Nablus2014 CIMPA Summer School

Pierre Nicodeme, Nicolas Pouyanne, Brigitte Chauvin, Jeremie Lumbroso, Basile Morcrette, Cécile Mailler

CIMPA Summer School **PROCEEDINGS**

Analysis of Random Structures

An Najah University, Nablus, Palestine
August 18–28, 2014

The school is based upon 7 courses of 6 hours

- 3 courses upon approaches of Analytic Combinatorics
- 2 courses upon Probabilistic Approaches
- 1 course using Analytic and Probabilistic Approaches
- 1 course upon Random Graphs

Organising Committee:

Pierre Nicodème	University Paris 13
Naji Qatanani	An Najah University

Speakers:

Cyril Banderier	University Paris 13
Brigitte Chauvin	University of Versailles
Cécile Mailler	University of Bath
Pierre Nicodème	University Paris 13
Nicolas Pouyanne	University of Versailles
Subhi Ruzieh	An Najah University
Sarah Eugene	Inria, Paris

Local organisation: An Najah University

Sponsors

Image credits
photo: Dome of the Rock of Jerusalem, user ghardman, Mar 22, 2007, on stack.xchng;
maths: a random exactly sampled directed animal on the square lattice, Axel Bacher, PhD Thesis, Oct 31, 2011.

Nablus 2014 Summer School, Palestine - Analysis of Random Structures

Foreword

Analysis of Random Structures, as studied by the world-wide network AofA (Analysis of Algorithms) and by the European ALEA network, relies on the interplay between analytic and probabilistic approaches. Philippe Flajolet (1948-2011) played a fundamental and inspiring role in the development of these methods and their scientific communities.

The Nablus 2014 CIMPA summer school was a unique opportunity to introduce both the analytic and the probabilistic approaches to the Palestinian students.

Contents

– Pierre NICODÈME,	A Glimpse at Analytic Combinatorics	1
– Nicolas POUYANNE,	A Glimpse at Urn Models	3
– Brigitte CHAUVIN,	A Note on Conditional Expectation	5
– Jérémie LUMBROSO and Basile MORCRETTE,	A Gentle Introduction to Analytic Combinatorics [a]	7
– Cécile MAILLER,	Markov Chains and Martingales applied to the analysis of random structures	37
– Nicolas POUYANNE,	Pólya Urn Models - analytic and probabilistic approaches (Slides)	65
– Brigitte CHAUVIN,	Random Trees and Probabilities	89
– Pierre NICODÈME,	Automata and Motif Statistics	113

[a]This course has been presented by Pierre Nicodème with the supplementary help of a set of slides of Conrado Martínez, Polytechnic University of Catalonia.

Contributors adresses [a]

- Brigitte Chauvin, LMV, University of Versailles
- Jérémie Lumbroso, Department of Computer Science, Princeton
- Cécile Mailler, Department of Mathematical Sciences, University of Bath
- Basile Morcrette, Lycée Janson de Sailly
- Pierre Nicodème, LIPN, University Paris13
- Nicolas Pouyanne, LMV, University of Versailles

[a]Sarah Eugene, Inria-Rocquencourt, `sarah.eugene@inria.fr`, monitored the exercises during the School.

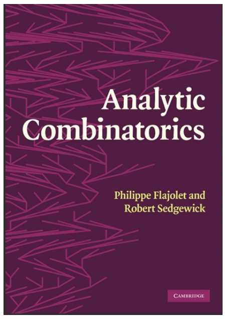
Acknowledgements

The organizers of the school thank the department of Mathematics of An Najah University of Nablus for its material support. They are also grateful for the high financial support of the laboratory LIPN of University Paris13. We mention in this regards more specifically Laure Petrucci, Director of the LIPN, Nathalie Tavares, administrative assistant, Andrea Sportiello, CNRS researcher, and Michael Fortier, system engineer and administrator.

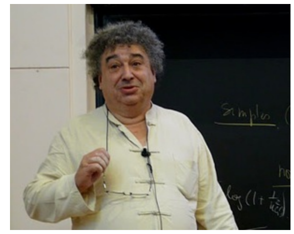
Proceedings editor: Pierre Nicodème

Summer School Nablus 2014, Palestine - Analysis of Random Structures
Contact: Pierre Nicodème, LIPN, http://www-lipn.univ-paris13.fr/CALIN/members/nicodeme

---



"Analytic combinatorics aims to enable precise quantitative predictions of the properties of large combinatorial structures. ..."



Amazon   As pdf   See also Philippe Flajolet's lectures and courses

"... The theory has emerged over recent decades as essential both for the analysis of algorithms and for the scientific models in many disciplines, including probability theory, statistical physics, computational biology and information theory. With a careful combination of symbolic enumeration methods and complex analysis, drawing heavily on generating functions, results of sweeping generality emerge that can be applied to fundamental structures such as permutations, sequences, strings, walks, trees, graphs and maps."
Foreword to "Analytic Combinatorics", Flajolet-Sedgewick 2009, Cambridge University Press.

Philippe Flajolet (1948-2011) laid the foundations of Analytic Combinatorics and extensively developed the methods and techniques used in this field.

Examples

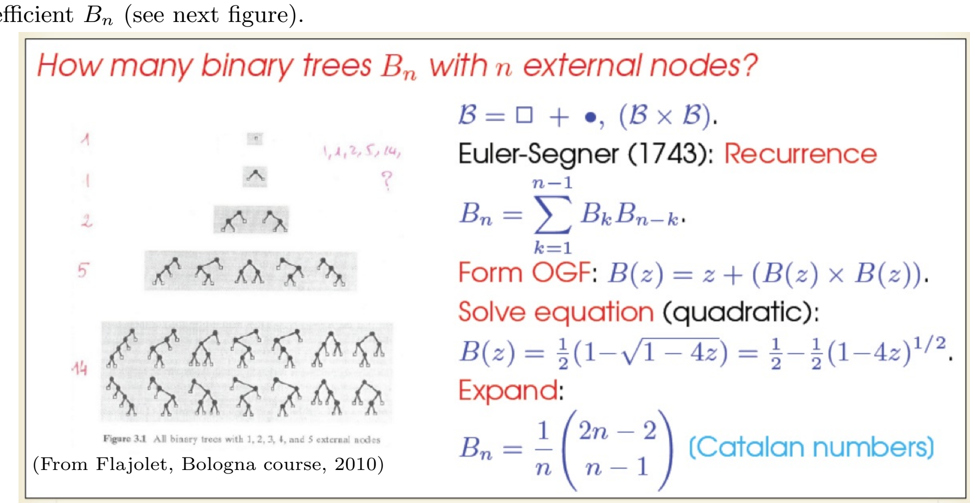
Binary trees. If you ask to a five or six years old child to draw binary trees with 1, 2, 3, 4, and 5 external nodes, and ask him about how many (different) ones there are, he will tell you the sequence (provided he or she does not get tired)

$$1, 1, 2, 5, 14 \ldots$$

Counting is also natural for mathematicians. Considering the sequence (B_n) enumerating binary trees and its OGF (ordinary generating function) $B(z)$, we have

$$(B_n) = (B_1, B_2, B_3, B_4, \ldots) = (1, 1, 2, 5, 14, \ldots) \quad \text{and} \quad B(z) = \sum_{n\geq 1} B_n z^n.$$

Now, if there are more than one external node in a binary tree, removing the root gives two subtrees that are equivalent (from a counting point of view) to any binary tree: there is a recursive decomposition that translates to a functional equation verified by the generating function $B(z)$, from which it is possible to extract the n-th Taylor coefficient B_n (see next figure).



(From Flajolet, Bologna course, 2010)

The example of binary trees is typical of the process of Analytic Combinatorics which works as follows.

1. Construct a symbolic equation on the combinatorial classes occurring in your problem (in the case of binary tree, these are the class $\mathcal{B}$ and the class $\square$ representing a leaf with OGF z).

2. Translate the symbolic equation into a functional equation on generating functions.

3. Extract the Taylor coefficient of interest; asymptotically, this is often done by complex analysis and Cauchy integrals or variants of these.

The counting is much more general than univariate counting as we see next.

Cycles in permutations. The *cycle construction* puts in equivalence classes sequences taken up to a circular shift; considering the permutations of the symmetric group $\mathfrak{S}_4$ of size 4!, we have

$$1234 \equiv 2341 \equiv 3412 \equiv 4123,\ 1243 \equiv 2341 \equiv \dots,\ 1324 \equiv \dots,\ 1342 \equiv \dots,\ 1423 \equiv \dots,\ 1432 \equiv \dots.$$

If $C_n = n!/n$ is the number of classes of the symmetric group $\mathfrak{S}_n$ quotiented by the cycle construction, the corresponding *exponential* generating function verifies

$$C(z) = \sum_{n\geq 0} \frac{C_n z^n}{n!} = \sum_{n\geq 0} \frac{z^n}{n} = \log\left(\frac{1}{1-z}\right).$$

Considering any permutation, we can decompose it as a set of cycles, as seen in the following example

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 \\ 11 & 12 & 13 & 17 & 10 & 15 & 14 & 9 & 3 & 4 & 6 & 2 & 7 & 8 & 1 & 5 & 16 \end{pmatrix},$$

one of the cycle being $4 \to 17 \to 16 \to 5 \to 10 \to 4$.

If $\mathcal{C}$ is a generic cycle, and $\mathcal{P}$ a generic permutation, the decompositions is written symbolically as

$$\mathcal{P} = \{\epsilon\} + \mathcal{C} + (\mathcal{C} \star \mathcal{C}) + (\mathcal{C} \star \mathcal{C} \star \mathcal{C}) + \dots \qquad \text{(Permutation = Set of Cycles)}.$$

As *Set* $\rightsquigarrow$ exp and *Cycle* $\rightsquigarrow$ log, using again exponential generating functions that count labelled objects, and moreover a variable u that counts the number of cycles, we have (being very sketchy)

$$P(z,u) = \sum_{\substack{n\geq 0\\ u\leq n}} \begin{bmatrix} n \\ k \end{bmatrix} u^k z^n = 1 + uC(z) + \frac{1}{2!}u^2C^2(z) + \frac{1}{3!}u^3C^3(z) + \dots = \exp\left(u \log\left(\frac{1}{1-z}\right)\right) = (1-z)^{-u},$$

where $\begin{bmatrix} n \\ k \end{bmatrix}$ is the Stirling cycle number that counts the number of permutations of size n with k cycles.

We obtain by the binary theorem

$$[z^n](1-z)^{-u} = \sum_{k\leq n} \begin{bmatrix} n \\ k \end{bmatrix} u^k = u(u+1)(u+2)\dots(u+n-1),$$

and, by logarithmic differentiation, the expected number of cycles $\mu_n = \sum_k \frac{k}{n!} \begin{bmatrix} n \\ k \end{bmatrix}$ in a random permutation of size n is the n-th harmonic number,

$$\mu_n = H_n \equiv 1 + \frac{1}{2} + \dots + \frac{1}{n} \qquad (\rightsquigarrow \mu_{100} \equiv H_{100} = 5.18738).$$

Second moment follows easily, and an asymptotic method known as *quasi-powers theorem* leads to a limiting Gaussian law. (There are equivalent probabilistic approaches.)

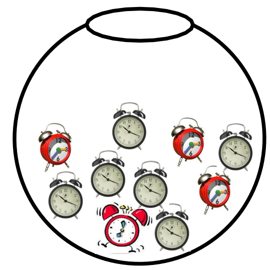
What can you learn from Analytic Combinatorics?

The projected courses will aim providing a thorough introduction to Flajolet-Sedgewick book "Analytic Combinatorics"; an additional course will be related to the Boltzmann random generation of objects. If you are a *mathematician* or a *physicist*, you cannot avoid being touched by the beauty of symbolic structures and by relatively simple mathematical concepts that lead to deep results with "real life" applications. If you are a *computer scientist* you will learn evaluating combinatorial structures that have algorithmic counterparts; *i.e* the (generalized) birthday paradox provides an analysis of collisions in data hashing.

Summer School Project for 2014 in Palestine - Random structures, Analytic and Probabilistic Approaches
Contact: Pierre Nicodème, CNRS, http://www-lipn.univ-paris13.fr/CALIN/members/nicodeme

Random structures: a probabilistic approach

Together with analytic combinatorics, methods coming from modern probability theory provide natural tools to study random structures. Being often of different nature, results from both complementary points of view enrich one another.



Example

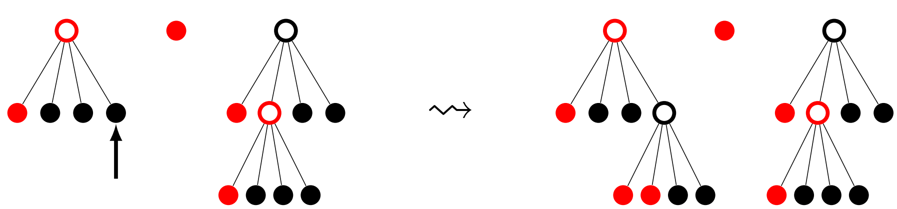
Pólya urns provide a rich model for many situations in algorithmics. In this model, one considers an urn that contains **red** and **black** balls (this can be generalized to any finite number of colors). One starts with an initial configuration. At any step of time, one chooses one ball at random in the urn, checks its color and puts it back into the urn. Depending on its color, one adds new balls of different colors according to some fixed replacement rule. The random process is defined by iterating this procedure.

Take for instance the urn process having $\begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix}$ as replacement matrix. This means that when a red ball is drawn, it is placed back into the urn together with 3 black ones; when one draws a black ball, one adds 2 red balls and 1 black one.

The composition sequence (*i.e.* the respective numbers of red and black balls it contains) of a Pólya urn is a Markov chain. This follows from the fact that the random composition at a given time depends only on the probability distribution of the preceding composition. This is the so-called ***forward*** point of view of the *growing* random structure that implies immediately, for example, hat the urn contains asymptotically 40% of red balls, with probability 1.

The forward point of view leads to represent all successive configuration in one global object: the *random process*, giving access to powerful probabilistic tools like
- *martingales*, after suitable rescaling of the urn process. Most of limit theorems come from this beautiful theory;
- *embedding in continuous time*, illustrated in our example by the underlying tree structure of the urn process as follows.

One can usefully represent the evolution of the urn by the growing of a tree. The leafs are colored red and black and represent the balls in the urn. Drawing a ball amounts to choosing a leaf. The corresponding added balls are represented as daughter leafs. In the figure below, one chooses the black pointed leaf in the tree on the left; one obtains the new tree drawn on the right.



In the discrete time urn, the subtrees are *not* stochastically independent. Embedding the process in continuous time consists in making the time intervals between two drawings random. When this random times are exponentially distributed, the subtrees of the continuous time urn process become independent. The resulting process is well-known by the probabilists: it is a branching process, giving rise to – Gaussian or not – limit laws.

After embedding in continuous time, the gained independence allows us to use the *recursive* properties of the random structure through the *divide and conquer* principle. This is to the ***backward*** point of view. Applied to generating functions, it is the base tool for analytic combinatorics methods. In the probabilistic domain, it translates the recursivity in terms of distributional equations on random variables, often of the type

$$W \stackrel{\mathcal{L}}{=} \sum A_i W^{(i)}$$

where the A_i are known random variables, the $W^{(i)}$ are independent copies of W, independent of the A_i as well. By means of Fourier analysis for instance, one derives properties of the limit distributional behavior of the random structure.

A note on conditional expectation [1]

The conditional probability of one event A with respect to another event B of non-zero probability is known as: $\mathbb{P}(A|B) := \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. If we consider a random variable X with (continuous) density, and we want to condition with respect to a value taken by X, it is not possible to apply the preceding formula since the event $\{X = x\}$ has null probability. With X and Y two random variables, the probability of Y conditioned to X may be viewed as taking a couple (X, Y), assuming known the value of X and doing a "prediction" of Y, *i.e* finding a function of X that approximates as well as possible Y. This is expressed in the following as $\mathbb{E}(Y|\mathcal{B}(X))$ where $\mathcal{B}(X)$ is the σ-algebra generated by X.

Mathematically, the conditional expectation of Y with respect to X is defined as the orthogonal projection of Y in the Hilbert space of square-integrable functions onto the space of $\mathcal{B}(X)$-measurable functions (see below).

