$\mu_N = \sum_{k \geq 0} \frac{kp_{Nk}}{P_N}$$

0 bits in bitstrings with a "vertical" OGF

OBGF $$B(z,u) = \frac{1}{1 - z(1+u)}$$

"Vertical" OGF $$q_k(z) = [u^k]B(z,u) = \frac{z^k}{(1-z)^{k+1}}$$

Enumeration $$P_N = [z^N]B(z,1) = 2^N$$

Cumulated cost $$Q_N = [z^N]\sum_k k\frac{z^k}{(1-z)^{k+1}}$$

$$= N2^{N-1}$$

Average # 1-bits in a random *N*-bit string $$P_N/Q_N = N/2 \checkmark$$

$$\sum_k kr^{k-1} = \frac{1}{(1-r)}2$$

$$\begin{aligned}\sum_k k\frac{z^k}{(1-z)^{k+1}} &= \frac{z}{(1-z)^2}\sum_k k\frac{z^{k-1}}{(1-z)^{k-1}} \\ &= \frac{z}{(1-z)^2}\frac{1}{(1-\frac{z}{1-z})^2} \\ &= \frac{z}{(1-2z)^2}\end{aligned}$$

Moment inequalities and concentration

Let X_N be the value of a parameter for a random object of size N with mean μ_N and std dev σ_N.

Markov inequality. $\Pr\{X_N \geq t\mu_N\} \leq 1/t$

Chebyshev inequality. $\Pr\{|X_N - \mu_N| \geq t\sigma_N\} \leq 1/t^2$

"The probability of being much larger than the mean must decay, and an upper bound on the rate of decay is measured in units given by the standard deviation."

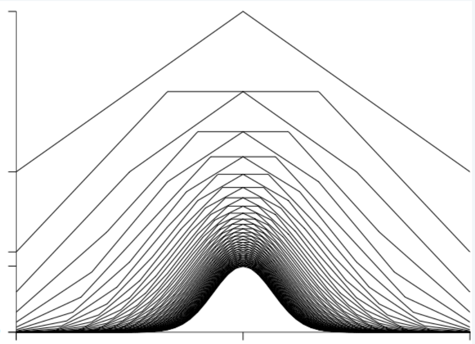
Def. A distribution is *concentrated* if $\sigma_N = o(\mu_N)$.

Proposition. If a distribution is concentrated,
then $X_N/\mu_N \to 1$ in probability: $\lim_{N\to\infty} \Pr\{1-\epsilon \leq \frac{X_N}{\mu_N} \leq 1+\epsilon\} = 1$

When a distribution is concentrated, the expected value is "typical".

Example: 100,000,000 random bits

Expected # 1 bits	$N/2$	50,000,000
Standard deviation	$\sqrt{N}/2$	5,000
Probability X_N is between 49,900,000 and 50,100,000		**.9975**



Moments for letters in M-words with OBGFs

Class	W_M, the class of all M-words
Size	$\|w\|$, the number of letters in w
Parameter	$occ(w)$, # of occurrences of a given letter in w

Example	4 3 5 5 2 4 1 1 2 3
OBGF	$W_M(z,u) = \sum_{w \in W_M} z^{\|w\|} u^{occ(w)}$

Construction $\quad B = SEQ\,(uZ + (M-1)Z)$

OBGF equation $\quad W_M(z,u) = \dfrac{1}{1-(M-1+u)z}$

Enumeration $\quad [z^N]W(z,1) = [z^N]\dfrac{1}{1-Mz} = M^N$

Cumulated cost $\quad [z^N]W_u(z,1) = [z^N]\dfrac{z}{(1-Mz)^2} = NM^{N-1}$

$$[z^N]W_u(z,1) = NM^{N-1}$$
$$[z^N]W_{uu}(z,1) = N(N-1)M^{N-2}$$

Mean # of occurences of a given letter in a random *M*-word with *N* letters $\quad \mu_N = \dfrac{[z^N]W_u(z,1)}{[z^N]W(z,1)} = N/M$ ✓

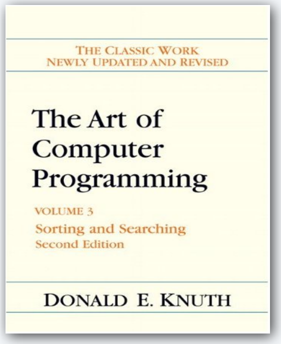
Variance $\quad \sigma_N^2 = [z^N]\dfrac{W_{uu}(z,1)}{[z^N]W(z,1)} + \mu_N - \mu_N^2 = N/M - N/M^2$

Standard deviation $\quad \sigma_N = \sqrt{N/M - N/M^2}$ ⟵ concentrated for fixed M

Application: Hashing algorithms

Goal: Provide efficient ways to

- *Insert* key-value pairs in a *symbol table*.
- *Search* the table for the pair corresponding to a given key.



Strategy

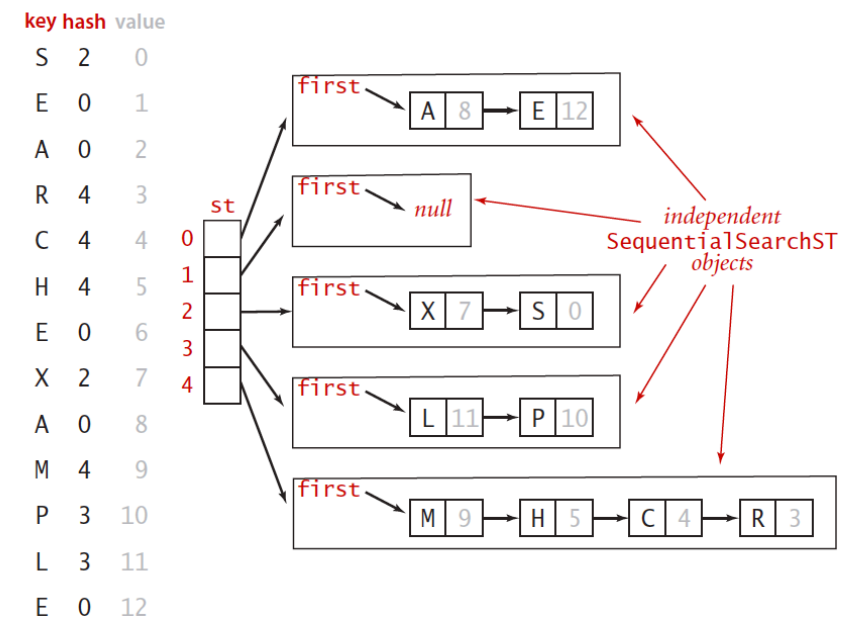
- Develop a *hash function* that maps each key into value between 0 and $M-1$.
- Maintain M lists of key-value pairs

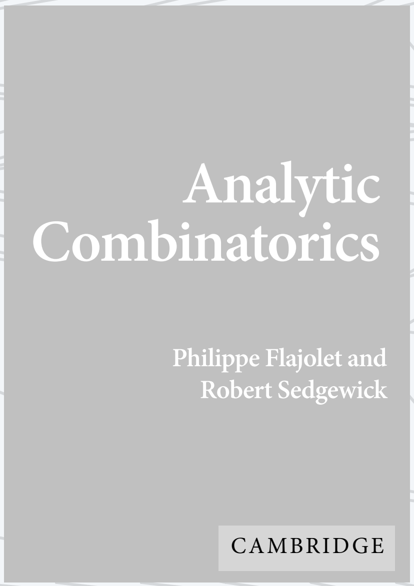
Q. Average list length for N keys?