Definition 1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let also $L^2(\mathcal{A})$ be the space of real-valued fonctions that are measurable on $(\Omega, \mathcal{A})$ and square-integrable with respect to the measure $\mathbb{P}$. It is a Hilbert space for the scalar product $\langle f, g\rangle = \int_\Omega fg \; d\mathbb{P}$.*
let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$ and let $L^2(\mathcal{B})$ be the space of real-valued fonctions that are measurable with respect to $\mathcal{B}$ and square-integrable. The orthogonal projection of $L^2(\mathcal{A})$ on $L^2(\mathcal{B})$ is called conditional expectation *with respect to $\mathcal{B}$ (or knowing $\mathcal{B}$).*

Notation. The conditional expectation of X knowing $\mathcal{B}$ is noted $\mathbb{E}^{\mathcal{B}}(X)$ or $\mathbb{E}(X|\mathcal{B})$.

A frequent particular case occurs when the σ-algebra $\mathcal{B}$ is one of the σ-algebras of a filtration $(\mathcal{F}_n)_{n\geq 0}$. Typically, when one considers a discrete-time process $(X_n)_{n\geq 0}$, and when $\mathcal{F}_n$ is the σ-algebra generated by the X_p for $p \leq n$. The σ-algebra $\mathcal{F}_n$ is called the σ-algebra of the past before n and $\mathbb{E}(X \mid \mathcal{F}_n)$ or $\mathbb{E}^{\mathcal{F}_n}(X)$ denotes the conditioning of X by the past before n.

Since L^2 is dense in L^1 for a finite positive measure, the last notion can be extended to all integrable functions. This leads to the following characterization that is in practice more useful that the definition:

Proposition 1 (characterization of the conditional expectation)
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathcal{B} \subset \mathcal{A}$. Then $\mathbb{E}(X|\mathcal{B})$ is the unique random variable such that:
- $\mathbb{E}(X|\mathcal{B})$ *is $\mathcal{B}$-measurable;*
- *for every $\mathcal{B}$-mesurable and bounded random variable Y, we have $\mathbb{E}(YX) = \mathbb{E}(Y\mathbb{E}(X|\mathcal{B}))$.*

It is necessary to remark that $\mathbb{E}(X|\mathcal{B})$ is a *random variable* $\mathcal{B}$-measurable; this is generally speaking not the case for a constant like $\mathbb{E}(X)$. The conditional expectation with respect to the trivial σ-algebra reduced to $\{\emptyset, \Omega\}$ is the usual simple expectation. If X is independent of $\mathcal{B}$, we get $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$.

[1] Translation to English by Pierre Nicodème of a note of Brigitte Chauvin written in French.

Proposition 2 (properties of the conditional expectation)

• *linearity :* $\forall a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B})$

• $|\mathbb{E}(X|\mathcal{B})| \leq \mathbb{E}(|X|\ |\mathcal{B})$

• *If* $\mathcal{C}$ *is a* σ*-algebra and if* $\mathcal{C} \subset \mathcal{B}$*, then* $\mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{C}) = \mathbb{E}(X|\mathcal{C})$

In particular, $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$

• *If* X *is integrable and* Z *is* $\mathcal{B}$*-measurable, then* $\mathbb{E}(XZ|\mathcal{B}) = Z\mathbb{E}(X|\mathcal{B})$*. Moreover, when* Z *is* $\mathcal{B}$*-measurable, we have* $\mathbb{E}(Z|\mathcal{B}) = Z$ *and* $\mathbb{E}(\mathbb{E}(X|Z)) = \mathbb{E}(X)$

<u>Link with the conditional probabilities</u>

Let A and B be two events, with $\mathbb{P}(B) \neq 0$. Let us choose as $\mathcal{B}$ the σ-algebra $\mathcal{B} = \{\emptyset, B, B^c, \Omega\}$. Then, one verifies with the characterization that

$$\mathbb{E}(\mathbb{1}_A|\mathcal{B}) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\mathbb{1}_B + \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)}\mathbb{1}_{B^c}$$

which gives

$$\mathbb{E}(\mathbb{1}_A|\mathcal{B}) = \mathbb{P}(A|B)\mathbb{1}_B + \mathbb{P}(A|B^c)\mathbb{1}_{B^c}.$$

A Gentle Introduction to Analytic Combinatorics

Jérémie Lumbroso Basile Morcrette

Oxford, September 5-7, 2012

"These notes were written by Jérémie Lumbroso and Basile Morcrette for the 1st French-British Young Research Workshop that took place in Oxford in 2012, and of which the purpose was to foster collaborations between French and British young researchers over topics common to them - probabilistic analyses, or analytic combinatorics. There have since been subsequent editions, most recently in Paris in 2014. Another edition is scheduled in Bath in 2015."

Contents

1 Introduction **9**
1.1 General Aim. 9
1.2 Catalan numbers, by hands 9

2 Unlabelled objects **10**
2.1 Basic definitions: combinatorial classes, generating functions 10
2.2 The symbolic method 11
2.2.1 Elementary constructions 11
2.2.2 Some direct examples 13
2.3 OGF as complex objects 14
2.4 Asymptotic of the coefficients (simple case) 14
2.5 General asymptotic scheme 15
2.6 Tree enumeration 15
2.6.1 Binary trees 16
2.6.2 Unary-Binary trees 16
2.6.3 General trees 16
2.6.4 Otter trees: the problem of symetries 16
2.6.5 Balanced 2-3 trees (external nodes): an example of substitution 16

3 Labelled objects and exponential generating functions **17**
3.1 Definition and examples 17
3.2 Construction of the sum 17
3.3 Construction of the product 18
3.4 Construction of the sequence 18
3.5 Construction of the set 19
3.6 Construction of the cycle 19
3.7 Examples of permutation classes 19
3.7.1 Permutations 19
3.7.2 Involutions 20
3.7.3 Derangements 20

4 Recursive classes. Asymptotic of trees **21**
4.1 Lagrange inversion . . . 21
4.1.1 Binary trees . . . 22
4.1.2 Unary-Binary trees . . . 22
4.1.3 Cayley trees . . . 22
4.2 Asymptotic for trees: analytic inversion . . . 22
4.2.1 Unary-Binary trees . . . 23
4.2.2 Cayley trees . . . 23

5 Other symbolic operators **23**
5.1 Boxed product . . . 23
5.2 Pointing and substitution . . . 23

6 Multivariate Generating functions using markers **24**
6.1 Definitions . . . 24
6.2 Symbolic method . . . 24
6.3 Distribution, mean, variance, moments . . . 25

7 Tree statistics **26**

8 Permutation statistics **27**
8.1 Prisoner's dilemma . . . 27
8.2 Average number of cycle . . . 28
8.3 Number of cycles of size r . . . 28

9 Statistic on mappings (or functional graphs) **29**
9.1 Expression of the BGFs . . . 30
9.2 Expected values . . . 31

10 Probability of being a connected graph **31**

11 Saddle-point method **32**
11.1 Exponential and 1/n! . . . 33
11.2 Number of involutions: asymptotics . . . 34

1 Introduction

1.1 General Aim.

- Study combinatorial structures in a simple, unified and automatic way.
- Do exact (with formal, symbolic methods) and asymptotic (with $\mathbb{C}$-analytic methods) counting.
- Examples of combinatorial structures: integers, words, permutations, trees, functional graphs.

1.2 Catalan numbers, by hands

Let's begin with one of the most famous objects in combinatorics. The approach presented here, is the typical approach one would use to find the enumeration of combinatorial objects from a recurrence, as it would be described for instance in Wilf's popular textbook [4, §1].
Consider C_n the number of binary trees of size n (i.e. with n internal nodes). A simple exhaustive study leads to the first terms $C_0 = 1$, $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, $C_4 = 14, \ldots$
A classical way of counting those numbers is to find a recurrence. A binary tree of size $n+1$ is composed of a root and two subtrees: its left child is a binary tree of size k, its right child is a binary tree of size $n-k$, and the choice of the integer k is in the set $\{0, 1, \ldots, n\}$. So, it is possible to write the recurrence scheme

$$C_{n+1} = \sum_{k=0}^{n} C_k \, C_{n-k} \, .$$

The hint is now to use a *generating function*: $C(z) = \sum_{n\geq 0} C_n \, z^n$, where the variable z is just some parameter. The sequence $(C_n)_{n\geq 0}$ is now encoded by the function $C(z)$. From the previous equation, we multiply each side by the monomial z^{n+1}, and then make the sum for $n = 0, 1, \ldots$.

$$\sum_{n\geq 0} C_{n+1} z^{n+1} = \sum_{n\geq 0} \sum_{k=0}^{n} C_k \, C_{n-k} \, z^{n+1} \, ,$$

which can be re-written

$$\sum_{n\geq 1} C_n z^n = z \sum_{n\geq 0} \sum_{k=0}^{n} \left(C_k \, z^k\right) \, \left(C_{n-k} \, z^{n-k}\right)$$

Now, using the generating function $C(z)$, we find the classical equation

$$C(z) - 1 = z \, C(z)^2$$

Solving this second order equation, and using the initial condition $C_0 = 1$ (which translates into $C(0) = 1$), the solution is

$$\boxed{C(z) = \frac{1 - \sqrt{1-4z}}{2z} \, .}$$

Finding the exact coefficients C_n is done by the formal power series expansion of $C(z)$. We use the classical Newton's generalised binomial theorem

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2} x^2 + \ldots + \frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!} x^k + \ldots \, ,$$

and find

$$C(z) = \sum_{n\geq 0} \frac{1}{n+1} \binom{2n}{n} z^n \, .$$

So we conclude saying the number of binary trees of size n is the Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}$. And if we want an asymptotic formula of C_n, we use the classical Stirling formula $n! \sim \sqrt{2\pi n}\, e^{-n}\, n^n$, and find

$$\boxed{C_n = \frac{1}{n+1}\binom{2n}{n} \sim \frac{4^n\, n^{-3/2}}{\sqrt{\pi}}\,.}$$

This course's aim is to directly get the framed results—the exact and asymptotic enumeration—from a symbolic specification of the combinatorial objects. In our current case, a binary tree can be symbolically specified as being: either a single leaf (noted $\circ$), or a node (noted $\bullet$), with a pair of binary trees (the left and right children), thus

$$\mathcal{B} = \circ \text{ or } (\bullet, \mathcal{B}, \mathcal{B})$$

which of course bears a striking resemblance with the functional equation satisfied by the generating function, $C(z) = 1 + zC(z)C(z)$...

2 Unlabelled objects

This section summarizes the main aspects of the first chapter of the reference book [2, §I].

2.1 Basic definitions: combinatorial classes, generating functions

Definition 1. A *combinatorial class* $\mathcal{A}$ (sometimes simply a *class*) is a finite or denumerable set on a which is defined a *size* function, $|\cdot| : \mathcal{A} \to \mathbb{Z}_{\geqslant 0}$, such that, for every size there is only a *finite* number of elements, that is

$$\forall n \in \mathbb{Z}_{\geqslant 0},\ a_n := |\{x \in \mathcal{A} \mid |x| = n\}| < \infty.$$

Remark. Following the common usage (as formalized in Flajolet and Sedgewick's reference text [2]), we will always denote combinatorial classes using upper-case calligraphic letters such as $\mathcal{A}$, subclasses containing only elements of a given size n as $\mathcal{A}_n$, and the counting sequences using the lower-case roman type, a_n.

As the definition suggests, for a given combinatorial class, there may be several different valid size functions. A well-known example in combinatorics is that of planar[1] binary trees: we can for instance enumerate them according to the number of internal nodes, the number of external nodes (also called *leaves*), or by counting both.
On the other hand, a trivial measure of size that would *not* be valid would be to count the number of children of the root (either 0, 1, or 2) as we would then have an infinite number of trees of "size" 1 and 2.

Definition 2. Let $\mathcal{A}$ be a combinatorial class, and let $(a_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be its *counting sequence*. We call $A(z)$ the *ordinary generating function* (or OGF) associated with $\mathcal{A}$,

$$A(z) := \sum_{n=0}^{\infty} a_n z^n.$$

In some cases, it is also sometimes convenient to consider the equivalent definition of generating function as the sum over the objects of combinatorial class $\mathcal{A}$

$$A(z) := \sum_{\alpha\in\mathcal{A}} z^{|\alpha|}.$$

[1]The term *planar* is here used to express that a combinatorial structure is embedded in the plane; in the case of binary trees, that means that we distinguish a left and a right child.

Combinatorial class	Counting sequence	OGF
Words on $\{0,1\}^\infty$	2^n	$W(z) = \dfrac{1}{1-2z}$
Integer compositions	2^{n-1}	$I(z) = \dfrac{1-z}{1-2z}$
Binary trees (counting internal node)	$\dfrac{1}{n+1}\dbinom{2n}{n}$	$B(z) = \dfrac{1-\sqrt{1-4z}}{2z}$
Permutations	$n!$	$P(z) = \sum_{n=0}^{\infty} n! z^n$

Table 1. Some standard combinatorial classes, their enumeration sequence, and their ordinary generating function (OGF). Note permutations do not have an analytic ordinary generating function, i.e., the radius of convergence of $P(z)$ is 0.

Exercise 1. Show that these two definitions are equivalent.

The generating function is a traditional object in combinatorics. But where it is usually considered as a formal object, algebraically manipulated, analytic combinatorics shows that there is considerable power in instead considering them as analytic objects.
Once given a generating function, our main goal will be to *extract its coefficients*. Let $f(z)$ be a generating function, we use the notation $[z^n]$ to note the coefficient of the variable z^n,

$$[z^n]f(z) = [z^n]\left(\sum_{i=0}^{\infty} f_i z^i\right) = f_n.$$

Here are some elementary but very fundamental operations on coefficients, which will also be revisited later on.

- **Scaling:** $[z^n]f(\lambda z) = \lambda^n [z^n]f(z)$, as

$$[z^n]f(\lambda z) = [z^n]\left(\sum_{i=0}^{\infty} f_i (\lambda z)^i\right) = [z^n]\left(\sum_{i=0}^{\infty} (f_i \lambda^i) z^i\right) = \lambda^n [z^n] f(z).$$

- **Right shifting:** $[z^n]z^k f(z) = [z^{n-k}]f(z)$, because

$$[z^n]z^k f(z) = [z^n]\left(\sum_{i=0}^{\infty} f_i z^{i+k}\right) = [z^n]\left(\sum_{i=k}^{\infty} f_{i-k} z^i\right) = [z^{n-k}]f(z).$$

2.2 The symbolic method

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be combinatorial classes with respective ordinary generating functions $A(z)$, $B(z)$ and $C(z)$. The symbolic method is the observation that some symbolic operations can *directly be translated* to ordinary generating functions.

2.2.1 Elementary constructions

The base elements are neutral objects, noted ε, which have no size and are thus translated as $z^{|\epsilon|} = z^0 = 1$, and atomic objects with size 1, noted $\mathcal{Z}$, and translated to OGFs as the variable z. In addition, we can distinguish however many kinds of neutral objects, for instance ε_1, ε_2, etc., which will all translate to 1, and however many kinds of atomic objects, which may translate either to the same variable z, or to some other variable z_1, z_2, etc. depending on whether it is important to distinguish the type of atom it contributes to.

Disjoint union. We write $\mathcal{A} = \mathcal{B} + \mathcal{C}$, if class $\mathcal{A}$ is defined as the disjoint union of $\mathcal{B}$ and $\mathcal{C}$: that is $\mathcal{A}$ contains all objects from $\mathcal{B}$ and $\mathcal{C}$, and objects keep their original sizes. Because the union is disjoint, there is no overlap in the enumeration, and this translates to the generating functions as

$$A(z) = B(z) + C(z).$$

Indeed, using the combinatorial definition of OGFs, since objects from $\mathcal{A}$ are either from $\mathcal{B}$ or $\mathcal{C}$,

$$A(z) = \sum_{\alpha\in\mathcal{A}} z^{|\alpha|} = \sum_{\alpha\in\mathcal{B}} z^{|\alpha|} + \sum_{\alpha\in\mathcal{C}} z^{|\alpha|} = B(z) + C(z).$$

Remark. Although we speak of "disjoint union", in practice, we never concern ourselves on whether the combinatorial classes are disjoint; instead we consider we are doing the union of unique copies of each class (for instance, imagine that $\mathcal{A} = \mathcal{B} + \mathcal{B}$ means that $\mathcal{A}$ is composed of either elements of $\mathcal{B}$ that are colored pink or purple—thus twice as many elements).