A. N/M ← Trivial

Q. *Typical* list length for N keys, for fixed M?

A. N/M, concentrated ← Useful

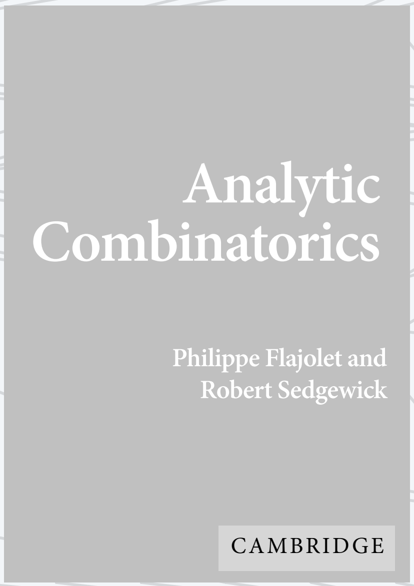




http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- **Moment calculations**
- OBGF examples
- Labelled classes



http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- Moment calculations
- **OBGF examples**
- Labelled classes

Number of parts in compositions

Q. How many compositions of N have k parts?

1

$C_{11} = 1$

1 + 1
2

$C_{21} = 1$

$C_{22} = 1$

cumulated cost: 3
average: 1.5

1 + 1 + 1
1 + 2
2 + 1
3

$C_{31} = 1$

$C_{32} = 2$

$C_{33} = 1$

cumulated cost: 8
average: 2

1 + 1 + 1 + 1
1 + 1 + 2
1 + 2 + 1
1 + 3
2 + 1 + 1
2 + 2
3 + 1
4

$C_{41} = 1$

$C_{42} = 3$

$C_{43} = 3$

$C_{44} = 1$

cumulated cost: 20
average: 2.5

$C_{41} + 2C_{42} + 3C_{43} + 4C_{44} = 20$

1 + 1 + 1 + 1 + 1
1 + 1 + 1 + 2
1 + 1 + 2 + 1
1 + 1 + 3
1 + 2 + 1 + 1
1 + 2 + 2
1 + 3 + 1
1 + 4
2 + 1 + 1 + 1
2 + 1 + 2
2 + 2 + 1
2 + 3
3 + 1 + 1
3 + 2
4 + 1
5

$C_{51} = 1$

$C_{52} = 4$

$C_{53} = 6$

$C_{54} = 4$

$C_{55} = 1$

cumulated cost: 48
average: 3

A. $C_{Nk} = \binom{N-1}{k-1}$

Number of parts in compositions

Class	C, the class of all compositions
Size	$\|c\|$, the number of •s in c
Parameter	parts(c), the number of parts in c

Example	1 + 3 + 1 + 5 + 2 = 12 ● | ●●● | ● | ●●●●● | ●● = ●●●●●●●●●●●●
OBGF	$C(z,u) = \sum_{c \in C} z^{\|c\|} u^{\mathrm{parts}(c)}$