Cartesian product. We write $\mathcal{A} = \mathcal{B} \times \mathcal{C}$, if class $\mathcal{A}$ is defined as all ordered pairs, $\alpha = (\beta, \gamma) \in \mathcal{A}$ where the first element β is from $\mathcal{B}$ and the second γ from $\mathcal{C}$ (*i.e* $\beta \in \mathcal{B}, \gamma \in \mathcal{C}$). The size function on $\mathcal{A}$ is then defined as $|\alpha| = |\beta| + |\gamma|$, thus

$$A(z) = B(z) \cdot C(z)$$

since

$$A(z) = \sum_{\alpha\in\mathcal{A}} z^{|\alpha|} = \sum_{\beta\in\mathcal{B}} \sum_{\gamma\in\mathcal{C}} z^{|\beta|+|\gamma|} = \left(\sum_{\alpha\in\mathcal{B}} z^{|\alpha|}\right) \cdot \left(\sum_{\alpha\in\mathcal{C}} z^{|\alpha|}\right) = B(z) \cdot C(z).$$

Remark. The size for Cartesian products is here the sum of the sizes of each object of a pair, and accordingly we say that we are dealing with *additive* combinatorial structures. Other rules for the Cartesian product are possible, for instance that the size of a pair be the *product* of each component; we would then be dealing with *multiplicative* combinatorial structures enumerated by *Dirichlet generating functions* (DGF),

$$D(s) = \sum_{n\geqslant 1} \frac{d_n}{n^s}.$$

These combinatorial structures are intimately tied to number theory, and in particular Riemann's zeta function features prominently as it is the DGF for the unit sequence (much like the quasi-inverse in additive combinatorics).

Sequence. We write $\mathcal{A} = \textsc{Seq}\,(\mathcal{B})$, if $\mathcal{A}$ is defined as all ordered sequences (of any size, including zero) of objects from $\mathcal{B}$,

$$\mathcal{A} := \{\varepsilon\} + \mathcal{B} + \mathcal{B} \times \mathcal{B} + \mathcal{B} \times \mathcal{B} \times \mathcal{B} + \dots$$

in other words we have

$$\mathcal{A} := \{(\beta_1, \dots, \beta_\ell) \mid \ell \geqslant 0,\ \beta_j \in \mathcal{B}\}.$$

Observe in order for $\mathcal{A}$ to be a well-defined class, it is necessary that $b_0 = 0$ (i.e. that there is no object in $\mathcal{B}$ with size zero), as then $\mathcal{A}$ would contain an infinity of objects of any given size. The translation to OGFs is

$$A(z) = \sum_{k=0}^{\infty} B(z)^k = \frac{1}{1 - B(z)}.$$

This operation is often referred to as the *quasi-inverse*.

$Structure$	OGF
$\{\varepsilon\}$	1
$\{\mathcal{Z}\}$	z
$\mathcal{A}+\mathcal{B}$	$A(z)+B(z)$
$\mathcal{A}\times\mathcal{B}$	$A(z)\cdot B(z)$
$\textsc{Seq}\,(\mathcal{A})$	$\dfrac{1}{1-A(z)}$

Table 2. Small dictionary of unlabelled combinatorial classes

Recursive classes. Finally we mention that, under certain conditions, combinatorial classes may be defined recursively, to allow for instance for the definition of branching structures. We will not go into the technical detail of these conditions (see [2, §I.2.3]), except to say that the general idea is that:

1. for every class there should be at least one terminal symbol (an atom or a neutral element);
2. a system should not allow for a same symbol to be expanded twice without increasing the size.

Example 1. This second point can be illustrated using a common mistake when specifying unary-binary trees (sometimes called Motzkin trees because they are in bijection with Motzkin paths, much like standard binary trees are in bijection with Dyck paths). If we define the class of unary binary tree as

$$\overline{\mathcal{U}} = \mathcal{Z} + \overline{\mathcal{U}} + \overline{\mathcal{U}}^2$$

that is, we define a tree is either a leaf, or an unary internal node or a binary internal node and *we count the leaves*, then the recursion is not well-founded, and there are two ways to see this.
Combinatorically, the problem is that since unary nodes (in particular) do not affect the size of a tree, it is possible to obtain an infinity of trees of the same size, simply by taking any unary-binary tree and increasing *ad infinitum* the number of unary binary nodes—without changing the size. We were able to get away with counting leaves in binary trees because binary nodes affect the number of leaves (in other words there is a direct correspondance between the number of internal nodes and external nodes).
Analytically, the problem is simply that the functional equation

$$\overline{U}(z) = z + \overline{U}(z) + \overline{U}(z)^2$$

does not admit any positive real solution.

The problem is solved by counting simultaneously the leaves by t and the internal nodes by z; this gives the equation

$$U(z,t) = t + zU(z,t) + zU^2(z,t).$$

2.2.2 Some direct examples

Example 2. Binary words on the alphabet $\{0,1\}$
A word is a finite sequence of 0 and 1.
$\mathcal{W}$ = $\textsc{Seq}(\{0\}+\{1\})$

$$W(z) = \frac{1}{1-(z+z)} \qquad \text{and} \qquad [z^n]W(z) = 2^n$$

Example 3. Number F_n of different ways to cover the segment [0,n] with bricks of size 1 and 2
Let a be an atomic class of size 1 and b an atomic class of size 2. Then, $\mathcal{F}$ = $\textsc{Seq}(a+b)$.

$$F(z) = \frac{1}{1-(z+z^2)} = 1 + z + 2z^2 + 3z^3 + 5z^4 + \dots$$

We identify it as the Fibonacci sequence F_n. The recurrence $F_{n+2} = F_{n+1} + F_n$ is directly linked to the equation $z^2 - z - 1 = 0$.

Example 4. Integer composition [2, §I.3]
The composition of an integer n is the sequence $x_1, x_2, \ldots, x_k$ such that $n = x_1 + x_2 + \ldots + x_k$, with $x_i \geq 1$.
An integer x is an atomic class of size x, represented by the OGF z^x. The class $\mathcal{I}$ of integers has the OGF $I(z) = z + z^2 + z^3 + \ldots = \frac{z}{1-z}$.
The class of compositions of integers $\mathcal{C}$ is described by $\mathcal{C}$ =$\textsc{Seq}(\mathcal{I})$. So,

$$C(z) = \frac{1}{1 - I(z)} = \frac{1}{1 - \frac{z}{1-z}} = \frac{1}{1-2z} - \frac{z}{1-2z}$$

$$C_n = [z^n]C(z) = [z^n]\frac{1}{1-2z} - [z^n]\frac{z}{1-2z} = 2^n - 2^{n-1} = 2^{n-1}$$

Remark. For each example (words, Fibonacci numbers, integer compositions), the exponential growth of the coefficients of the OGF is directly linked to the singularity of the generating function (a singularity of a function is a point where the function is not well defined, when it grows to infinity).

2.3 OGF as complex objects

Until now, an OGF is simply a formal sum of monomials. Let's now consider[2] the OGF as a univariate function of the complex variable z.

$$f(z) = \sum_{n \geq 0} f_n z^n$$

When it is possible to write f as a Taylor expansion $f(z) = \sum_{n \geq 0} \tilde{f}_n (z - z_0)^n$, we say that f is analytic at the point z_0. In combinatorics, almost all generating functions are analytic at 0. The function f has a radius of convergence R defined by

$$R = \sup\{r \text{ such that } f(z) \text{ is analytic for } |z| < r\}$$

An other way to see the radius of convergence is

$$R^{-1} = \limsup_n |f_n|^{1/n}$$

It means that when n grows to infinity, we have $f_n \sim R^{-n}\theta(n)$ where $\theta(n)$ is a subexponential function of n. The definition impose that it must exist a singularity on the circle $|z| < r$. Furthermore, a classical theorem in complex analysis (due to Pringsheim) says: If the coefficients f_n are non negative, then there exists a singularity at the point of the real line $z = R$.

2.4 Asymptotic of the coefficients (simple case)

Lemma 1. *(Schützenberger) All the combinatorial constructions upon (ε, $\mathcal{Z}$, +, $\times$, $\textsc{Seq}$) leads to generating functions that are rational.*

Indeed, ϵ and $\mathcal{Z}$ translates to 0 and z that are trivial rational expressions; moreover the operators $+, \times$ and $\textsc{Seq}$ transform a pair of rational functions, or a rational function, to another rational function (where a polynomial is a rational function of denominator 1).
Let f be an OGF. It is possible to write f as a quotient of two polynomials $A(z)$ and $B(z)$. And so, finding the singularities of f is equivalent to finding the zeros of the denominator $B(z)$. The rational function f has a partial fraction expansion:

$$f(z) = \text{polynomial} + \sum_{(\rho, r), B(\rho)=0} \frac{c}{(1 - z/\rho)^r} \qquad (r \in \mathbb{N})$$

[2]This material is covered partially in [2, §IV.1 p.225] for the complex nature of the OGF, and then the exponential growth is explained in §IV.3 p.238 and in particular §IV.3.2 p.243.

Finding the asymptotics of the coefficients f_n is equivalent to the study of the asymptotics of $(1-z/\rho)^{-r}$.

$$\begin{aligned}
[z^n]\frac{1}{(1-z/\rho)^r} &= \rho^{-n}[z^n](1-z)^{-r} \\
&= \rho^{-n}\binom{n+r-1}{r-1} \\
&= \rho^{-n}\frac{(n+r-1)(n+r-2)\dots(n+1)}{(r-1)!} \\
&\sim \frac{\rho^{-n}n^{r-1}}{(r-1)!}
\end{aligned}$$

Finally, f_n is a sum of terms of the form $c\,\rho^{-n}n^{r-1}$. (This is a version of Theorem VI.1 p.381 in [2], when $\rho = 1$.)

Conclusive remarks

- the singularity which is the closest to the origin give the exponential growth in the asymptotics. The singularity of minimal modulus is called *dominant singularity*.
- the subexponential term of this asymptotic is given by the multiplicity of the dominant singularity.

Example 5. Find the asymptotics of the coefficients of

$$f(z) = (1-z^2/2)^{-5}(1-z^3)^{-1}(1-2z)^{-5}(1-z-z^2)^{-1}\,.$$

Singularities:=$\{\sqrt{2}, -\sqrt{2}, 1, 1/2, \phi, \overline{\phi}\}$ Dominant singularity: $z = 1/2$ Multiplicity: 5. So, $f_n = [z^n]f(z) \sim c\,2^n n^4$.

2.5 General asymptotic scheme

With more detailed complex analysis, it is possible to get the asymptotic of other generating functions (not necessarily rational). This is Theorem VI.2 p.385 in [2], also seen in the special case where the singularity is $\rho = 1$ (using the property of scaling, $[z^n]f(\rho z) = \rho^n[z^n]f(z)$, we can always get back to this case).

Theorem 1. *(Subexponential asymptotic term). For $\alpha \in \mathbb{R}\backslash\{0,-1,-2,\dots\}$, and $k \in \mathbb{N}$,*

$$[z^n]\frac{1}{(1-z)^\alpha}\log^k\left(\frac{1}{1-z}\right) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\log^k(n),$$

where Γ is the classical generalized factorial function: $\Gamma(x) = \int_0^\infty e^{-t}\,t^{x-1}dt$.

Theorem 2. *(Transfer lemma, Th. VI.3 p.390 [2])*
If $f(z) \sim_{z\to 1} g(z)$, then $f_n \sim g_n$.
If $f(z) =_{z\to 1} O(g(z))$, then $f_n = O(g_n)$.
If $f(z) =_{z\to 1} o(g(z))$, then $f_n = o(g_n)$.

This powerful theorem expresses that it is enough to know the comparative behaviour of two functions in the neighbourhood of their smallest singularity (here assumed to be 1).
The intuition is that a function's behaviour around its singularity is extremal and dictated exactly by its singularity.

Remark. For a more detailled lemma (with all hypothesis), see [2]. Moreover, instead of getting only a first order equivalent, it is also possible to have a more precise asymptotic expansion with several error terms.

2.6 Tree enumeration

The topic here is fully covered in [2, §I.5].

2.6.1 Binary trees (number of internal nodes). $\mathcal{B} = \varepsilon + \mathcal{Z}\times\mathcal{B}\times\mathcal{B}$

So, $B(z) = 1 + zB(z)^2$. We solve the equation and find $B(z) = \frac{1-\sqrt{1-4z}}{2z}$.
The singularity is at $z = 1/4$, and the order is $-1/2$.
Near $z = 1/4$, we can write $B(z) \sim -2\frac{1}{(1-4z)^{-1/2}}$. So,

$$B_n \sim -2\frac{4^n n^{-3/2}}{\Gamma(-1/2)} \sim \frac{4^n\, n^{-3/2}}{\sqrt{\pi}} \qquad \left(\Gamma(-1/2) = -2\sqrt{\pi}\right)$$

2.6.2 Unary-Binary trees (internal and external nodes). $\mathcal{U} = \mathcal{Z} + \mathcal{Z}\times\mathcal{U} + \mathcal{Z}\times\mathcal{U}\times\mathcal{U}$

$U(z) = z + zU(z) + zU(z)^2 = z\phi(U(z))$, where $\phi(t) = 1 + t + t^2$.

Exercise 2. Find the generating function, an expression for the coefficients and an asymptotic value.

2.6.3 General trees $\mathcal{A} = \mathcal{Z}\times\textsc{Seq}(\mathcal{A})$

$$A(z) = \frac{z}{1 - A(z)} \qquad \text{so,} \qquad A(z) = z + A(z)^2$$

$$A(z) = \frac{1-\sqrt{1-4z}}{2} \qquad A_n \sim \frac{4^{n-1} n^{-3/2}}{\sqrt{\pi}}$$

Remark. We notice that $zB(z) = A(z)$. Then, $[z^{n-1}]B(z) = [z^n]A(z)$, and $B_{n-1} = A_n$. The bijection between binary trees and general trees is here proved thanks to the symbolic method!

2.6.4 Otter trees: the problem of symetries

An Otter tree $\mathcal{T}$ is a rooted binary non-planar unlabelled tree.

$$T(z) = z + z^2 + z^3 + 2z^4 + 3z^5 + 6z^6 + 11z^7 + \dots$$

An Otter tree is just a leaf, or it is a node with two Otter subtrees. But there is a symmetry at this node, so we put a factor $1/2$ in the counting of those configurations. But with this correction, when the two subtrees are exactly the same, it it now counted just a half time. So we add the other half for those subtrees. Then,

$$T(z) = z + \frac{1}{2}T(z)^2 + \frac{1}{2}T(z^2).$$

2.6.5 Balanced 2-3 trees (external nodes): an example of substitution

Balanced 2–3 trees are trees where each node is:

- a leaf,
- an internal node with two or three sons,

and all leaves are at the same distance from the root.
The combinatorial specification is:

$$\mathcal{E} = \mathcal{Z} + \mathcal{E} \circ [\{\mathcal{Z}\times\mathcal{Z}\} + \{\mathcal{Z}\times\mathcal{Z}\times\mathcal{Z}\}] \rightsquigarrow E(z) = z + E(z^2 + z^3),$$

since trees with depth h are transformed to trees of depth $h+1$ by substituting each leaf by an internal node and two or three leaves.

3 Labelled objects and exponential generating functions

We now discuss the topic of labelled objects, introduced in [2, §II.1 and 2].
As noted, for instance in Table 1, the class of permutations does not have an analytic OGF, because the coefficients $n!$ grow exponentially faster than z^n and thus the radius of convergence the ordinary generating function is zero.
This combinatorial explosion is a common trait shared by all combinatorial classes that are *labelled*—that is, of which the atoms are endowed with a permutation of n, the size. Permutations are such a class (a permutation is a sequence of labelled atoms), as are arrangements (a subset of labelled atoms), and more complex objects such as graphs.

3.1 Definition and examples

The solution is to enumerate these objects using *exponential generating functions*, in which the coefficient is normalized by $n!$.

Definition 3. Let $\mathcal{A}$ be a *labelled* combinatorial class, and let $(a_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be its *counting sequence*. We call $A(z)$ the *exponential generating function* (or EGF) associated with $\mathcal{A}$,

$$A(z) := \sum_{n=0}^{\infty} a_n \frac{z^n}{n!}.$$

And with EGFs there is also a combinatorial definition,

$$A(z) := \sum_{\alpha\in\mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!}.$$

Notice that now, extracting the coefficient leads to a factorial factor:

$$a_n = n![z^n]A(z)$$

Example 6. $\mathcal{P}$ = {Permutations}

$$P(z) = \sum_{n\geq 0} n!\frac{z^n}{n!} = \frac{1}{1-z}$$

It looks like a sequence of atoms. Indeed, a permutation can be viewed as a linear graph of size n:

$$\sigma(1) — \sigma(2) — \sigma(3) — \;\ldots\; — \sigma(n)$$

Example 7. $\mathcal{U}$: non connected graphs (graphs with no edge). For all n, $U_n = 1$.

$$U(z) = \sum_{n\geq 0} \frac{z^n}{n!} = e^z$$

Example 8. $\mathcal{K}$: Complete graphs (all edges). It is the same EGF, $K(z) = e^z$.

Example 9. $\mathcal{C}$: Cyclic graphs (with a given orientation in the plan). $C_n = (n-1)!$. So,

$$C(z) = \sum_{n\geq 1}(n-1)!\frac{z^n}{n!} = \sum_{n\geq 1}\frac{z^n}{n} = \log\left(\frac{1}{1-z}\right).$$

3.2 Construction of the sum

The disjoint union is the same construction as the unlabelled case. If $\mathcal{A} = \mathcal{B} + \mathcal{C}$, then the EGF of $\mathcal{A}$ is $A(z) = B(z) + C(z)$.

3.3 Construction of the product

Starting with two labelled structures β and γ, the classical Cartesian product does not provide a well labelled structure. The set of labels of a well-labelled structure of size n is exactly the set of integers $[1, n]$.
So, from a couple (β, γ), we define a re-labelled structure (β', γ') where the labels are exactly $\{1, \ldots, |\beta| + |\gamma|\}$, and the relative order of labels of each element is preserved. We define

$$\beta \star \gamma = \{\text{ all couples } (\beta', \gamma') \text{ well relabelled}\}$$

The class $\beta \star \gamma$ contains exactly $\binom{|\beta|+|\gamma|}{|\beta|}$ distinct elements. Then we can define the *labelled* product

$$\mathcal{A} = \mathcal{B} \star \mathcal{C} = \bigcup_{\beta \in \mathcal{B}, \gamma \in \mathcal{C}} \beta \star \gamma$$

Lemma 2. $A(z) = B(z) \cdot C(z)$

Proof.

$$\begin{aligned} A(z) &= \sum_{\alpha \in \mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!} \\ &= \sum_{\beta \in \mathcal{B}} \sum_{\gamma \in \mathcal{C}} \sum_{\alpha \in \beta \star \gamma} \frac{z^{|\beta|+|\gamma|}}{(|\beta| + |\gamma|)!} \\ &= \sum_{\beta \in \mathcal{B}} \sum_{\gamma \in \mathcal{C}} \binom{|\beta| + |\gamma|}{|\beta|} \frac{z^{|\beta|} z^{|\gamma|}}{(|\beta| + |\gamma|)!} \\ &= \sum_{\beta \in \mathcal{B}} \sum_{\gamma \in \mathcal{C}} \frac{z^{|\beta|} z^{|\gamma|}}{|\beta|! \, |\gamma|!} \\ &= B(z) \cdot C(z) \end{aligned}$$

□

Remark. $\mathcal{B} \star \mathcal{B} := \mathcal{B}^2$ does not contain elements (β, β): the re-labelling make the two βs different.

3.4 Construction of the sequence

Since we have the two constructions, sum and labelled product, it is possible to construct the sequence as before. For any labelled class $\mathcal{B}$ where $b_0 = 0$,

$$\mathcal{A} = \textsc{Seq}\,(\mathcal{B}) = \{\alpha \text{ s.t. } \exists k \geq 0,\, \alpha = (\beta_1, \ldots, \beta_k) \text{ finite re-labelled sequence, } \beta_i \in \mathcal{B}\}$$

$$\textsc{Seq}\,(B) = \{\varepsilon\} + \mathcal{B} + \mathcal{B} \star \mathcal{B} + \mathcal{B} \star \mathcal{B} \star \mathcal{B} + \ldots$$

The corresponding EGF is

$$A(z) = \sum_{k \geq 0} B(z)^k = \frac{1}{1 - B(z)}$$

Definition 4. k components sequence : $\textsc{Seq}_k(\mathcal{A}) = \mathcal{A}^k$

3.5 Construction of the set

A k components set is defined as:

$$\textsc{Set}_k(\mathcal{B}) := \{\text{sets with } k \text{ elements of } \mathcal{B}\}$$

This class can be viewed as an equivalence class:

$$\textsc{Set}_k(\mathcal{B}) = \frac{\textsc{Seq}_k(\mathcal{B})}{\mathfrak{R}}$$

where $\mathfrak{R}$ is the following equivalence relation:
$(\beta_1, \ldots, \beta_k)\mathfrak{R}(\beta'_1, \ldots, \beta'_k)$ iff there exists a permutation $\sigma \in \mathfrak{S}_k$ such that $\beta_{\sigma(i)} = \beta'_i$.
We notice that the ratio of cardinalities is:

$$\frac{|\textsc{Set}_k(\mathcal{B})|}{|\textsc{Seq}_k(\mathcal{B})|} = \frac{1}{k!}\,.$$

Then, we define the $\textsc{Set}$ constructor:

$$\mathcal{A} := \textsc{Set}\,(\mathcal{B}) = \bigcup_{k\geq 0} \textsc{Set}_k(\mathcal{B})\,,$$

and the corresponding EGF is

$$A(z) = \sum_{k\geq 0} \frac{1}{k!} A(z)^k = \exp(B(z))\,.$$

3.6 Construction of the cycle

For any labelled class $\mathcal{B}$ with $b_0 = 0$ and $k \geq 1$, the class of k components cycle is

$$\textsc{Cyc}_k(\mathcal{B}) := \{\text{cycles with } k \text{ elements of } \mathcal{B}\}$$

This class can be viewed as an equivalence class:

$$\textsc{Cyc}_k(\mathcal{B}) = \frac{\textsc{Seq}_k(\mathcal{B})}{\mathfrak{T}}\,,$$

where $\mathfrak{T}$ is the following equivalence relation:
$(\beta_1, \ldots, \beta_k)\mathfrak{T}(\beta'_1, \ldots, \beta'_k)$ iff there exists a cyclic permutation $\tau \in \mathfrak{S}_k$ such that $\beta_{\tau(i)} = \beta'_i$.
We notice that the ratio of cardinalities is:

$$\frac{|\textsc{Cyc}_k(\mathcal{B})|}{|\textsc{Seq}_k(\mathcal{B})|} = \frac{1}{k}\,.$$

Then, we define the $\textsc{Cyc}$ constructor:

$$\mathcal{A} := \textsc{Cyc}\,(\mathcal{B}) = \bigcup_{k\geq 0} \textsc{Cyc}_k(\mathcal{B})\,,$$

and the corresponding EGF is

$$A(z) = \sum_{k\geq 1} \frac{1}{k} A(z)^k = \log\left(\frac{1}{1-B(z)}\right)\,.$$

3.7 Examples of permutation classes

3.7.1 Permutations

$$P(z) = \frac{1}{1-z} = \exp\left(\log\left(\frac{1}{1-z}\right)\right)$$

This corresponds to the symbolic equation:

$$\mathcal{P} = \textsc{Set}\,(\textsc{Cyc}\,(\mathcal{Z}))$$

This express the classical decomposition of a permutation in a product of cycles with disjoint supports.

$Structure$	EGF
$\{\varepsilon\}$	1
$\{\mathcal{Z}\}$	z
$\mathcal{A}+\mathcal{B}$	$A(z)+B(z)$
$\mathcal{A}\star\mathcal{B}$	$A(z)\cdot B(z)$
$\textsc{Seq}\,(\mathcal{A})$	$\dfrac{1}{1-A(z)}$
$\textsc{Set}\,(\mathcal{A})$	$\exp(A(z))$
$\textsc{Cyc}\,(\mathcal{A})$	$\log\left(\dfrac{1}{1-A(z)}\right)$

Table 3. Small dictionary of labelled combinatorial classes

3.7.2 Involutions

An involution σ is a permutation such that $\sigma^2 = Id$. It can be viewed as a product of permutations of size 1 and 2 with disjoint supports, that is a set of cycles of size 1 or 2. All permutations are defined by: $\mathcal{P} = \textsc{Set}\,(\textsc{Cyc}\,(\mathcal{Z}))$. Involutions are specified by $\mathcal{I} = \textsc{Set}(\textsc{Cyc}_{\leq 2}(\mathcal{Z}))$. Then, the EGF is

$$\begin{aligned} I(z) &= \exp\left(z+\frac{z^2}{2}\right) \\ &= \sum_{n\geq 0}\frac{1}{n!}\left(z+z^2/2\right)^n \\ &= \sum_{n\geq 0}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\frac{1}{2^k}z^{2k}z^{n-k} \\ &= \sum_{n\geq 0}\frac{z^n}{n!}\sum_{k=0}^{n}\binom{n}{k}\frac{1}{2^k}z^k \end{aligned}$$

Extracting the coefficient,

$$\begin{aligned} [z^n]I(z) &= \frac{1}{n!}\binom{n}{0}\frac{1}{2^0}+\frac{1}{(n-1)!}\binom{n-1}{1}\frac{1}{2^1}+\ldots+\frac{1}{(n-k)!}\binom{n-k}{k}\frac{1}{2^k}+\ldots \\ &= \sum_{i=0}^{\lfloor n/2\rfloor}\frac{1}{(n-i)!}\binom{n-i}{i}\frac{1}{2^i} \end{aligned}$$

Finally, the exact number of involutions of size n is $I_n = \displaystyle\sum_{i=0}^{\lfloor n/2\rfloor}\frac{n!}{i!\,(n-2i)!\,2^i}$.

Remark. Finding an asymptotic for those formula will be develop later (*Saddle-point* analysis).

3.7.3 Derangements

A derangement is a permutation without fix points

$$\mathcal{D} = \textsc{Set}\,(\textsc{Cyc}_{>1}(\mathcal{Z}))$$

$$D(z) = \exp\left(\frac{z^2}{2}+\frac{z^3}{3}+\ldots\right) = \exp\left(\log\left(\frac{1}{1-z}\right)+z\right) = \frac{e^{-z}}{1-z}$$

$$d_n = n![z^n]D(z) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k (n-k)! = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

Remark. The probability for a random permutation of being a derangement is:

$$\frac{d_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \longrightarrow_{n\to\infty} e^{-1}$$

Remark. It can be directly done by singularity analysis. The singularity of D(z) is at $z = 1$. At this point, the asymptotic expansion of D(z) is

$$D(z) \sim_{z=1} \frac{e^{-1}}{1-z}, \qquad \text{so,} \qquad d_n \sim \frac{n!}{e}.$$

4 Recursive classes. Asymptotic of trees

(Covered in I.5 and II.5 of the book.)
In the previous examples of class of trees (binary, unary-binary, general), we saw that the generating function is often (or almost) of the form $A(z) = z\phi(A(z))$. This formula express the classical recursive definition of tree structures.
For example,

- $\phi(t) = 1 + t + t^2$, we have unary-binary trees;
- $\phi(t) = 1/(1 - t)$ is for general trees;

Example 10. The Cayley tree is a rooted labelled non planar tree. Its recursive definition is a node and a set of subtrees. So, $\mathcal{T} = \mathcal{Z} \star \textsc{Set}(\mathcal{T})$.

$$T(z) = z \exp(T(z)).$$

For Cayley trees, $\phi(t) = e^t$.

$$\boxed{\text{How to get easily exact and asymptotic formula?}}$$

4.1 Lagrange inversion

Theorem 3. *If $A(z) = z\phi(A(z))$, then the tree equation has a unique solution which satisfies:*

$$[z^n]A(z) = \frac{1}{n}[y^{n-1}]\phi(y)^n;$$

$$[z^n]A(z)^k = \frac{k}{n}[y^{n-k}]\phi(y)^n.$$

Remark. This theorem needs some analytic hypothesis on the function ϕ, which are always verified for classical tree examples.

Proof.

Lemma 3. *If $f(z) = \sum_{n\geq 0} f_n z^n$ is analytic, then we have by the Cauchy formula*

$$f_n = \frac{1}{2i\pi} \oint f(z) \frac{dz}{z^{n+1}}.$$

If $z = \dfrac{A(z)}{\phi(A(z))} = \dfrac{y}{\phi(y)}$, then by differentiation, $dz = \dfrac{dy}{\phi(y)} - \dfrac{y\phi'(y)}{\phi(y)^2}dy$.
Then, the coefficient a_n can be written:

$$\begin{aligned}
[z^n]A(z) = & \quad \frac{1}{2i\pi}\oint y\frac{\phi(y)^{n+1}}{y^{n+1}}\left(\frac{dy}{\phi(y)} - \frac{y\phi'(y)}{\phi(y)^2}dy\right) \\
= & \quad \frac{1}{2i\pi}\oint \frac{\phi(y)^n}{y^n}dy - \frac{1}{2i\pi}\oint \frac{\phi^{n-1}\phi'}{y^{n-1}}dy \\
= & \quad [y^{n-1}]\phi(y)^n - \frac{1}{n}[y^{n-2}](\phi(y)^n)'
\end{aligned}$$

If we write $\phi(y)^n = \sum \alpha_p y^p$, then $(\phi(y)^n)' = \sum p\alpha_p y^{p-1}$.

Therefore, $[z^n]A(z) = \alpha_{n-1} - \frac{1}{n}(n-1)\alpha_{n-1} = \frac{1}{n}\alpha_{n-1}$.

Finally, $[z^n]A(z) = \frac{1}{n}[y^{n-1}]\phi(y)^n$. □

4.1.1 Binary trees $\quad \mathcal{B} = \varepsilon + \mathcal{Z} \times \mathcal{B} \times \mathcal{B}$

$B(z) = 1 + zB(z)^2$ does not fit to the specification but if we set $C(z) = B(z) - 1$, then $C(z) = z(1 + C(z))^2$. Thanks to the Lagrange inversion,

$$[z^n]C(z) = \frac{1}{n}[y^{n-1}](1+y)^{2n} = \frac{1}{n}\binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}.$$

4.1.2 Unary-Binary trees. $\quad U(z) = z(1 + U(z) + U(z)^2)$

$$u_n = [z^n]T(z) = \frac{1}{n}[y^{n-1}](1+y+y^2)^n = \frac{1}{n}\sum_{n_1+n_2+n_3=n, n_2+2n_3=n-1}\binom{n}{n_1, n_2, n_3}$$

4.1.3 Cayley trees. $\quad \mathcal{T} = \mathcal{Z} \star \textsc{Set}\,(\mathcal{T})$

The tree equation is $T(z) = ze^z$.

$$[z^n]T(z) = \frac{1}{n}[y^{n-1}]e^{nz} = \frac{1}{n}\frac{n^{n-1}}{(n-1)!} = \frac{n^{n-1}}{n!}\,.$$

Finally, $T_n = n![z^n]T(z) = n^{n-1}$.

4.2 Asymptotic for trees: analytic inversion

The following is based on the implicit function theorem (see [2] Prop. IV.5 p.278 and Thm VI.6 p.404).

Theorem 4. *If $Y(z) = z\phi(Y(z))$, with ϕ an analytic function of radius of convergence R, and if there exists a unique τ, $0 < \tau < R$ such that $\phi(\tau) = \tau\phi'(\tau)$, then, $Y(z)$ is analytic at $z = 0$, its radius of convergence is $\rho = 1/\phi'(\tau)$, and $Y(z)$ has an asymptotic expansion near its singularity ρ,*

$$Y(z) \sim_{z=\rho} \tau - \gamma\sqrt{1 - z/\rho}$$

where $\gamma = \sqrt{2\phi(\tau)/\phi''(\tau)}$.

4.2.1 Unary-Binary tree $U(z) = z(1 + U(z) + U(z)^2)$

We need $1 + \tau + \tau^2 = \tau(1 + 2\tau)$, which implies $\tau^2 = 1$. So, $\rho = 1/3$ and $\gamma = \sqrt{3}$.
So, for z near $1/3$, we have $U(z) \sim 1 - \sqrt{3}\sqrt{1 - 3z}$
Finally, the singularity analysis leads to the asymptotic

$$U_n \sim \frac{\sqrt{3}}{2} \frac{3^n\, n^{-3/2}}{\sqrt{\pi}}\,.$$

4.2.2 Cayley tree $T(z) = ze^{T(z)}$

The equation $e^\tau = \tau e^\tau$ implies $\tau = 1$. So, the radius of convergence is $\rho = e^{-1}$, and $\gamma = \sqrt{2}$. Finally,

$$T(z) \sim_{z=e^{-1}} 1 - \sqrt{2}\sqrt{1 - ez}$$

The singularity analysis implies

$$T_n = n![z^n]T(z) \sim n!\, \frac{e^n\, n^{-3/2}}{\sqrt{2\pi}}$$

Remark. Besides, we know that $T_n = n^{n-1}$, so it is possible to re-discover the Stirling formula

$$n^{n-1} \sim n! \frac{e^n n^{-3/2}}{\sqrt{2\pi}}\,.$$

5 Other symbolic operators

5.1 Boxed product

Let us defined a modified labelled product, when $\mathcal{B}$ is a class with no element of size 0, ($b_0 = 0$).
$\mathcal{A} = \mathcal{B}^\square \star \mathcal{C}$ is the subset of $\mathcal{B} \star \mathcal{C}$ with labels such that the smallest label is in the $\mathcal{B}$ component. The generating function of $\mathcal{A}$ is given by

$$A(z) = \int_0^z \left(\frac{d}{dt}B(t)\right) C(t)dt\,.$$

Example 11. records in permutation, increasing binary trees.