Construction $\qquad C = SEQ\,(\,u\ SEQ_{>0}\,(\,Z\,)\,)$

OBGF equation from symbolic method

$$C(z,u) = \frac{1}{1 - u\dfrac{z}{1-z}} = \frac{1-z}{1-z(u+1)}$$

"Horizontal" OGF for parts in a composition of N

$$c_N(u) \equiv [z^N]C(z,u) = (u+1)^N - (u+1)^{N-1}$$

Enumeration

$$c_N(1) = 2^N - 2^{N-1} = 2^{N-1}$$

Cumulated cost

$$c'_N(1) = N2^{N-1} - (N-1)2^{N-2} = (N+1)2^{N-2}$$

Average # parts in a random composition of *N*

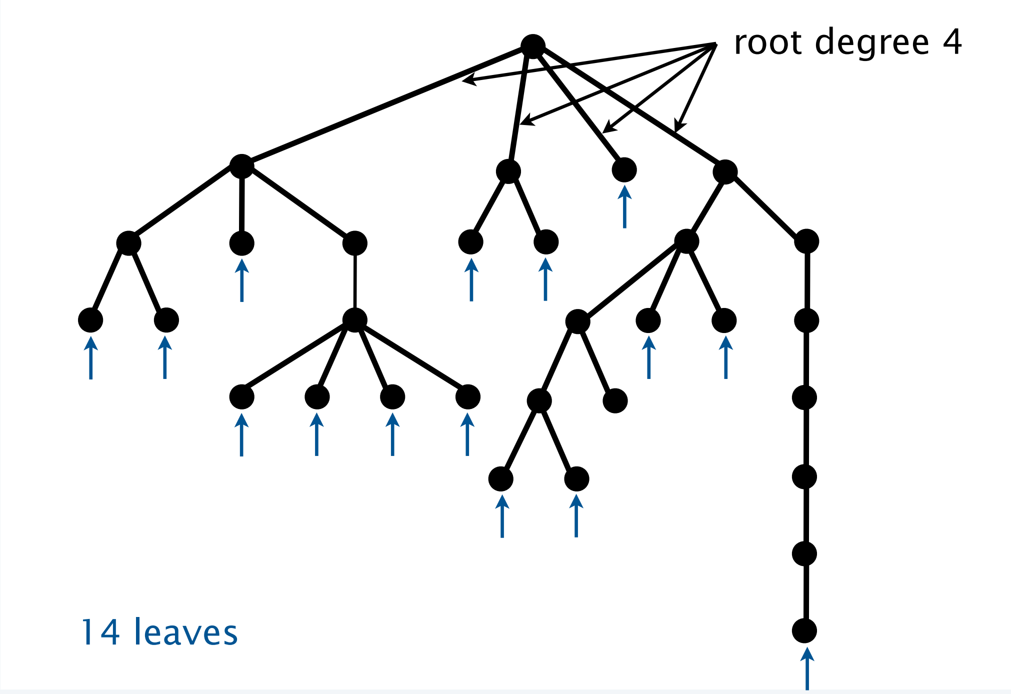
$$c'_N(1)/c_N(1) = \frac{N+1}{2}$$

✓

Tree parameters

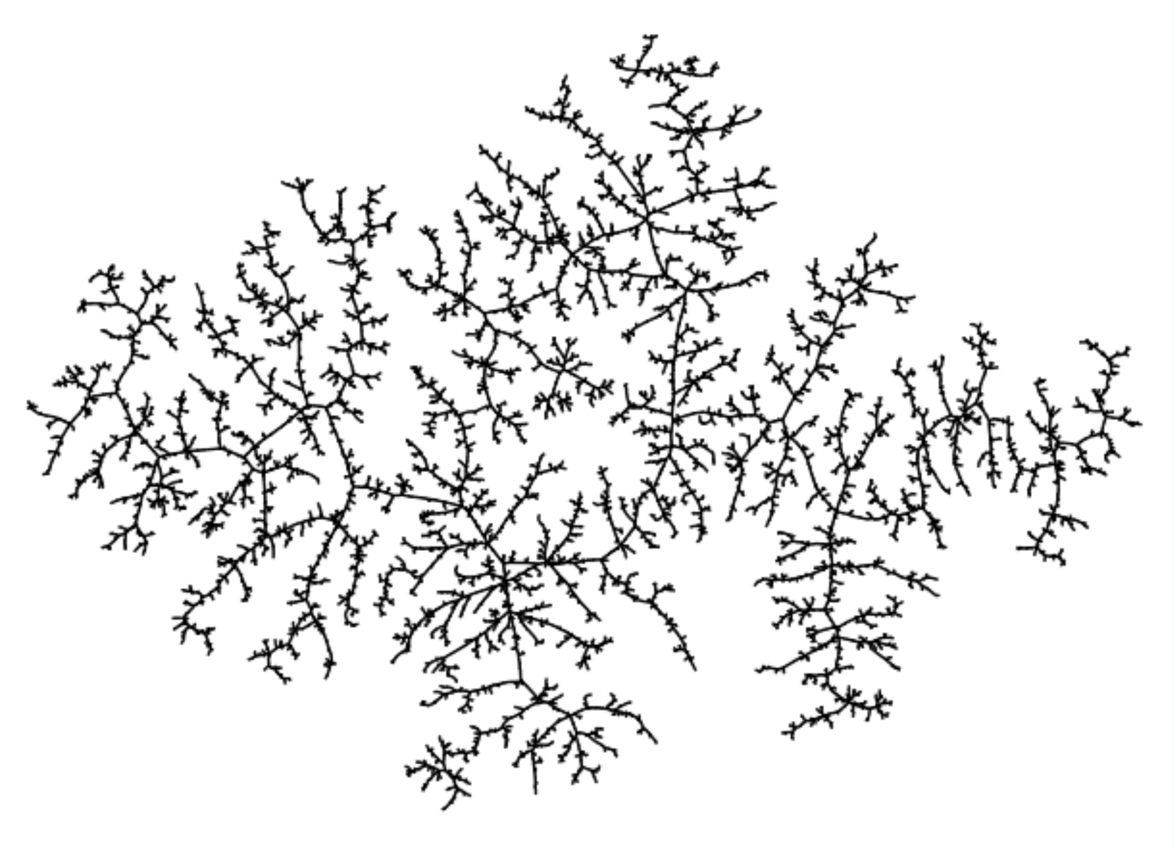
Q. What is the expected *root degree* of a random tree with N nodes ?

Q. How many *leaves* in a random tree with N nodes ?



Leaves in a random tree

Q. How many *leaves* in a random tree with N nodes ?

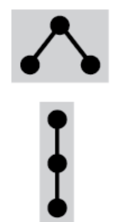


Leaves in random trees

Q. How many trees with N nodes and k leaves ?

$G^L_{10} = 1$

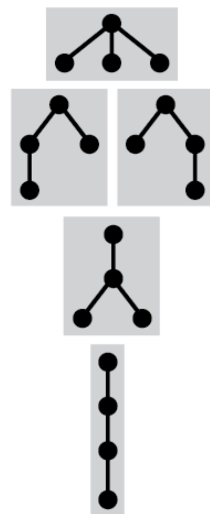
$G^L_{21} = 1$



$G^L_{31} = 1$

$G^L_{32} = 1$

cumulated cost: 3
average: 1.5

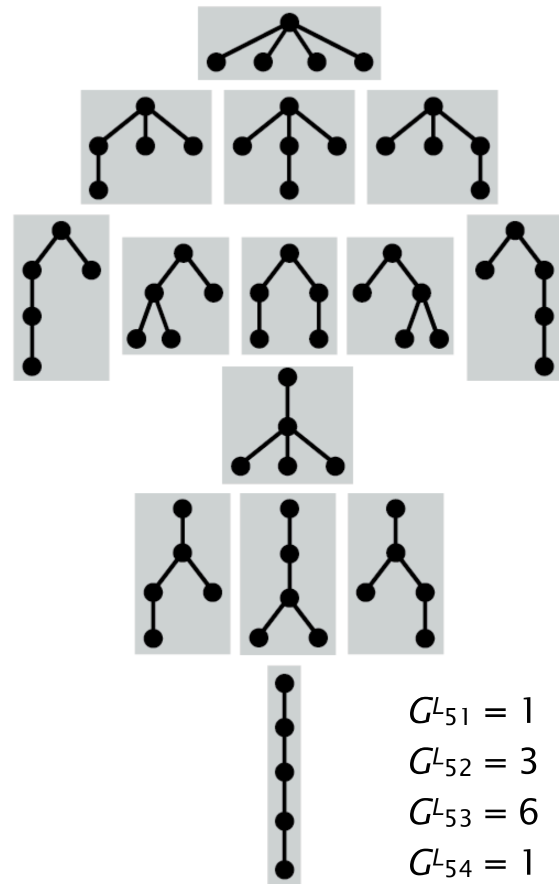


$G^L_{41} = 1$

$G^L_{42} = 3$

$G^L_{43} = 1$

cumulated cost: 10
average: 2



$G^L_{51} = 1$

$G^L_{52} = 3$

$G^L_{53} = 6$

$G^L_{54} = 1$

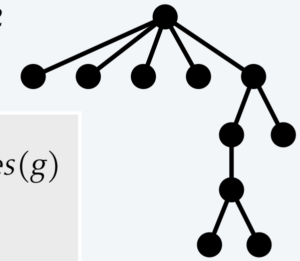
cumulated cost: 35
average: 2.5

A. $N/2$ (next slide)

Leaves in random trees

Class	*G*, the class of all ordered trees
Size	$\|g\|$, the number of •s in *g*
Parameter	*leaves*(*g*), the number of leaves in *g*

Example



OBGF	$G^L(z,u) = \sum_{g\in G} z^{\|g\|}u^{leaves(g)}$

Construction $G^L = u\,Z + Z \times SEQ_{>0}(\,G^L\,)$

OBGF equation from symbolic method $$G^L(z,u) = zu + \frac{zG^L(z,u)}{1-G^L(z,u)}$$

Enumeration OGF $$G^L(z,1) = G(z)$$

$$[z^N]G(z) = \frac{1}{N}\binom{2N-2}{N-1}$$

Cumulated cost OGF $$G_u^L(z,1) = \frac{z}{2}\Big(1+\frac{1}{\sqrt{1-4z}}\Big)$$

$$[z^N]\frac{1}{\sqrt{1-4z}} = \binom{2N}{N}$$

Average # leaves in a random tree $$\frac{[z^N]G_u^L(z,1)}{[z^N]G(z)} = \frac{N}{2} \quad \text{for } N \geq 2 \ ✓$$