5.2 Pointing and substitution

Those two operations are the same in labelled and unlabelled world.

Pointing. This operator written Θ points a distinguished atom.
$\mathcal{A} = \Theta\mathcal{B}$ means $\mathcal{A}_n = [1, n] \times \mathcal{B}_n$. Constructing an object of size n in $\mathcal{A}$ is choosing an object of size n in $\mathcal{B}$ and point one of the n atoms of this object. Clearly, we have $a_n = nb_n$, so

$$A(z) = z\frac{d}{dz}B(z)\,.$$

Substitution $\mathcal{A} = \mathcal{B} \circ \mathcal{C}$ means substitute every atom of $\mathcal{B}$ by elements of $\mathcal{C}$. It translates directly into $A(z) = B(C(z))$.

6 Multivariate Generating functions using markers

In this course we consider a very simple extension of our combinatorial objects to allow for the analysis of special parameters in function of the size of an object. For simplicity, we will restrain ourselves to a simple type of parameter that can be expressed in terms of *markers* (see [2] III.1 p.152), but the technique is powerful enough to consider much more advanced parameters, for instance recursive (see [2] III.5 p.181) or extremal (see [2] III.8 p.214) ones.

6.1 Definitions

Definition 5. A parameter χ for a combinatorial class $\mathcal{A}$ is a function $\chi : \mathcal{A} \longrightarrow \mathbb{N}$.

Example 12. Number of letters in a word, height of a tree, number of disconnected nodes in a graph.

Definition 6. Let $\mathcal{A}$ be a class and χ be a parameter on $\mathcal{A}$. The bivariate generating function (BGF) associated to this couple $(\mathcal{A}, \chi)$ is

$$A(z,u) := \sum_{\alpha\in\mathcal{A}} z^{|\alpha|} u^{\chi(\alpha)} \text{ (unlabelled)} \qquad A(z,u) := \sum_{\alpha\in\mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!} u^{\chi(\alpha)} \text{ (labelled)}$$

Equivalently, we have

$$A(z,u) = \sum_{n,k\geq 0} a_{n,k} z^n u^k \text{ (unlabelled)} \qquad A(z,u) = \sum_{n,k\geq 0} a_{n,k} \frac{z^n}{n!} u^k \text{ (labelled)}$$

where

$$a_{n,k} = |\{\alpha \in \mathcal{A} \text{ such that } |\alpha| = n,\ \chi(\alpha) = k\}| \ .$$

Notation $[z^n u^k]A(z,u) = a_{n,k}$ (unlabelled) and $\dfrac{a_{n,k}}{n!}$ (labelled).

Remark. When u is set to 1, we obtain the univariate OGF or EGF.
$A(z,1) = \sum_n \sum_k a_{n,k} z^n 1^k = \sum_n a_n z^n = A(z)$ (in case of an OGF).

6.2 Symbolic method

All previous symbolic constructions are preserved when we use multivariate generating functions. Now, in the specifications, we are allowed to add markers, stickers (•) on the objects.
In the unlabelled world, we still have a direct correspondence for Union, Product, Sequence. In the labelled world, we also have a direct correspondence for Union, Product, Sequence, Set, Cycle.

Example 13. (Binary words)
We want to count the number of ones in a binary word (with alphabet $\{0,1\}$).
$\mathcal{W} = \text{SEQ}\,(\mathcal{Z}_0 + \bullet\mathcal{Z}_1)$, and the bivariate generating function is $W(z,u) = \dfrac{1}{1-(z+uz)}$.

$$w_{n,k} := [z^n u^k]W(z,u) = [u^k][z^n](1 - z(1+u))^{-1} = [u^k](1+u)^n = \binom{n}{k},$$

where $w_{n,k}$ is the number of words of size n with k ones.
$W(z,1) = (1-2z)^{-1}$, so $[z^n]W(z,1) = 2^n$.
The distribution is now easy to compute:

$$\mathbb{P}_n[\text{drawing a word with } k \text{ ones}] = \frac{\binom{n}{k}}{2^n} = \frac{[z^n u^k]W(z,u)}{[z^n]W(z,1)} \ .$$

6.3 Distribution, mean, variance, moments

What is said here applies to all multivariate generating functions, even those obtained with more powerful techniques than markers (see III.2 p.156).

Definition 7. (Distribution.) Considering a class $\mathcal{A}$ and a parameter χ, let $A(z,u)$ be its BGF. The distribution of the parameter χ, uniformly with respect to the size, is given by

$$\mathbb{P}_n[\chi = k] = \frac{[z^n u^k]A(z,u)}{[z^n]A(z,1)}\,.$$

Remark. We always consider that objects of the same size have the same probability to be chosen. For a class $\mathcal{A}$, we consider therefore a uniform distribution over $\mathcal{A}_n$.

Definition 8. (Mean.) For a class $\mathcal{A}$, a parameter χ and the associated BGF $A(z,u)$, the expected value of the parameter χ is given by

$$\mathbb{E}_n[\chi] = \frac{[z^n]\left(\frac{d}{du}A(z,u)\right)|_{u=1}}{[z^n]A(z,1)}\,.$$

Proof.

$$\begin{aligned}\frac{[z^n]\left(\frac{d}{du}A(z,u)\right)|_{u=1}}{[z^n]A(z,1)} &= \frac{[z^n]\left(\sum_{n,k} k\, a_{n,k} z^n u^{k-1}\right)|_{u=1}}{[z^n]\sum_n a_n z^n} = \frac{[z^n]\sum_{n,k} k\, a_{n,k} z^n}{a_n} \\ &= \frac{\sum_k k\, a_{n,k}}{a_n} = \sum_k k \frac{a_{n,k}}{a_n} \\ &= \sum_k k\, \mathbb{P}_n[\chi = k] = \mathbb{E}_n(\chi)\end{aligned}$$

□

Definition 9. (Moments) For a class $\mathcal{A}$, a parameter χ and the associated BGF $A(z,u)$, the factorial moment of order r of the parameter χ is given by

$$\mathbb{E}_n[\chi(\chi-1)\dots(\chi-r+1)] = \frac{[z^n]\left(\frac{d^r}{du}A(z,u)\right)|_{u=1}}{[z^n]A(z,1)}\,.$$

In particular, the variance is given by

$$\mathbb{V}_n(\chi) = \mathbb{E}_n[\chi(\chi-1)] + \mathbb{E}_n[\chi] - \mathbb{E}_n[\chi]^2\,.$$

Example 14. (Binary words)
$W(z,u) = (1 - z(1+u))^{-1}$.

$$\begin{aligned}[z^n]\left(\frac{d}{du}A(z,u)\right)\Bigg|_{u=1} &= [z^n]\left(\frac{z}{(1-z(1+u))^2}\right)\Bigg|_{u=1} = [z^n]\frac{z}{(1-2z)^2} \\ &= [z^{n-1}]\frac{1}{(1-2z)^2} = 2^{n-1}[z^{n-1}]\frac{1}{(1-z)^2} = 2^{n-1}n\end{aligned}$$

Finally, $\mathbb{E}_n[\text{number of ones}] = \frac{2^{n-1}n}{2^n} = \frac{n}{2}$, which is hopefully the result we expected.

Example 15. (Giving back the change). We have only coins of size 1, 2, and 5. The problem is to know what is the expected number of coins we receive, in general, when we are returned a total amount of n, and when the probability of drawing a coin is the same, whatever the size $(1, 2, 5)$ of the coin. The specification is in the unlabelled world, and the back given money is just a sequence of coins of size 1, then a sequence of coins of size 2, and finally a sequence of coins of size 5. On the specification, we choose to mark the number of coins of size 2.

$$\mathcal{D} = \text{SEQ}(\mathcal{Z}) \times \text{SEQ}\left(\bullet\mathcal{Z}^2\right) \times \text{SEQ}\left(\mathcal{Z}^5\right)$$

So, the corresponding generating function is:

$$D(z, u) = \frac{1}{(1-z)} \frac{1}{(1-uz^2)} \frac{1}{(1-z^5)}\,.$$

The cumulative function $C(z) := \frac{d}{du} D(z, u)|_{u=1}$ is given by

$$C(z) = \frac{z^2}{(1-z^2)^2(1-z)(1-z^5)}\,.$$

All the poles of this function are on the circle of convergence $|z| = 1$. But, the singularity $z = 1$ is the only dominant singularity because of its multiplicity (which is 4.). So, the subexponential term of asymptotic is $n^4 - 1 = n^3$. The constant factor is given by the asymptotic equivalent near the singularity $z = 1$,

$$C(z) \sim_{z=1} \frac{1}{(1-z)^4(1+z)^2(1+z+z^2+z^3+z^4)} \sim \frac{1}{2^2 \cdot 5} \frac{n^3}{3!}\,.$$

With the same technique of singularity analysis, we find $[z^n]D(z) \sim \frac{1}{2 \cdot 5} \frac{n^2}{2}$
So the expected number of coins of size 2 verifies $\mathbb{E}_n[\text{coins of size 2}] \sim \frac{n}{6}$.
The same analysis can be done for the expected number of coins of size 1 and 5, and we find:

$$\mathbb{E}_n[\text{coins of size 1}] \sim \frac{n}{3} \qquad , \qquad \mathbb{E}_n[\text{coins of size 5}] \sim \frac{n}{15}\,.$$

So, the expected number of coins is $\mathbb{E}_n[\text{number of coins}] \sim \frac{n}{3}(1 + 1/2 + 1/5) \sim \frac{17n}{30}$.

7 Tree statistics

Example 16. (Root degree of a rooted tree or "Cayley tree", [2] Ex III.12 p.179).
The aim of this problem is to find the average number of children at the root of a Cayley tree.
Specification:

$$\mathcal{T}^\bullet = \mathcal{Z} \star \text{SET}(\bullet\mathcal{T})$$
$$\mathcal{T} = \mathcal{Z} \star \text{SET}(\mathcal{T})$$

So the generating functions satisfy

$$T(z, u) = z \exp(uT(z))$$
$$T(z) = z \exp(T(z))$$

The derivative is $\frac{d}{du} T(z, u) = zT(z) \exp(uT(z))$. So, for $u = 1$, we have an expression for the cumulative function

$$\frac{d}{du} T(z, u)|_{u=1} = T(z) z \exp(T(z)) = T(z)^2\,.$$

Using the Lagrange inversion, we find the coefficient of z^n:

$$[z^n] \frac{d}{du} T(z, u)|_{u=1} = [z^n] T(z)^2 = \frac{2}{n} [y^{n-2}] e^{ny} = \frac{2}{n} \frac{n^{n-2}}{(n-2)!}\,.$$

Finally, since, $T(z, 1) = T(z) = \sum_n n^{n-1} \frac{z^n}{n!}$, the expected number of children at the root is given by

$$\mathbb{E}_n[\text{children at the root}] = \frac{[z^n] \frac{d}{du} T(z, u)|_{u=1}}{[z^n] T(z)} = \frac{2n^{n-2}}{n(n-2)!} \cdot \frac{n!}{n^{n-1}} = 2(1 - \frac{1}{n})$$

Conclusion: in general, a rooted tree has 2 children at the root!

Remark. Note that a nice direct proof (volunteered by Colin McDiarmid during the lecture in Oxford) exists, which uses the well-known fact that in a graph $G = (V, E)$, where V is the set of vertices and E the set of edges, $\sum_{v \in V} \deg(v) = 2|E|$. Let r be the root,

$$
\begin{aligned}
\mathbb{E}_n[\deg(r)] &= \sum_{v \in V} \mathbb{P}_n[v \text{ is root}] \deg(v) \\
&= \frac{1}{n} \sum_{v \in V} \deg(v) && \text{[all vertices equiprobably the root]} \\
&= \frac{2|E|}{n} && \text{[total degree formula]} \\
&= 2\left(1 - \frac{1}{n}\right) && \text{[a tree has } n-1 \text{ edges].}
\end{aligned}
$$

Indeed, direct methods can generally be simpler (especially for the toy examples considered in this course to illustrate our methods), but analytic combinatorics generally presents the advantage of providing a generic "one size fits all" method to tackle combinatorial problems which can be specified.

8 Permutation statistics

We can use all the concepts previously presented (EGF, BGF, symbolic method and singularity analysis) for the study of some statistics on permutations.

8.1 Prisoner's dilemma

Puzzle A hundred prisoners, each uniquely identified by a number between 1 and 100, have been sentenced to death. The director of the prison gives them a last chance. He has a cabinet with 100 drawers (numbered 1 to 100). In each, he'll place at random a card with a prisoner's number (all numbers different). Prisoners will be allowed to enter the room one after the other and open, then close again, 50 drawers of their own choosing, but will not in any way be allowed to communicate with one another afterwards. The goal of each prisoner is to locate the drawer that contains his own number. If *all* prisoners succeed, then they will all be spared; if at least one fails, they will all be executed.
There are two mathematicians among the prisoners. The first one, a pessimist, declares that their overall chances of success are only of the order of $1/2^{100} \simeq 8 \cdot 10^{-31}$. The second one, a combinatorialist, claims he has a strategy for the prisoners, which has a greater than 30% chance of success. Who is right?

Remark. This problem, described in [2] Notes II.15 p.124 and III.10 p.176, takes its origin from a paper by Gál and Miltersen on data structures [3, 5]. The optimality of the strategy was recently proven in 2006 by Curtin and Warshauer [1].

Solution The better strategy goes as follows. Each prisoner will first open the drawer which corresponds to his number. If his number is not there, he'll use the number he just found to access another drawer, then find a number there that points him to a third drawer, and so on, hoping to return to his original drawer in at most 50 trials. (The last opened drawer will then contain his number.) This strategy globally succeeds provided the initial permutation σ defined by σ_i (the number contained in drawer i) has *all* its cycles of length at most 50. The probability of the event is

$$p = [z^{100}] \exp\left(\frac{z}{1} + \frac{z^2}{2} + \cdots + \frac{z^{50}}{50}\right) = 1 - \sum_{j=51}^{100} \frac{1}{j} \simeq 0.31182\,78206.$$

Do the prisoners stand a chance against a malicious director who would not place the numbers in drawers at random? For instance, the director might organize the numbers in a cyclic permutation. [Hint: randomize the problem by renumbering the drawers according to a randomly chosen permutation.]