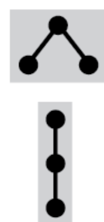
concentrated: σ_N is $O(\sqrt{N})$

Root degree in random trees

Q. How many trees with N nodes and root degree k ?

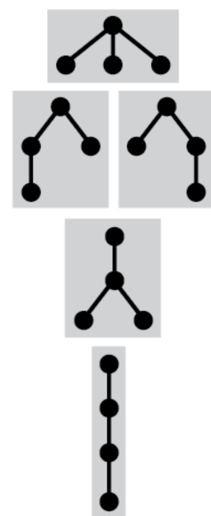
$G^D{}_{10} = 0$

$G^D{}_{21} = 1$



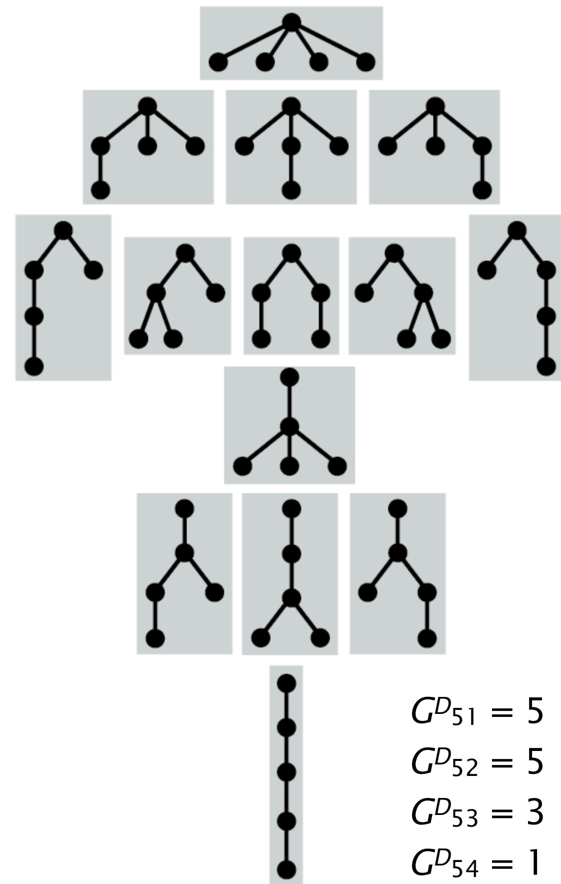
$G^D{}_{31} = 1$
$G^D{}_{32} = 1$

cumulated cost: 3
average: 1.5



$G^D{}_{41} = 2$
$G^D{}_{42} = 2$
$G^D{}_{43} = 1$

cumulated cost: 9
average: 1.8



$G^D{}_{51} = 5$
$G^D{}_{52} = 5$
$G^D{}_{53} = 3$
$G^D{}_{54} = 1$

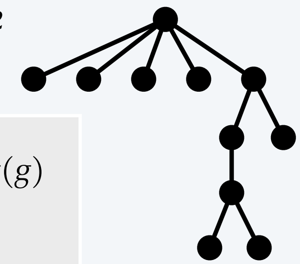
cumulated cost: 28
average: 2

A. (next slide)

Root degree in random trees

Class	*G*, the class of all ordered trees
Size	$|g|$, the number of •s in g
Parameter	*deg*(g), the degree of the root of g

Example



OBGF	$G^L(z,u) = \sum_{g\in G} z^{|g|}u^{deg(g)}$

Construction $\quad G^D = Z \times SEQ_{>0}(\, uG^D \,)$

OBGF equation from symbolic method $\quad G^D(z,u) = \dfrac{z}{1-uG(z)}$

Enumeration OGF $\quad G^D(z,1) = G(z)$

Cumulated cost OGF $\quad G_u^D(z,1) = \dfrac{zG(z)}{(1-G(z))^2} = (1-z)\dfrac{G(z)}{z} - 1$!!

Average # leaves in a random tree $\quad \dfrac{[z^N]G_u^D(z,1)}{[z^N]G(z)} = \dfrac{G_{N+1}}{G_N} - 1$

~ 3

$$\frac{\frac{1}{N+1}\binom{2N}{N}}{\frac{1}{N}\binom{2N-2}{N-1}} = \frac{2N(2N-1)N}{(N+1)NN}$$
$$= 4 - \frac{6}{N+1}$$

N	$3 - \dfrac{6}{N+1}$
1	0
2	1
3	1.5
4	1.8
5	2 ✓

Rhyming schemes

Q. How many ways to *rhyme a poem* ?

There was a small boy of Quebec	**A**	*TWO roads diverged in a yellow wood,*	**A**
Who was buried in snow to his neck	**A**	*And sorry I could not travel both*	**B**
When they said, "Are you friz?"	**B**	*And be one traveler, long I stood*	**A**
He replied, " Yes, I is —	**B**	*And looked down one as far as I could*	**A**
But we don't call this cold in Quebec!	**A**	*To where it bent in the undergrowth;*	**B**

Rhyming schemes

Q. How many ways to *rhyme an N-line poem with k rhymes* ?

A

$S_{11} = 1$

A	B
A	A

$S_{21} = 1$

$S_{22} = 1$

A	B	C
A	B	B
A	B	A
A	A	B
A	A	A

$S_{31} = 1$

$S_{32} = 3$

$S_{33} = 1$

A	B	C	D
A	B	C	C
A	B	C	B
A	B	B	C
A	B	C	A
A	B	A	C
A	A	B	C
A	A	B	B
A	B	A	B
A	B	B	A
A	B	B	B
A	B	A	A
A	A	B	A
A	A	A	B
A	A	A	A

$S_{41} = 1$

$S_{42} = 7$

$S_{43} = 6$

$S_{44} = 1$

Rhyming schemes

Class	*S*, the class of all rhyming patterns
Size	number of lines
Parameter	number of rhymes with k lines

Example	**A B C A D A B E**
OBGF	$S(z,u) = \sum_{s \in S} z^{\|s\|} u^{rhymes(s)}$

"Vertical" construction $Z_A \times SEQ(Z_A) \times Z_B \times SEQ(Z_A + Z_B) \times Z_C \times SEQ(Z_A + Z_B + Z_C) \times \ldots$

Vertical OGF

$$S_k(z) = \frac{z^k}{(1-z)(1-2z)\ldots(1-kz)}$$

"Stirling numbers of the 2nd kind " (stay tuned)

Average # *k*-rhyming patterns in an *N*-line poem

$$\sum_{N\geq 0}\sum_{k\geq 0} \left\{ {N \atop k} \right\} z^N u^k \sim \frac{k^N}{k!}$$

details omitted
(see page 63)

OBGF of Stirling numbers of the 2nd kind (partition numbers)

$$\sum_{N\geq 0}\sum_{k\geq 0}\left\{ {N \atop k} \right\} z^N u^k$$

$$= \sum_{N\geq 0} B_N(u) z^N$$

(horizontal OGF)
"Bell polynomials"

$$= \sum_{k\geq 0} \frac{z^k}{(1-z)(1-2z)\ldots(1-kz)} u^k$$

(vertical OGF)