8.2 Average number of cycle

Recall that the class of permutation can be seen as a set of cycles: $\mathcal{P} = \textsc{Set}(\textsc{Cyc}(\mathcal{Z}))$. We want to count the number of cycles, so the specification becomes $\mathcal{P} = \textsc{Set}(\bullet\textsc{Cyc}(\mathcal{Z}))$. The corresponding BGF is

$$P_c(z,u) = \exp\left(u \log\left(\frac{1}{1-z}\right)\right) = (1-z)^{-u} \text{ while } P(z) = \frac{1}{1-z} = \sum \frac{n! z^n}{n!}.$$

The average number of cycles is given by

$$\mathbb{E}_n[\text{number of cycles}] = \frac{n![z^n]\frac{d}{du}P_c(z,u)|_{u=1}}{n![z^n]P(z)} = [z^n]\left.\frac{d}{du}P_c(z,u)\right|_{u=1}.$$

$$\Omega(z) := \frac{d}{du}P_c(z,u)|_{u=1} = \log\left(\frac{1}{1-z}\right)\exp\left(u\log\left(\frac{1}{1-z}\right)\right)|_{u=1} = \frac{1}{1-z}\log\left(\frac{1}{1-z}\right).$$

So,

$$\begin{aligned}
\mathbb{E}_n[\text{number of cycles}] = [z^n]\Omega(z) &= [z^n]\left(\sum_i z^i\right)\left(\sum_k \frac{z^k}{k}\right) \\
&= [z^n]\sum_p z^p\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{p}\right) \\
&= \sum_{i=1}^{n}\frac{1}{i} = H_n \sim_{n\to\infty} \log(n).
\end{aligned}$$

8.3 Number of cycles of size r

Let d_r be the number of cycles of size r in a permutation of size n. In the specification of a permutation, we now want to mark only the cycles of size r.

$$\mathcal{P}_{d_r} = \textsc{Set}((\textsc{Cyc}(\mathcal{Z})\backslash\{\textsc{Cyc}_r(\mathcal{Z})\}) + \{\bullet\textsc{Cyc}_r(\mathcal{Z})\}).$$

The corresponding BGF is

$$P_{d_r}(z,u) = \exp\left(\log\left(\frac{1}{1-z}\right) - \frac{z^r}{r} + u\frac{z^r}{r}\right) = \frac{1}{1-z}\exp\left((u-1)\frac{z^r}{r}\right).$$

$[u^k z^n]P_{d_r}(z,u) = \frac{n![u^k z^n]P_{d_r}(z,u)}{n![z^n](1-z)^{-1}}$ is the probability that a permutation of size n has exactly k cycles of size r. This function $P_{d_r}(z,u)$ has a singularity at $z=1$, so using the transfer lemma (Theorem 2),

$$\begin{aligned}
[u^k z^n]P_{d_r}(z,u) &\sim [u^k z^n]\frac{1}{1-z}e^{-1/r}e^{u/r} \\
&\sim e^{-1/r}\left([u^k]e^{u/r}\right)\left([z^n]\frac{1}{1-z}\right) \\
&\sim \frac{1}{k!}\frac{1}{r^k}e^{-1/r}.
\end{aligned}$$

So, we conclude saying the number of cycles of size r in a permutation of size n follows a Poisson law of parameter $\frac{1}{r}$.

$$\mathbb{P}_n[d_r = k] \sim \frac{1}{k!}\frac{1}{r^k}e^{-1/r} \qquad \text{so,} \qquad d_r \sim \mathcal{P}\text{oisson}\left(\frac{1}{r}\right).$$

Remark. (Expected number of cycles of size r)
In order to find this quantity, we have several option. As we know d_r follow a Poisson law of parameter r^{-1} when $n \to \infty$, we can directly say that $\mathbb{E}_n(d_r) \sim r^{-1}$.
Or, we can use the asymptotic of the cumulative function $C_{d_r}(z) = \frac{d}{du} P_{d_r}(z,u)|_{u=1}$.

$$C_{d_r}(z) = \frac{1}{1-z}\frac{z^r}{r} = \frac{1}{r}\frac{z^r}{1-z} .$$

So,

$$\mathbb{E}_n(d_r) = \frac{n![z^n]C_{d_r}(z)}{n!} = [z^n]C_{d_r}(z) = \frac{1}{r}[z^{n-r}]\frac{1}{1-z} = \frac{1}{r}, \qquad \text{for } r \in \{1, \ldots n\} .$$

This expression is exact, so it is possible to conclude on the average number of cycles in a permutation:

$$\mathbb{E}_n[\text{number of cycles}] = \sum_{r=1}^{n} \mathbb{E}_n(d_r) = \sum_{r=1}^{n} \frac{1}{r} \sim_{n\to\infty} \log(n) .$$

9 Statistic on mappings (or functional graphs)

This topic is broached in the book in several parts: decomposing the functional graph structure into a symbolic specification is explained in II.5.2 p.129; an analysis of various parameters is explained in VII.3.3 p.462.
We define $\mathcal{M}$ the class of mappings (or functions) by

$$\mathcal{M}_n = \{f : \{1, \ldots, n\} \to \{1, \ldots, n\}\} .$$

We will represent a mapping of $\mathcal{M}_n$ by a graph with n vertices, and there is an edge between two vertices, from i to j, if $f(i) = j$. The class of graphs we obtain is called *functional graphs*, and it can be viewed as graph where every vertex has outdegree 1.
Starting from a vertex x, let us apply several times the function f: x, $f(x)$, $f^2(x)$, $\ldots$ At some point, since the domain is finite, this construction will loop back on itself. Repeating the process for all vertices, we thus construct the whole graph. It is generally composed of several connected components; each component is an oriented cycle of points (possibly reduced at only one point), and at each point of the cycle is hung some (possibly empty) tree structure, where the edges of a tree are oriented in direction of the root. These tree structures are rooted non-planar trees (without order on its children), so they are Cayley trees. The specification derives from this description:

$$\begin{aligned} \mathcal{M} &= \textsc{Set}\left(\textsc{Cyc}\left(\mathcal{T}\right)\right) \\ \mathcal{T} &= \mathcal{Z} \star \textsc{Set}\left(\mathcal{T}\right) \end{aligned}$$

The corresponding generating functions are

$$\begin{aligned} M(z) &= \exp\left(\log\left(\frac{1}{1-T(z)}\right)\right) = \frac{1}{1-T(z)} \\ T(z) &= z \cdot \exp(T(z)) \end{aligned}$$

We study the following statistics on this structure of functional graph:

1. γ_1 is the number of cycles (connected components);
2. γ_2 is the number of cyclic points (vertices of the cycles);
3. γ_3 is the number of points without preimages (leaves of the Cayley trees).

So we will consider three bivariate generating functions, called $M_i(z,u)$ for $i = 1, 2, 3$. The goal of this study is to find the expected value of each parameter γ_i. We know the expression of the expectation:

$$\mathbb{E}_n[\gamma_i] = \frac{n![z^n]C_i(z)}{n![z^n]M(z)} ,$$

where $C_i(z)$ is the corresponding cumulative function $C_i(z) := \frac{d}{du} M_i(z,u)|_{u=1}$. The total number of mappings is $m_n = n^n$, and therefore the expression of the expectation reduces to

$$\mathbb{E}_n[\gamma_i] = \frac{n!}{n^n} [z^n] C_i(z)\,.$$

9.1 Expression of the BGFs

We have to find the symbolic specification for each parameter γ_i.

Number of cycles: γ_1

$$\mathcal{M}_1 = \textsc{Set}\left(\bullet\textsc{Cyc}\left(\mathcal{T}\right)\right) \qquad \text{so} \qquad M_1(z,u) = \exp\left(u \log\left(\frac{1}{1-T(z)}\right)\right)\,.$$

So,

$$C_1(z) = \frac{d}{du} M_1(z,u)\bigg|_{u=1} = \frac{1}{1-T(z)} \log\left(\frac{1}{1-T(z)}\right)\,.$$

Number of cyclic points: γ_2

$$\mathcal{M}_2 = \textsc{Set}\left(\textsc{Cyc}\left(\bullet\mathcal{T}\right)\right) \qquad \text{so} \qquad M_2(z,u) = \exp\left(\log\left(\frac{1}{1-uT(z)}\right)\right)\,.$$

So,

$$C_2(z) = \frac{d}{du} M_2(z,u)\bigg|_{u=1} = \frac{T(z)}{(1-T(z))^2}\,.$$

Number of points without preimages: γ_3

As stated previously, a functional mapping may be viewed as a set of cycles of Cayley trees. These Cayley trees may be reduced to a root-leaf. The leaves of these trees do not have a preimage, except if they are root-leaves, since the latter belong to a cycle; we must therefore take care of removing the root-leaves when counting the points without preimages. $\mathcal{M}_3 = \textsc{Set}\left(\textsc{Cyc}\left(\widehat{\mathcal{T}}\right)\right)$ where $\widehat{T}$ is the class of Cayley trees where the leaves <u>but not the root</u> are marked. Let $\widetilde{\mathcal{T}}$ be the class of Cayley trees where all leaves <u>and the root</u> are marked. The specification is

$$\begin{aligned}
\mathcal{M}_3 &= \textsc{Set}\left(\textsc{Cyc}\left(\widehat{\mathcal{T}}\right)\right)\\
\widehat{\mathcal{T}} &= \widetilde{\mathcal{T}} \ \setminus \{\bullet\mathcal{Z}\}\\
\widetilde{\mathcal{T}} &= \left(\mathcal{Z} \star \textsc{Set}(\widetilde{\mathcal{T}}) \ \setminus \{\mathcal{Z}\}\right) + \{\bullet\mathcal{Z}\}
\end{aligned}$$

The corresponding bivariate generating functions are

$$\begin{aligned}
M_3(z,u) &= \exp\left(\log\left(\frac{1}{1-\widehat{T}(z,u)}\right)\right) = \frac{1}{1-\widehat{T}(z,u)}\\
\widehat{T}(z,u) &= \widetilde{T}(z,u) - uz\\
\widetilde{T}(z,u) &= z\exp\left(\widetilde{T}(z,u)\right) + (u-1)z\,.
\end{aligned}$$

So, the cumulative function can be expressed and we find

$$C_3(z) = \frac{d}{du} M_3(z,u)\bigg|_{u=1} = \frac{zT(z)}{(1-T(z))^3}\,.$$

9.2 Expected values

All three cumulative are expressed in terms of the tree function $T(z)$. The asymptotic behavior is dictated by this function. But, we have already study this function and its singularities (section 3.2, analytic inversion theorem for trees). We know that the dominant singularity of $T(z)$ it at $z = e^{-1}$, and near this singularity, $T(z)$ admits an asymptotic development

$$T(z) \underset{z=e^{-1}}{\sim} 1 - \sqrt{2}\sqrt{1-ez}\,.$$

Number of cycles: γ_1

$$\begin{aligned}\mathbb{E}_n[\gamma_1] = \frac{n!}{n^n}[z^n]C_1(z) &= \frac{n!}{n^n}[z^n]\frac{1}{1-T(z)}\log\left(\frac{1}{1-T(z)}\right)\\ &\sim \frac{n!}{n^n}[z^n]\frac{1}{\sqrt{2}\sqrt{1-ez}}\log\left(\frac{1}{\sqrt{2}\sqrt{1-ez}}\right)\\ &\sim \frac{n!}{n^n}\frac{e^n}{2\sqrt{2}}[z^n]\frac{1}{(1-z)^{1/2}}\log\left(\frac{1}{1-z}\right)\\ &\sim \frac{n!}{n^n}\frac{e^n}{2\sqrt{2}}\frac{n^{-1/2}}{\Gamma(1/2)}\log(n) \sim \boxed{\frac{1}{2}\log(n)}\end{aligned}$$

Number of cyclic points: γ_2

$$\begin{aligned}\mathbb{E}_n[\gamma_2] = \frac{n!}{n^n}[z^n]C_2(z) &= \frac{n!}{n^n}[z^n]\frac{T(z)}{(1-T(z))^2}\\ &\sim \frac{n!}{n^n}[z^n]\frac{1}{2(1-ez)}\\ &\sim \frac{n!}{n^n}\frac{e^n}{2}[z^n]\frac{1}{(1-z)} \sim \boxed{\sqrt{\frac{\pi n}{2}}}\end{aligned}$$

Number of points without preimages: γ_3

$$\begin{aligned}\mathbb{E}_n[\gamma_3] = \frac{n!}{n^n}[z^n]C_3(z) &= \frac{n!}{n^n}[z^n]\frac{zT(z)}{(1-T(z))^3}\\ &\sim \frac{n!}{n^n}[z^n]\frac{e^{-1}}{2\sqrt{2}(1-ez)^{3/2}}\\ &\sim \frac{n!}{n^n}\frac{e^n e^{-1}}{2\sqrt{2}}[z^n]\frac{1}{(1-z)^{3/2}} \qquad \sim \frac{n!e^n e^{-1}}{n^n\cdot 2\sqrt{2}}\frac{n^{1/2}}{\Gamma(3/2)} \sim \boxed{\frac{n}{e}}\end{aligned}$$

10 Probability of being a connected graph

This section is treated as Example II.5 p.138 in [2].

Generating function are used here only as formal objects. Indeed, the functions are implicit and their radius of convergence is 0. However it is still possible to use them in computations.

Let $\mathcal{G}$ be the class of labelled graphs. Take $G \in \mathcal{G}$ a graph with n vertices. We have $\binom{n}{2}$ possible edges, and for each edge, we decide to choose it or not. So the total number of labelled graphs with n vertices is $g_n = 2^{\binom{n}{2}}$. This gives for the generating function:

$$G(z) = \sum_{n\geq 0} 2^{\binom{n}{2}}\frac{z^n}{n!}\,.$$

Let $\mathcal{K}$ be the subclass of $\mathcal{G}$ of connected graphs. As a graph is the set of its connected components, the symbolic method provides the following equation $\mathcal{G} = \textsc{Set}(\mathcal{K})$. With $K(z)$ the EGF of $\mathcal{K}$, this translates to $G(z) = \exp(K(z))$. By inversion, we can formally write

$$K(z) = \log\left(1 + \sum_{n\geq 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right).$$

And using the formal definition of the log, $\log(1+u) = u - u^2/2 + u^3/3 + \ldots$, we can express the number k_n of connected graphs with n vertices as

$$\begin{aligned}
k_n &= n![z^n]K(z) = n![z^n]\log\left(1 + \sum_{n\geq 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right) \\
&= n![z^n]\left(\sum_{n\geq 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right) - \frac{1}{2}n![z^n]\left(\sum_{n\geq 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right)^2 + \frac{1}{3}n![z^n]\left(\sum_{n\geq 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right)^3 + \ldots \\
&= 2^{\binom{n}{2}} - \frac{1}{2}\sum_{n_1+n_2=n}\binom{n}{n_1,\, n_2}2^{\binom{n_1}{2}}2^{\binom{n_2}{2}} + \frac{1}{3}\sum_{n_1+n_2+n_3=n}\binom{n}{n_1,\, n_2,\, n_3}2^{\binom{n_1}{2}}2^{\binom{n_2}{2}}2^{\binom{n_3}{2}} + \ldots
\end{aligned}$$

In these sums, there are only a few dominant terms. Indeed, the sequence $\left(2^{\binom{n}{2}}\right)_n$ increases exponentially:

$$2^{\binom{n+1}{2}} = 2^n 2^{\binom{n}{2}}.$$

So, in the first sum, only the first and the last term are meaningful with regard to the asymptotic; (that is $n_1 = 1$ and $n_2 = n-1$, or $n_1 = n-1$ and $n_2 = 1$). The others terms and the other sums are all included into a $o\left(2^{\binom{n}{2}}2^{-n}\right)$. So,

$$k_n = 2^{\binom{n}{2}}\left(1 - 2n2^{-n} + o(2^{-n})\right).$$

Finally, almost all labelled graphs of size n are connected:

$$\mathbb{P}_n[\text{a graph is connected}] = \frac{k_n}{g_n} \underset{n\to\infty}{\sim} 1 - 2n2^{-n} \underset{n\to\infty}{\longrightarrow} 1.$$

11 Saddle-point method

What can we say about the asymptotic of coefficients of a generating function without singularities ?

Let $f(z) = \sum_{n\geq 0} f_n z^n$ be a generating function with no singularities: it means that $f(z)$ is analytic in $\mathbb{C}$. The only formula we can use is the Cauchy formula for coefficients:

$$f_n = \frac{1}{2i\pi}\oint \frac{f(z)\,dz}{z^{n+1}}.$$

where the integral is evaluated around some contour which encompasses 0. The theory says that any contour around 0 can be used. The saddle-point method relies on a good choice of contour in order to make an approximation, and an asymptotic expansion.
The integrand is $g(z) = \frac{f(z)}{z^{n+1}}$. This function has a pole at $z = 0$. Furthermore, let us assume that f is a $\mathbb{C}$-analytic (or *entire* function) with positive coefficients, and that $g(z)$ grows to infinity when $|z|$ tends to infinity. Let us recapitulate the geography of the problem. The real function $\frac{f(z)}{z^{n+1}}$ has a peak at $z = 0$ and an other peak when $z \to \infty$. Therefore,

between these two peaks, there exists a point ρ where $g'(\rho) = 0$. This point has the smallest height among the points of $[x, g(x)]$, with $x \in]0, \infty[$. At $z = \rho$, the derivative of $g(z)$ viewed as a function of z complex also vanishes. The graph of the function $|g(z)|$ in the neighborhood of $z = \rho$ is looking like a saddle (or a pass in mountains). The point $(\rho, g(\rho))$ is called a *saddle-point* [3]

Definition 10. (Saddle-point.) A saddle-point z_0 of a function f is a point such that $f(z_0) \neq 0$ and $f'(z_0) = 0$.