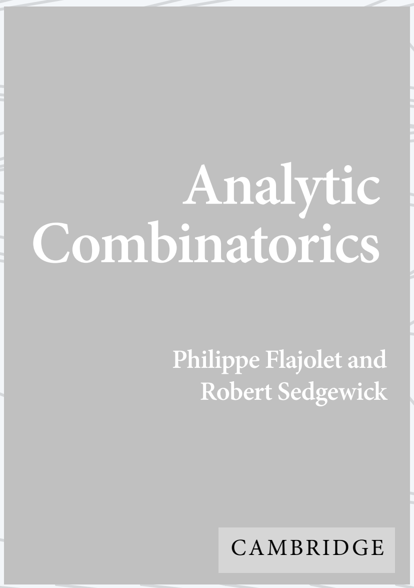
vertical OGF coefficients

$N \downarrow$ $k \rightarrow$	1	2	3	4	5	6	7
1	1						
2	1	1					
3	1	3	1				
4	1	7	6	1			
5	1	15	25	10	1		
6	1	31	90	65	15	1	
7	1	63	301	350	140	21	1

horizontal OGF coefficients

$[u^k]B_4(u)$

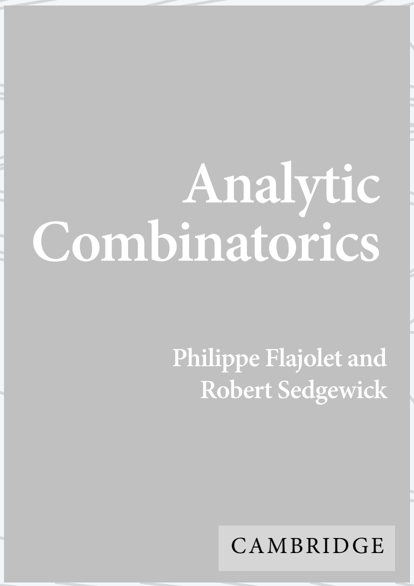
$$[z^N]\frac{z^3}{(1-z)(1-2z)(1-3z)}$$



http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- Moment calculations
- **OBGF examples**
- Labelled classes



http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- Moment calculations
- OBGF examples
- **Labelled classes**

Basic definitions (combinatorial parameters for labelled classes)

Def. A *labelled combinatorial class* is a set of *labelled* combinatorial objects and an associated size function that may have an associated parameter.

Def. The *exponential bivariate generating function* (EBGF) associated with a *labelled* class is the power series

$$A(z,u) = \sum_{a \in A} \frac{z^{|a|}}{|a|!} u^{cost(a)}$$

size function

parameter value

object name

class name

Fundamental (elementary) identity

$$A(z,u) \equiv \sum_{a \in A} \frac{z^{|a|}}{|a|!} u^{cost(a)} = \sum_{N \geq 0} \sum_{k \geq 0} \frac{A_{Nk}}{N!} z^N u^k$$

Terminology.

The variable *z marks* size
The variable *u marks* the parameter

Q. How many objects of size N with value k?

A. $A_{Nk} = N![z^N][u^k]A(z,u)$

Terminology.

BGF: bivariate GF.
MGF: multivariate GF

might add arbitrary number of markers

With the symbolic method, we specify the class *and at the same time* characterize the EBGF

The symbolic method for EBGFs (basic constructs)

Suppose that A and B are classes of unlabelled objects with *EBGFs* $A(z,u)$ and $B(z,u)$ where z marks size and u marks a parameter value. Then

operation	notation	semantics	OGF
disjoint union	$A + B$	disjoint copies of objects from A and B	$A(z,u) + B(z,u)$
labelled product	$A \star B$	ordered pairs of copies of objects, one from A and one from B	$A(z,u)B(z,u)$
sequence	$SEQ(A)$	sequences of objects from A	$\dfrac{1}{1 - A(z,u)}$

Construction immediately gives BGF equation, as for enumeration.

Extends immediately to mark multiple parameters simultaneously with MGFs.

Number of different letters in 3-words

Q. How many different letters in a 3-word of length N?

1
2
3

$W_{11} = 3$

cumulated cost: 3
average: 1.5

1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

$W_{21} = 3$
$W_{22} = 6$

cumulated cost: 15
average: 1.667

1	1	1		2	1	1		3	1	1
1	1	2		2	1	2		3	1	2
1	1	3		2	1	3		3	1	3
1	2	1		2	2	1		3	2	1
1	2	2		2	2	2		3	2	2
1	2	3		2	2	3		3	2	3
1	3	1		2	3	1		3	3	1
1	3	2		2	3	2		3	3	2
1	3	3		2	3	3		3	3	3

$W_{31} = 3$
$W_{32} = 18$
$W_{33} = 6$

cumulated cost: 57
average: 2.111

Number of different letters in M-words

Class	W_M, the class of all M-words
Size	$\|w\|$, the length of w
Parameter	*lets*(w), the # of different letters in w

Example	3 1 4 6 4 1 2 2 3 4 4 1
EBGF	$W_M(z,u) = \sum_{w \in W_M} \frac{z^{\|w\|}}{\|w\|!} u^{lets(w)}$

Construction $\quad W_M = SEQ_M(\, E + u\ SET_{>0}(\, Z\,)\,)$

EBGF equation from symbolic method $\quad W_M(u,z) = (1 + u(e^z - 1))^M$

Enumeration EGF $\quad W_M(1,z) = e^{zM}$

Cumulated cost EGF $\quad W_u(1,z) = Me^{z(M-1)}(e^z - 1) = Me^{zM} - Me^{z(M-1)}$

Average # different letters in a random *M*-word of length *N*

$$\mu_N = \frac{N![z^N]W_u(1,z)}{N![z^N]W(1,z)} = M(1 - (1 - \frac{1}{M})^N)$$

N	μ_N
1	1
2	1.667
3	2.111

✓

Number of different letters with a given frequency in M-words

Class	W_M, the class of all M-words
Size	$\|w\|$, the length of w
Parameter	$f_k(w)$, the # of different letters in w

Example	3 1 4 6 4 1 2 2 3 4 4 1
EBGF	$W_M(z,u) = \sum_{w \in W_M} \frac{z^{\|w\|}}{\|w\|!} u^{f_k(w)}$

Construction

$$W_M = SEQ_M\,(\,SET_{\neq k}(\,Z\,) + u\ SET_k(\,Z\,)\,)$$

EBGF equation from symbolic method

$$W_M(u,z) = \Big(e^z + (u-1)\frac{z^k}{k!}\Big)^M$$

Enumeration EGF

$$W_M(1,z) = e^{zM}$$

Cumulated cost EGF

$$W_u(1,z) = Me^{z(M-1)}\frac{z^k}{k!}$$

Average # letters that appear *k* times in a random *M*-word of length *N*

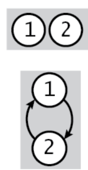
$$\frac{N![z^N]W_u(1,z)}{N![z^N]W(1,z)} = M\binom{N}{k}\Big(\frac{1}{M}\Big)^k\Big(1-\frac{1}{M}\Big)^{N-k}$$

occupancy distribution ✓

Cycles in random permutations

Q. How many permutations of N elements *have k cycles* ?

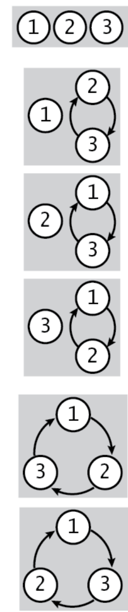
$P_{11} = 1$



$P_{21} = 1$

$P_{22} = 1$

cumulated cost: 3
average: 1.5

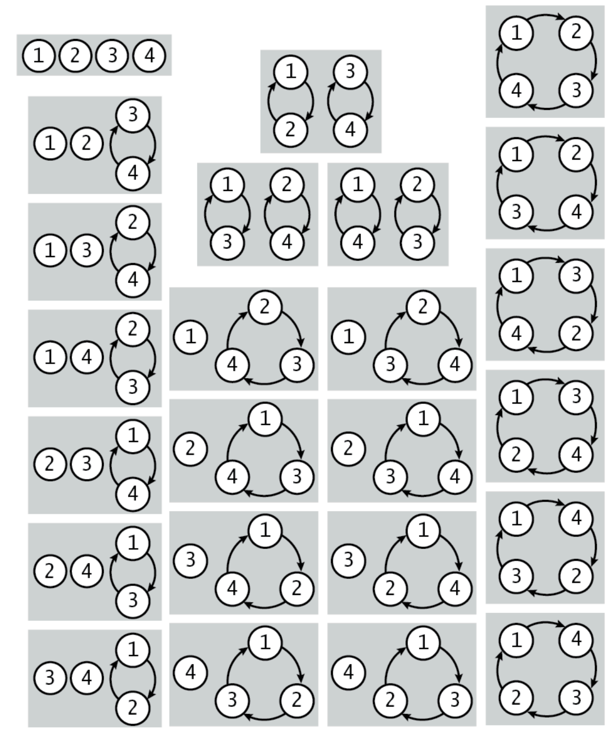


$P_{31} = 2$

$P_{32} = 3$

$P_{33} = 1$

cumulated cost: 11
average: 1.8333



$P_{41} = 6$

$P_{42} = 11$

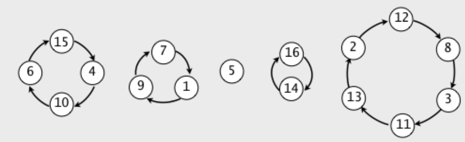
$P_{43} = 6$

$P_{44} = 1$

cumulated cost: 50
average: 2.0833

Cycles in random permutations

Class	P, the class of all permutations
Size	$\|p\|$, the length of p
Parameter	$cyc(p)$, the number of cycles in p

Example	
EBGF	$P(z,u) = \sum_{p \in P} \frac{z^{\|p\|}}{\|p\|!} u^{cyc(p)}$

Construction $\qquad P = SET(\,u\, CYC(\,Z\,)\,)$

EBGF equation from symbolic method $\qquad P(z,u) = e^{u \ln \frac{1}{1-z}} = (1-z)^{-u}$

Enumeration EGF $\qquad P(z,1) = \frac{1}{1-z}$

Cumulated cost EGF $\qquad P_u(z,1) = \frac{1}{1-z} \ln \frac{1}{1-z}$

Average # cycles in a random permutation $\qquad \frac{N![z^N]P_u(z,1)}{N![z^N]P(z,1)} = H_N$

concentrated: σ_N is $O(\sqrt{\log N})$

N	H_N
1	1
2	1.5
3	1.833
4	2.083

✓

EBGF of Stirling numbers of the 1st kind (cycle numbers)

$$\sum_{N\geq 0}\sum_{k\geq 0}\begin{bmatrix}N\\k\end{bmatrix}\frac{z^N}{N!}u^k$$

$$= \sum_{N\geq 0} u(u+1)\dots(u+N-1)\frac{z^N}{N!} \quad \text{(horizontal EGF)}$$

$$= \sum_{k\geq 0}\frac{1}{k!}\Big(\ln\frac{1}{1-z}\Big)^k u^k \quad \text{(vertical EGF)}$$

$$= \frac{1}{(1-z)^u} \quad \text{(EBGF)}$$

$N \backslash k$	1	2	3	4	5	6	7
1	1						
2	1	1					
3	2	3	1				
4	6	11	6	1			
5	24	50	35	10	1		
6	120	274	225	85	15	1	
7	720	1764	1624	735	175	21	1

vertical OGF coefficients

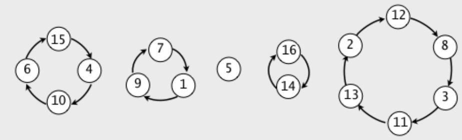
horizontal OGF coefficients

$[u^k]u(u+1)(u+2)(u+3) \longrightarrow$ (row 4)

$N![z^N]\frac{1}{3!}\Big(\ln\frac{1}{1-z}\Big)^3$ (column 3)

Number of cycles of a given length in random permutations

Class	P, the class of all permutations
Size	$\lvert p \rvert$, the length of p
Parameter	$cyc_r(p)$, # of cycles of length r in p

Example	
EBGF	$P(z,u) = \sum_{p \in P} \frac{z^{\lvert p \rvert}}{\lvert p \rvert !} u^{cyc_r(p)}$

Construction $\quad P = SET(\, CYC_{\neq r}(\, Z\,) + u\; CYC_{=r}(\, Z\,)\,)$

EBGF equation from symbolic method

$$P(z,u) = e^{\ln \frac{1}{1-z} - \frac{z^r}{r} + \frac{uz^r}{r}} = \frac{e^{(u-1)z^r/r}}{1-z}$$

Enumeration EGF

$$P(z,1) = \frac{1}{1-z}$$

Cumulated cost EGF

$$P_u(z,1) = \frac{z^r}{r}\frac{1}{1-z}$$

Average # r-cycles in a random permutation

$$\frac{N![z^N]P_u(z,1)}{N![z^N]P(z,1)} = \frac{1}{r}$$

Set partitions

Q. How many ways to *partition* a set of size of N ?

```
{1}
```

$S_1 = 1$

```
{1} {2}
{1 2}
```

$S_2 = 2$

```
{1} {2} {3}
{1} {2 3}
{2} {1 3}
{3} {1 2}
{1} {2} {3}
```

$S_3 = 5$

```
{1} {2} {3} {4}
{1} {2 3 4}
{2} {1 3 4}
{3} {1 2 4}
{4} {1 2 3}
{1 2} {3} {4}
{1 3} {2} {4}
{1 4} {2} {3}
{2 3} {1} {4}
{2 4} {1} {3}
{3 4} {1} {2}
{1 2} {3 4}
{1 3} {2 4}
{1 4} {2 3}
{1 2 3 4}
```

$S_4 = 15$

Set partitions

Q. How many ways to partition a set of size of N *into k subsets*?

{1}

$S_{11} = 1$

{1} {2}
{1 2}

$S_{21} = 1$
$S_{22} = 1$

cumulated cost: 3
average: 1.5

{1} {2} {3}
{1} {2 3}
{2} {1 3}
{3} {1 2}
{1 2 3}

$S_{31} = 1$
$S_{32} = 3$
$S_{33} = 1$

cumulated cost: 11
average: 2

{1} {2} {3} {4}
{1} {2 3 4}
{2} {1 3 4}
{3} {1 2 4}
{4} {1 2 3}
{1 2} {3} {4}
{1 3} {2} {4}
{1 4} {2} {3}
{2 3} {1} {4}
{2 4} {1} {3}
{3 4} {1} {2}
{1 2} {3 4}
{1 3} {2 4}
{1 4} {2 3}
{1 2 3 4}