Saddle-point approximation is used when, along a suitable contour going through the saddle-point, the integrand is negligible except in a small neighborhood of the saddle-point. In this case, it is easy to evaluate contour integrals of the form $\oint e^{h(z)} dz$. Indeed, for such integrals, we locate the saddle-point z_0 where $h'(z_0) = 0$, and then, around this saddle-point, we use the Taylor expansion of $h(z)$

$$h(z) = h(z_0) + \frac{1}{2} h''(z_0)(z - z_0)^2 + O((z - z_0)^3) .$$

So, for the evaluation of the contour integral, we cut the contour into two parts: a part $\mathcal{C}_1$ in a small neighborhood of the saddle-point z_0, and the other part $\mathcal{C}_2$ (the rest of the contour encompassing 0). For the part $\mathcal{C}_1$, we use the Taylor expansion of $h(z)$, then the constant term $e^{h(z_0)}$ can be extracted of the integral, and the rest of the integral is easy to evaluate (directly related to $\int e^{-t^2} dt$). At this point, it is often possible to show that the integral on the part $\mathcal{C}_2$ is exponentially negligible.

Saddle-point technique Let $f(z) = \sum f_n z^n$. Let us note $\exp(h(z)) = \frac{f(z)}{z^{n+1}}$.
Find ζ_n such that $h'(\zeta_n) = 0$, that is

$$\zeta_n \frac{f'(\zeta_n)}{f(\zeta_n)} = n + 1 .$$

This gives an asymptotic expression for the coefficients,

$$f_n \sim \frac{f(\zeta_n)}{\zeta_n^{n+1} \sqrt{2\pi h''(\zeta_n)}} .$$

11.1 Exponential and 1/n!

If $f(z) = e^z$, we already know that $[z^n]f(z) = \frac{1}{n!}$. The function f has no singularity so we can, as an exercise, use the saddle-point method. Let $h(z) = \log \frac{f(z)}{z^{n+1}} = z - (n+1)\log(z)$.
So, $h'(z) = 1 - \frac{n+1}{z}$, and $h''(z) = \frac{n+1}{z^2}$. $h'(\zeta_n) = 0$ implies $\zeta_n = n + 1$.
So, we can deduce an asymptotic for the factorial

$$\frac{1}{n!} \sim \frac{e^{n+1}}{(n+1)^{n+1} \sqrt{2\pi/(n+1)}} .$$

Then, we put one factor $(n + 1)$ inside the square root, put the factor n^n outside, and use the equivalent $(1 + 1/n)^n \sim e$, and we find

$$\frac{1}{n!} \sim \frac{e^n}{n^n \sqrt{2\pi n}} .$$

[3] We assumed that the coefficients of $f(z)$ are positive, which implies that there is only one saddle-point on the real positive axis; as a counter-example, think of $\sin(z)/z^{n+1}$. Moreover, this saddle-point will be dominant; see an example Section 11.2 below. In general, the function $f(z)/z^{n+1}$ has many saddle-points.

11.2 Number of involutions: asymptotics

Remember that the generating function of the involutions is $I(z) = \exp\left(z + \frac{z^2}{2}\right)$. This gives directly

$$\frac{I_n}{n!} = [z^n] \exp\left(z + \frac{z^2}{2}\right) = \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{1}{i!\,(n-2i)!\,2^i}.$$

We want to find an asymptotic equivalent of I_n.
The function $I(z)$ has no singularity, so we use the saddle-point method. Let $\exp(h(z)) = I(z)/z^{n+1}$, which gives

$$h(z) = z + \frac{z^2}{2} - (n+1)\log(z)\,, \quad h'(z) = 1 + z - \frac{n+1}{z}\,, \quad h''(z) = 1 + \frac{n+1}{z^2}\,.$$

The derivative cancels for the roots of $z^2 + z - (n+1)$. The positive saddle-point is $-1/2 + 1/2\sqrt{1+4(n+1)}$. When n tends to infinity, it is sufficient to know an asymptotic equivalent of the saddle-point, namely

$$\zeta_n \sim \sqrt{n} - 1/2\,.$$

We obtain an expression for $[z^n]I(z)/n!$, the probability that a permutation is an involution,

$$\frac{I_n}{n!} \sim \frac{e^{I(\zeta_n)}}{\zeta_n^{n+1}\sqrt{2\pi h''(\zeta_n)}} \sim \frac{e^{n/2+\sqrt{n}-1/4}\, n^{-n/2}}{2\sqrt{\pi n}}.$$

References

[1] Eugene Curtin and Max Warshauer. The Locker Puzzle. *The Mathematical Intelligencer*, 28:28–31, 2006. 10.1007/BF02986999.

[2] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009. Also available electronically from the authors' home pages.

[3] Anna Gál and Peter Miltersen. The Cell Probe Complexity of Succinct Data Structures. In Jos Baeten, Jan Lenstra, Joachim Parrow, and Gerhard Woeginger, editors, *Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 332–344. Springer Berlin / Heidelberg, 2003.

[4] Herbert S. Wilf. *Generatingfunctionology*. Academic Press, Inc., 2nd edition, 1994. Also available electronically from the author's home page.

[5] Peter Winkler. Seven puzzles you think you must not have heard correctly. Preprint, 2006. Paper presented at the Seventh Gathering for Gardner (in honour of Martin Gardner).

Markov chains and Martingales applied to the analysis of discrete random structures.

Cécile Mailler

Nablus, August 18–28, 2014

Contents

1 Discrete time Markov chains **39**
1.1 Definitions and first properties . . . 39
1.2 Stationary probability and reversibility . . . 40
1.3 Recurrence and transience . . . 42
1.4 Ergodic theorems . . . 45

2 Discrete time martingales **46**
2.1 Definitions and first properties . . . 46
2.2 Stopping theorems . . . 48
2.3 Doob's inequalities . . . 50
2.4 Convergence of martingales . . . 50

3 Continuous time Markov processes **53**
3.1 Definitions . . . 53
3.2 Ergodicity . . . 54
3.3 Queues . . . 55

4 Continuous time martingales **56**
4.1 Definitions and first properties . . . 56
4.2 Stopping times . . . 57
4.3 Doob's inequalities . . . 58
4.4 Convergence of continuous time martingales . . . 58

5 Exercises **60**

Introduction

Probability and theoretical Computer Science interact in many ways: from stochastic algorithms such as `ethernet` to analysis of algorithms on average. This course aims at presenting two very classical objects in probability theory: Markov chains and martingales through their applications in Computer Science. Our goal is not to give the complete theory, but only to give definitions, basic results and numerous examples. Not all proofs will be developed.

Let us start with a story. John gets out of a bar in Manhattan and wants to go to his hotel. He his so drunk though, that at each crossing, he does not remember where he comes from and choose one road out of the four at random. The next crossing he visits thus only depends on where he is now and what will be his decision, but it does not depend on the past. This is the heuristic of a Markov chain: *the future only depends on the present and not on the past*. Random walks are the classical example of Markov chains, and we will prove in this course that, John will almost surely reach his hotel in finite time – whereas a drunken fish in a 3D undersea Manhattan would almost surely never find his hotel.

A martingale models a fair game: let us say you play heads-or-tails against you banker. Each time you toss a coin, if its heads, you win one peso, if its tail, you loose one peso. If the coin is fair, your expected wealth after the next toss is equal to your actual wealth. This is the heuristic definition of a martingale.

The course is divided into 4 sections: the two first ones concern discrete time Markov chains and martingales, while the two last ones detail continuous time versions of both objects. The discrete time objects being less intricate, we will study them in full detail. Instead of studying continuous time Markov chains in full generality, we will focus on queuing processes, very useful in Computer Science and which study is more basic. In all sections, our aim will be to state convergence results for the considered stochastic processes.

Prerequisites for this course are elementary probability: in particular conditional expectation, convergence of sequences of random variables. It could also be useful to know about σ-algebras, even if a heuristic description should be enough.

This course does not aim to be exhaustive. Many references are available to go further: one can for example cite the following

[Norris] J. R. Norris: **Markov Chains**. *Cambridge University Press*, 1998.

[Williams] D. Williams: **Probability with Martingales**. *Cambridge University Press*, 1991.

[Steward] W. J. Steward: **Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modelling**. *Princeton University Press*, 2009.

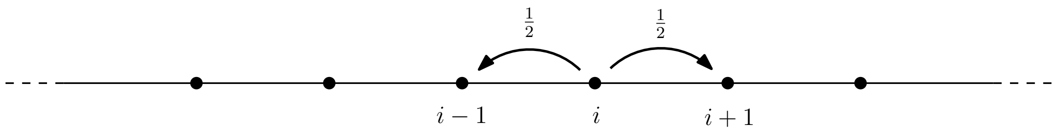


FIGURE 1 – *The simple random walk on $\mathbb{Z}$.*

1 Discrete time Markov chains

1.1 Definitions and first properties

Markov chains can be defined on any space: discrete or continuous. In this course, we will only treat with discrete state spaces, but one has to keep in mind that Markov chains exists as well on $\mathbb{R}$, for example. But in the following, E will always be a discrete space.

Definition 1.1

A matrix $P = (p_{x,y})_{x,y\in E}$ is a **stochastic matrix** *if, for all $x \in E$,*

$$\sum_{y\in E} p_{x,y} = 1.$$

Definition 1.2

Let P be a stochastic matrix on E. A sequence $(X_n)_{n\geq 1}$ of random variables taking value in E is a **Markov chain** *of initial law μ_0 and transition matrix P if*

(i) X_0 has law μ_0, and,

(ii) for all $n \geq 0$, for all $x \in E$,

$$\mathbb{P}(X_{n+1} = x \,|\, X_n, \ldots, X_0) = \mathbb{P}(X_{n+1} = x \,|\, X_n) = p_{X_n,x}.$$

Proposition 1.3

Let $(X_n)_{n\geq 1}$ be a **Markov chain** *of initial law μ_0 and transition matrix P. Then, for all $n \geq 0$, for all $x_0, \ldots, x_n \in E$,*

$$\mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \mu(x_0)p_{x_0,x_1} \ldots p_{x_{n-1},x_n}.$$

Example 1.1: The simple random walk on $\mathbb{Z}$ (cf. Figure 1).

Wild Bill Hickok plays *heads or tails* against his banker. His honesty is so much renowned that his banker allows him an infinite credit: he will eventually pay his dept after arresting some wanted outlaw. At time 0, Bill owns x_0 dollars. Each time Bill tosses a coin, he earns one dollar if its heads and looses one if its tails.

If we denote by X_n the number of dollars Wild Bill owns after he has tossed his n^{th} coin, the sequence $(X_n)_{n\geq 0}$ is a Markov chain on $E = \mathbb{Z}$. Its initial law is $\mu_0 = \delta_{x_0}$ and its transition probabilities are defined as follows: for all $i \in \mathbb{Z}$,

$$\begin{array}{lcll} p_{i,i+1} & = & 1/2 & \\ p_{i,i-1} & = & 1/2 & \\ p_{i,j} & = & 0 \quad \text{for all} & j \notin \{i-1, i+1\}. \end{array}$$

Example 1.2: Umbrellas management in England.

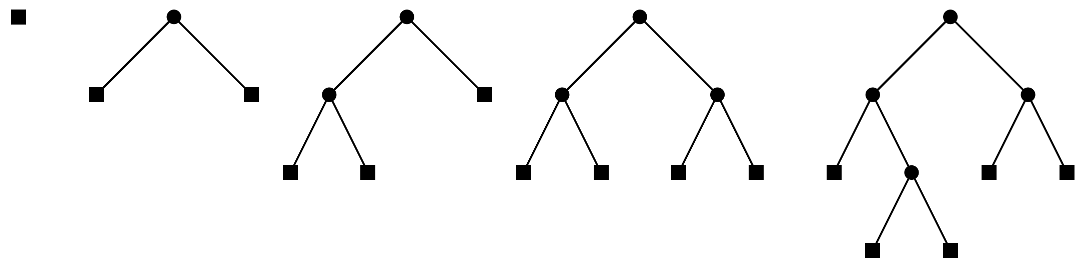


FIGURE 2 – *A realisation of the random BST from time 1 (on the left) to time 5 (on the right).*

I own n umbrellas (n is reasonably large because I live in England). At the beginning of the year, all my umbrellas are at home. Every morning, I go from home to work and every evening from work to home. If it rains when I leave home, and only if it rains, I take one umbrella with me. If it rains when I leave work, and only if it rains, I take one umbrella with me. And each time I leave a building, it rains with probability p (independently).

If we denote by X_n the number of umbrella I have at home at the n^{th} night of the year, then X_n is a Markov chain on $E = \{0, \dots, n\}$. Can you find its probability transitions? For all $i \in \{1, \dots, n-1\}$

$$\begin{array}{lcl} p_{i,i-1} & = & \\ p_{i,i+1} & = & \\ p_{i,i} & = & \\ p_{i,j} & = & 0 \text{ if } j \notin \{i-1, i, i+1\} \end{array}.$$

And don't forget the extremal cases $i = 0$ and $i = n$.

Example 1.3: Ehrenfest's urn

Snowy and Snoopy have fleas: in total, there are N fleas. Each day, a flea chosen at random amongst the N fleas jumps from one dog to the other.

Let us denote by X_n the number of fleas on Snowy on the n^{th} day. The sequence X_n is a Markov chain of transition probabilities

$$\begin{array}{lcl} p_{i,i-1} & = & i/N \\ p_{i,i+1} & = & 1 - i/N \\ p_{i,j} & = & 0 \text{ if } j \notin \{i-1, i+1\} \end{array}$$

Example 1.4: The Binary Search Tree (cf. Figure 2)

The random BST is defined as follows: At time 1, it is a single node. At each step, a leaf of the tree is picked up uniformly at random and becomes an internal node with two leaves as children.

If we denote by T_n the random binary search tree at time n, then $(T_n)_{n\geq 0}$ is a Markov chain on E, the space of binary trees. Can you understand its transition probabilities?

Theorem 1.4 (*Markov property*)

Let (X_n) be a Markov chain of transition matrix P and initial law μ_0. Then, for all $m \geq 1$, $(X_{m+n} \,|\, X_0, \dots, X_m)_{n\geq 0}$ is a Markov chain of transition matrix P and initial law δ_{X_m}.

1.2 Stationary probability and reversibility

Definition 1.5

A probability measure π on E is a **stationary probability** *of a Markov chain of transition matrix P if and only if*

$$\pi P = \pi,$$

i.e. *for all* $x \in E$, $\sum_{y \in E} \pi_y p_{y,x} = \pi_x$.

The existence of such a stationary probability is not guaranteed; it is for example interesting to prove that the simple random walk on $\mathbb{Z}$ does not admit a stationary probability.

Example 1.5: Umbrellas management in England.

The probability transitions of the umbrellas management problem (cf. Example 1.2) are given by: for all $i \in \{1, \dots, N-1\}$

$$\begin{array}{llllll} p_{i,i+1} & = p(1-p) & p_{0,1} & = p & p_{N,N-1} & = p(1-p) \\ p_{i,i-1} & = p(1-p) & p_{0,0} & = 1-p & p_{N,N} & = 1-p(1-p) \\ p_{i,i} & = 1-2p(1-p) & & & & \\ p_{i,j} & = 0 \text{ if } j \notin \{i-1, i, i+1\} & & & & \end{array}$$

Thus, to be a stationary probability of this Markov chain, π has to verify

$$\pi_0 = p(1-p)\pi_1 + (1-p)\pi_0$$

$$\pi_N = p(1-p)\pi_{N-1} + (1-p(1-p))\pi_N$$

and, for all $i \in \{1, \dots, N-1\}$,

$$\pi_i = p(1-p)\pi_{i-1} + (1-2p(1-p))\pi_i + p(1-p)\pi_{i+1}.$$

It implies that

$$\pi_0 = (1-p)\pi_1 \quad \text{and} \quad \pi_N = \pi_{N-1},$$

and, for all $i \in \{1, \dots, N-1\}$, $2\pi_i = \pi_{i-1} + \pi_{i+1}$, which implies

$$\pi_i = \frac{1}{N-p} \quad \text{for all } i \in \{1, \dots, N\} \text{ and } \pi_0 = \frac{1-p}{N-p}.$$

The unique stationary probability of this Markov chain is this *almost uniform law* on $\{0, \dots, N\}$.

Example 1.6: Ehrenfest's urn.

To be a probability distribution on the Ehrenfest's urn defined in Example 1.3, π has to verify:

$$\pi_0 = \frac{1}{N}\pi_1, \quad \pi_N = \frac{1}{N}\pi_{N-1},$$

and, for all $i \in \{1, \dots, N-1\}$,

$$\pi_i = \left(1 - \frac{i-1}{N}\right)\pi_{i-1} + \frac{i+1}{N}\pi_{i+1}.$$

One can check that if, for all $i \in \{0, \dots, N\}$,

$$\pi_i = \frac{1}{2^N}\binom{N}{i},$$

then, π is a stationary probability of the Ehrenfest's urn.