$S_{41} = 1$
$S_{42} = 7$
$S_{43} = 6$
$S_{44} = 1$

cumulated cost: 37
average: 2.466

Number of subsets in set partitions

Class	*S*, the class of all set partitions
Size	size of the set
Parameter	number of subsets in the partition

Example	{1} {2 5 6} {3 7 8} {4}
EBGF	$S(z,u) = \sum_{s\in S} \frac{z^{\lvert s\rvert}}{\lvert s\rvert!} u^{subsets(s)}$

Construction $S = SET(\, u\ SET_{>0}(\, Z\,)\,)$

EBGF equation from symbolic method $$S(z,u) = e^{u(e^z - 1)}$$

Enumeration EGF $$S(z,1) = e^{e^z - 1}$$

Cumulated cost EGF $$S_u(z,1) = (e^z - 1)e^{(e^z - 1)}$$

Average # subsets in a random set partition $$\frac{N![z^N]S_u(z,1)}{N![z^N]S(z,1)}$$ ← need complex asymptotics (stay tuned)

EBGF of Stirling numbers of the 2nd kind (partition numbers)

$$\sum_{N\geq 0}\sum_{k\geq 0}\left\{ {N \atop k} \right\}\frac{z^N}{N!}u^k$$

$$= \sum_{N\geq 0} B_N(u)\frac{z^N}{N!}$$ (horizontal EGF) "Bell polynomials"

$$= \sum_{k\geq 0}(e^z - 1)^k\frac{u^k}{k!}$$ (vertical EGF)

$$= e^{u(e^z - 1)}$$ (EBGF)

vertical EGF coefficients

horizontal EGF coefficients $\longrightarrow$ row 4

$N \backslash k$	1	2	3	4	5	6	7
1	1						
2	1	1					
3	1	3	1				
4	1	7	6	1	$\longleftarrow [u^k]B_4(u)$		
5	1	15	25	10	1		
6	1	31	90	65	15	1	
7	1	63	301	350	140	21	1

$$N![z^N]\frac{1}{3!}(e^z - 1)^3$$

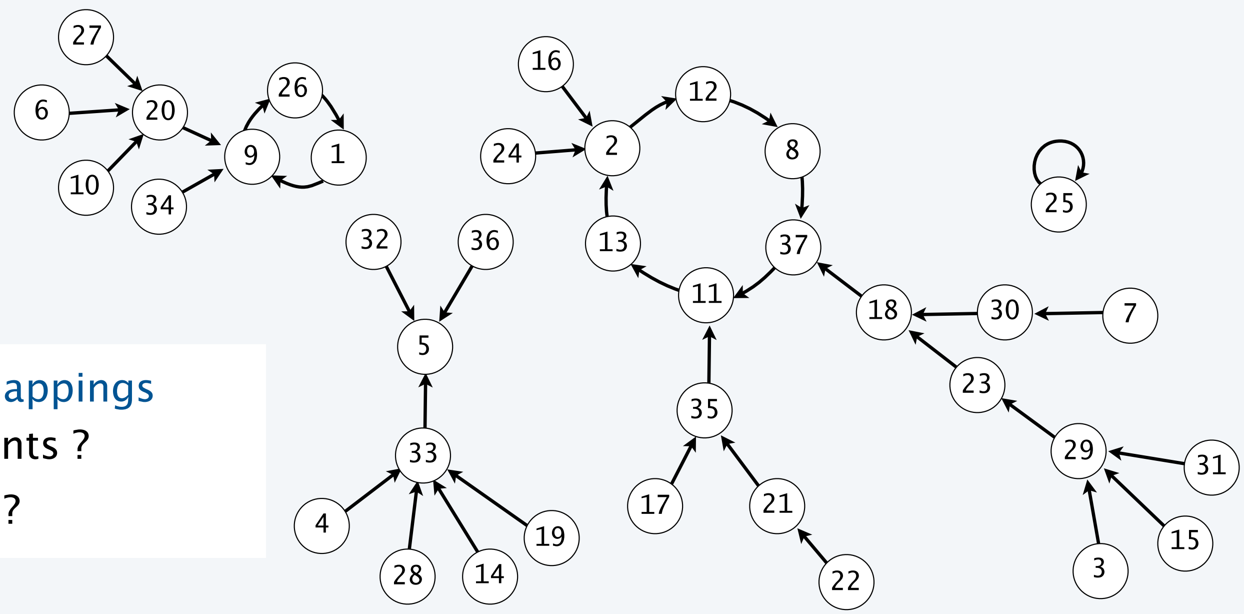
Mappings

Def. A *mapping* is a function from the set of integers from 1 to N onto itself.

Example

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
9	12	29	33	5	20	30	37	26	20	13	8	2	33	29	2	35	37	33	9	35	21	18	2	25	1	20	33	23	18	29	5	5	9	11	5	11

Every mapping corresponds to a digraph
- N vertices, N edges
- Outdegrees: all 1
- Indegrees: between 0 and N



Natural questions about random mappings
- How many connected components ?
- How many nodes are on cycles ?

Mapping EGFs (see lecture on EGFs)

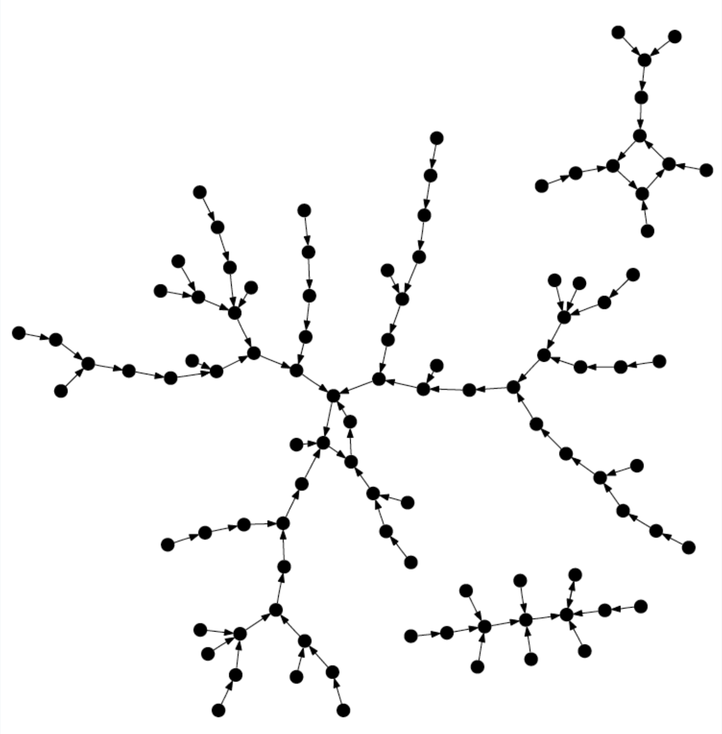
Combinatorial class	C, the class of Cayley trees	⟵ labelled, rooted, unordered
Construction	$C = Z \star (SET(C))$	⟵ "a tree is a root connected to a set of trees"
EGF equation	$C(z) = ze^{C(z)}$	

Combinatorial class	Y, the class of mapping components	
Construction	$Y = CYC(C)$	⟵ "a mapping component is a cycle of trees"
EGF equation	$Y(z) = \ln \dfrac{1}{1 - C(z)}$	

Combinatorial class	C, the class of Cayley trees	
Construction	$M = SET(CYC(C))$	⟵ "a mapping is a set of components"
EGF equation	$M(z) = \exp\left(\ln \dfrac{1}{1 - C(z)}\right) = \dfrac{1}{1 - C(z)}$	

Mapping parameters

are available via EBGFs based on the same constructions



Ex 1. Number of components

Construction $M = SET(uCYC(C))$

EGF equation $M(z) = \exp\left(u \ln \frac{1}{1 - C(z)}\right) = \frac{1}{(1 - C(z))^u}$

Ex 2. Number of trees (nodes on cycles)

Construction $M = SET(CYC(uC))$

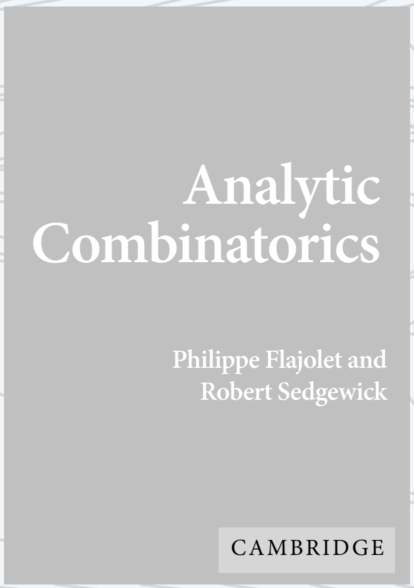
EGF equation $M(z) = \exp\left(\ln \frac{1}{1 - uC(z)}\right) = \frac{1}{1 - uC(z)}$

Q. Moments? Coefficients? Other parameters?

A. Stay tuned for general theorems from complex aysmptotics.

"We shall now stop supplying examples that could be multiplied ad libitum, since such calculations greatly simplify when interpreted in the light of asymptotic analysis"

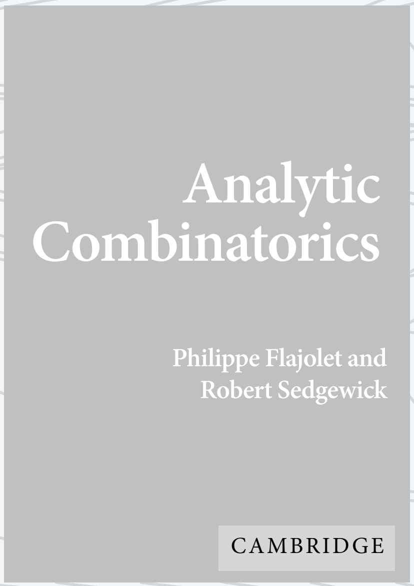
— Philippe Flajolet, 2007



http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- Moment calculations
- OBGF examples
- **Labelled classes**



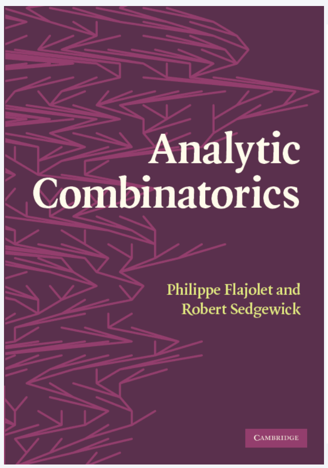
http://ac.cs.princeton.edu

3. Combinatorial parameters and MGFs

- Basics
- Moment calculations
- OBGF examples
- Labelled classes
- **Exercises**

Note III.17

Leaves in Cayley trees



▷ **III.17.** *Leaves and node-degree profile in Cayley trees.* For Cayley trees, the bivariate EGF with u marking the number of leaves is the solution to

$$T(z,u) = uz + z(e^{T(z,u)} - 1).$$

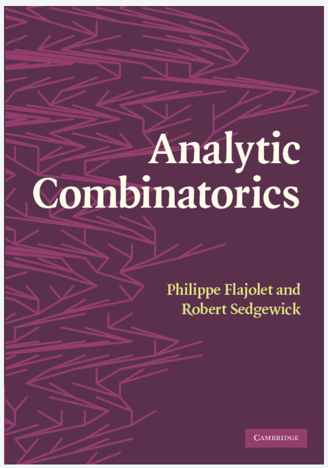
(By Lagrange inversion, the distribution is expressible in terms of Stirling partition numbers.) The mean number of leaves in a random Cayley tree is asymptotic to ne^{-1}. More generally, the mean number of nodes of outdegree k in a random Cayley tree of size n is asymptotic to

$$n \cdot e^{-1} \frac{1}{k!}.$$

Degrees are thus approximately described by a Poisson law of rate 1. ◁

Note III.21

After Bhaskara Acharya



▷ **III.21.** *After Bhaskara Acharya (circa 1150*AD). Consider all the numbers formed in decimal with digit 1 used once, with digit 2 used twice,. . . , with digit 9 used nine times. Such numbers all have 45 digits. Compute their sum S and discover, much to your amazement that S equals

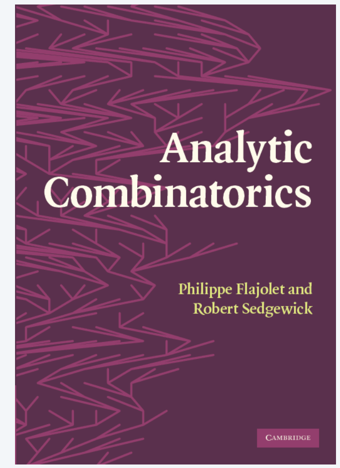
$$45875559600006153219084769286\underline{3999999999999999}541244403999938467809152307136000000.$$

This number has a long run of nines (and further nines are hidden!). Is there a simple explanation? This exercise is inspired by the Indian mathematician Bhaskara Acharya who discovered multinomial coefficients near 1150AD. ◁

4587555960000615321908476928639999999999999999541244403999938467809152307136000000

Assignments

1. Read pages 151-219 in text.



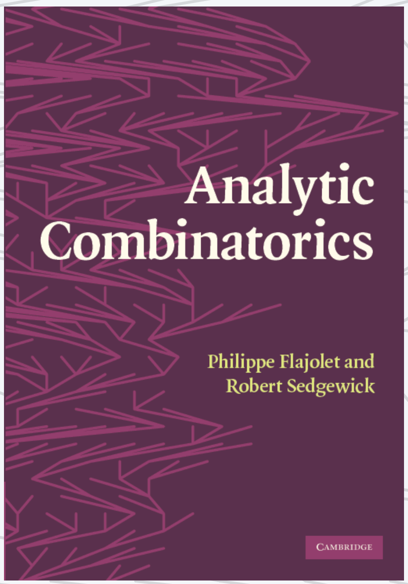
2. Write up solutions to Notes III.17 and III.21.

3. Programming exercise.

Program III.1. Write a program that generates 1000 random permutations of size N for $N = 10^3$, 10^4, ... (going as far as you can) and plots the distribution of the number of cycles, validating that the mean is concentrated at H_N.

ANALYTIC COMBINATORICS

PART TWO



http://ac.cs.princeton.edu

3. Combinatorial Parameters and MGFs