Definition 1.6

A Markov chain of transition matrix P is reversible according to a probability measure π if and only if, for all $x, y \in E$,

$$\pi_x p_{x,y} = \pi_y p_{y,x}.$$

Lemma 1.7

If a Markov chain is reversible according to a probability measure π, then π is an stationary probability of this Markov chain.

Proof. Recall that π is invariant for a Markov chain of transition matrix P if and only if $\pi P = \pi$. Consider a Markov chain of transition matrix P and assume it is reversible according to π. Then,

$$\sum_{y\in E} \pi_y p_{y,x} = \sum_{y\in E} \pi_x p_{x,y} = \pi_x,$$

which implies that π is invariant for the considered Markov chain. □

If $(X_n)_{n\geq 0}$ is a Markov chain reversible according to π and with initial distribution π, then, for all $n \in \mathbb{N}$, the random vectors $(X_0, \dots, X_n)$ and $(X_n, \dots, X_0)$ have the same law.

1.3 Recurrence and transience

Definition 1.8

An **absorbing state** *of a Markov chain $(X_n)_{n\geq 0}$ is a state $x \in E$ such that $p_{x,x} = 1$.*

Let $(X_n)_{n\geq 0}$ be a Markov chain of initial law μ_0 and of transition matrix P. For all $n \geq 1$, let $p_{x,y}^{(n)} = \mathbb{P}(X_n = y|X_0 = x) = \mathbb{P}_x(X_n = y)$. Then, the n^{th} power of the transition matrix P is given by

$$P^n = \left(p_{x,y}^{(n)}\right)_{x,y\in E}.$$

Definition 1.9

A Markov chain of transition matrix $P = (p_{x,y})_{x,y\in E}$ is **irreducible** *if and only if, for all $x, y \in E$, the probability that a Markov chain starting from x eventually reaches y is positive, i.e. if and only if, for all $x, y \in E$, there exists $n \geq 0$ such that $p_{x,y}^{(n)} > 0$.*

The examples of Markov chain introduced in Section 1 are all irreducible, except the Binary Search Tree Markov chain.

The reaching time of a state $x \in E$ is defined and denoted as follows:

$$\tau_x = \inf\{n \geq 1 \,|\, X_n = x\}.$$

Definition 1.10

Let $(X_n)_{n\geq 0}$ be a Markov chain, a state $x \in E$ is

- **recurrent** *for this Markov chain if $\mathbb{P}(\tau_x < +\infty) = 1$;*
- **transient** *for this Markov chain if $\mathbb{P}(\tau_x = +\infty) = 1$.*

A Markov chain is recurrent (resp. transient) if all its states are recurrent (resp. transient).

For all $x \in E$, let us denote by $N_x = \sum_{n\geq 0} \mathbb{1}_{X_n = x}$ the number of visits of the Markov chain $(X_n)_{n\geq 0}$ at state x.

Proposition 1.11

Let $(X_n)_{n\geq 0}$ be a Markov chain of transition matrix P. Then:

(i) If $x \in E$ is transient, then $\mathbb{P}_x(N_x < +\infty) = 1$, $\sum_{n\geq 0} p_{x,x}^{(n)} < +\infty$, and, conditioned on $\{X_0 = x\}$, N_x is a geometric random variable of parameter $\mathbb{P}_x(\tau_x = +\infty)$.

(ii) If x is recurrent then $\mathbb{P}_x(N_x = +\infty) = 1$ and $\sum_{n\geq 0} p^{(n)}_{x,x} = +\infty$.

(iii) If the Markov chain $(X_n)_{n\geq 0}$ is irreducible, then it is either recurrent or transient. In the first case, for all $x \in E$, $\mathbb{P}(N_x = +\infty) = 1$. In the second case, for all $x \in E$, $\mathbb{P}(N_x < +\infty) = 1$.

Proof. First of all, remark that

$$\mathbb{P}_x(\tau_x = +\infty) = \mathbb{P}_x(N_x = 1).$$

For all $m \geq 1$, let us denote by $\tau_x^{(m)}$ the time of the m^{th} visit of the chain into x: $\tau_x^{(1)} := \tau_x$, and

$$\tau_x^{(m)} := \inf\{i > \tau_x^{(m-1)} \mid X_i = x\}.$$

Remark that, for all $m \geq 1$,

$$\begin{aligned}
\mathbb{P}_x(N_x > m) &= \sum_{s\geq m} \mathbb{P}_x(N_x > m \text{ and } \tau_x^{(m)} = s) \\
&= \sum_{s\geq m} \mathbb{P}_x\left(\sum_{i=1}^{s} \mathbb{1}_{X_i=x} = m \text{ and } X_s = x \text{ and } \sum_{i\geq s+1} \mathbb{1}_{X_i=x} > 1\right) \\
&= \sum_{s\geq m} \mathbb{P}_x\left(\sum_{i=1}^{s} \mathbb{1}_{X_i=x} = m \text{ and } X_s = x\right) \mathbb{P}_x\left(\sum_{i\geq 1} \mathbb{1}_{X_i=x} > 1\right) \\
&= \mathbb{P}_x(N_x \geq m)\mathbb{P}_x(N_x > 1).
\end{aligned}$$

Thus, if we denote by $p := \mathbb{P}_x(\tau_x = +\infty) = \mathbb{P}_x(N_x = 1)$, we get, for all $m \geq 0$,

$$\mathbb{P}_x(N_x > m) = (1-p)^m.$$

it immediately implies that

$$\mathbb{P}_x(N_x = m) = p(1-p)^{m-1}.$$

Finally, note that

$$\mathbb{E}N_x = \sum_{i\geq 1} \mathbb{P}_x(X_i = x) = \sum_{i\geq 1} p^{(i)}_{x,x}.$$

(i) If $x \in E$ is transient, then $p > 0$, and conditioned on $\{X_0 = x\}$, N_x is geometrically distributed with parameter p, which implies that its expectation is finite.

(ii) If x is recurrent, then $p = 0$, $\mathbb{P}_x(N_x = +\infty) = 1$ and the expectation of N_x is infinite.

(iii) Let x and y in E. Note that since the chain is irreducible, there exist $n_1, n_2 > 0$ such that $p^{(n_1)}_{x,y} > 0$ and $p^{(n_2)}_{y,x} > 0$. In addition, for all $n \geq 0$,

$$p^{(n+n_1+n_2)}_{y,y} \geq p^{(n_2)}_{y,x} p^{n}_{x,x} p^{(n_1)}_{x,y},$$

which implies that the two series $\sum_{n\geq 1} p^{(n)}_{x,x}$ and $\sum_{n\geq 1} p^{(n)}_{y,y}$ have the same behaviour. Therefore, an irreducible chain is either recurrent or transient.

If the chain is transient, then, for all $x \in E$,

$$\mathbb{P}(N_x = +\infty) = \sum_{s\geq 0} \mathbb{P}(\tau_x = s)\mathbb{P}_x(N_x = +\infty) = 0.$$

The recurrent case is more complicated and left to the reader. □

Example 1.7: The simple random walk on $\mathbb{Z}$ is recurrent (cf. Example 1.1)

For all $n \geq 0$,

$$p_{0,0}^{(2n)} = \binom{2n}{n}\frac{1}{2^{2n}} = \mathtt{Cat}_n 4^{-n},$$

with $\mathtt{Cat}_n = \frac{1}{n+1}\binom{2n}{n}$. Recall that $\mathtt{Cat}_n \sim n^{-3/2}4^n$ when $n \to +\infty$, thus,

$$\sum_{n\geq 0} p_{00}^{(n)} = +\infty,$$

which implies, by Proposition 1.11 lemma that 0 is recurrent. Since the simple random walk is irreducible, we can conclude that the whole chain is recurrent.

Remark: It can be proved that the simple random walk on $\mathbb{Z}^2$ is recurrent as well, but that the simple random walk on $\mathbb{Z}^3$ is transient. In fact, for all $d \geq 3$, the simple walk on $\mathbb{Z}^d$ is transient.

Definition 1.12

Let $(X_n)_{n\geq 0}$ be a Markov chain of transition matrix P. The **period** *of a state $x \in E$ is the* GCD *of $\{n > 0 \,|\, p_{x,x}^{(n)} > 0\}$. A state is said to be* **aperiodic** *if its period is 1 and periodic otherwise. A Markov chain is aperiodic if all its states are aperiodic.*

Proposition 1.13

Let $(X_n)_{n\geq 0}$ be a Markov chain of transition matrix P, then:

(i) If $x \in E$ is aperiodic, then $p_{x,x}^{(n)} > 0$ for all n large enough.

(ii) If $(X_n)_{n\geq 0}$ is irreducible, it is aperiodic as soon as one of its states is aperiodic.

Proof. (i) Assume that $x \in E$ is aperiodic. Let $I = \{n \geq 1 \mid p_{x,x}^{(n)} > 0\}$. Remark that I is stable by addition. There exists $K > 0$, $n_1, \ldots, n_K > 0$ and $a_1, \ldots, a_K \in \mathbb{Z}$ such that $n_i \in I$ for all $i \in \{1, \ldots, K\}$, and

$$1 = \sum_{i=1}^{K} a_i n_i.$$

Let $n_1 = \sum_{a_i>0} a_i n_i$ and $n_2 = -\sum_{a_i<0} a_i n_i$. We know that $n_1, n_2 \in I$ and $n_1 - n_2 = 1$.

Let $n \geq n_2^2$, then, there exists $q \geq n_2$ and $0 \leq r < n_2$ such that

$$n = qn_2 + r = qn_2 + r(n_1 - n_2) = (q - r)n_2 + rn_1,$$

which implies that any $n \geq n_2$ belongs to I.

(ii) Assume that $x \in E$ is aperiodic, then, for all n large enough, $p_{x,x}^{(n)} > 0$. For all $y \in E$, there exists $n_1, n_2 \geq 1$ such that $p_{x,y}^{(n_1)} > 0$ and $p_{y,x}^{(n_2)} > 0$. Thus, for all $n \geq 1$,

$$p_{y,y}^{(n+n_1+n_2)} \geq p_{y,x}^{(n_2)} p_{x,x}^{(n)} p_{x,y}^{(n_1)},$$

which implies that, for all n large enough, $p_{y,y}^{(n)} > 0$ and thus that y is also aperiodic. □

Recall that $\tau_x = \inf\{n \geq 1 \,|\, X_n = x\}$. For all $x \in E$, we define $\nu(x) = \dfrac{1}{\mathbb{E}_x \tau_x} \in [0, 1]$. Remark that if $(X_n)_{n\geq 0}$ is an irreducible, transient Markov chain, then, for all $x \in E$, $\nu(x) = 0$.

Definition 1.14

A recurrent state x of the Markov chain is **positive recurrent** *if $\nu(x) > 0$ and* **null recurrent** *if $\nu(x) = 0$. A Markov chain is called positive (resp. null) recurrent if all its states are positive (resp. null) recurrent.*

Example 1.8: Consider a Markov chain on $\mathbb{N}$ starting at 0 and verifying

$$p_{0,1} = 1, \qquad \text{and } \forall i \geq 1 \quad p_{i,i+1} = \frac{i}{i+1}, \quad p_{i,0} = \frac{1}{i+1};$$

We have

$$\mathbb{P}(\tau_0 = 1) = 0, \quad \mathbb{P}(\tau_0 = n) = \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n} \ (n \geq 2),$$

which implies

$$\sum_{n\geq 0} \mathbb{P}(\tau_0 = n) = 1 \qquad \text{the state 0 is recurrent,}$$

$$\mathbb{E}\tau_0 = \sum_{n=2}^{\infty} n \times \frac{1}{n(n-1)} = \infty \qquad \text{the state 0 is null recurrent.}$$

1.4 Ergodic theorems

An event A is almost sure for a Markov chain if, for all state $x \in E$, $\mathbb{P}_x(A) = 1$, i.e. if $\mathbb{P}(A) = 1$ for any initial distribution μ_0.

Theorem 1.15

Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on E.

(i) $(X_n)_{n\geq 0}$ is either transient, either positive recurrent, or null recurrent.

(ii) If $(X_n)_{n\geq 0}$ is transient or null recurrent, then, she has no invariant probability, and $\nu = 0$.

(iii) For all $x \in E$, we have, almost surely when n tends to infinity,

$$\frac{1}{n} \sum_{m=0}^{n} \mathbb{1}_{X_m = x} \to \nu(x).$$

This result tells you the following: if you are able to exhibit an invariant probability for a Markov chain, then this Markov chain is recurrent. It thus apply for example for the Ehrenfest urn (cf. Example 1.3) or for the umbrellas Markov chain (cf. Example 1.2) which are thus both recurrent. Remark that knowing that a Markov chain admits no invariant probability is not enough to conclude that it is not recurrent: the simple random walk on $\mathbb{Z}$, for example is recurrent but has no stationary distribution.

Theorem 1.16 (*Ergodic Theorem*)

Let $(X_n)_{n\geq 0}$ be an irreducible, positive recurrent Markov chain on E, then:

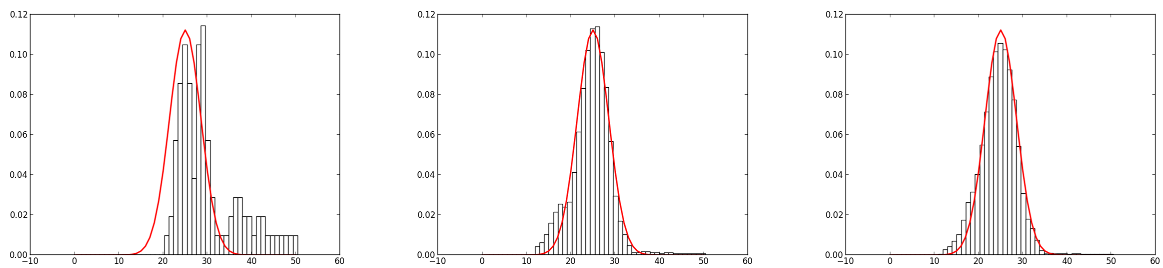
1. *ν is a probability distribution on E and is the unique invariant probability of $(X_n)_{n\geq 0}$. We moreover have that $\nu(x) > 0$ for all $x \in E$.*
2. *For all function $f : E \to \mathbb{R}$ such that $f \geq 0$ or $\int_E f(x) d\nu(x) < +\infty$, we have,*

$$\frac{1}{n} \sum_{m=0}^{n} f(X_k) \to \int_E f(x) d\nu(x).$$

3. *If, in addition, $(X_n)_{n\geq 0}$ is aperiodic, then $X_n \to \nu$ in law when n tends to infinity, and thus, $\mathbb{P}(X_n = x) \to \nu(x)$ for all $x \in E$ when n tends to $+\infty$.*

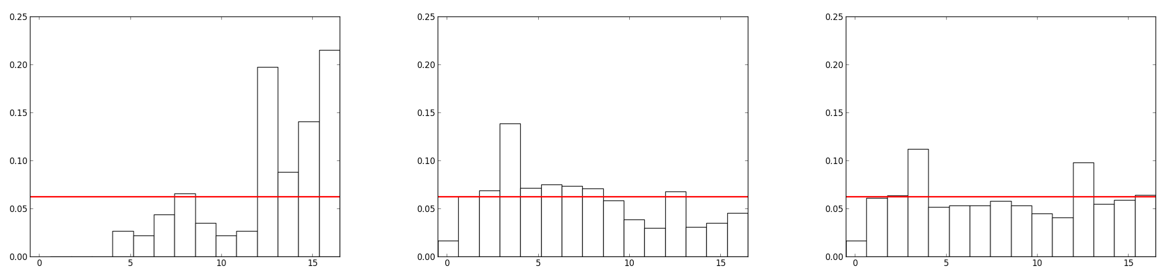
Example 1.9: Ehrenfest's urn.

The Ehrenfest urn is an irreducible positive recurrent chain on $\{1, \dots, N\}$? Therefore, Theorem 1.16 applies as it can be seen on the following simulations: The figure below is the histogram of the number of fleas on Snoopy from time 0 to time 100 (resp. 2000, resp. 5000), when $N = 50$, starting from Snoopy having 50 fleas on it at time 0. The blue curve is the stationary distribution of this Markov chain.



Example 1.10: Umbrellas.

The umbrellas Markov chain described in Example 1.2 is an irreducible positive recurrent chain on $\{1, \dots, N\}$? Therefore, Theorem 1.16 applies as it can be seen on the following simulations: The figure below is the histogram of the number of umbrellas at home between days 1 and 200 (resp. 5000, resp. 10000), when $N = 16$. The red curve is the uniform law on $\{0, \dots, N\}$.



Remark: One can prove that both for a transient and for a null recurrent irreducible Markov chain, $\lim_{n\to+\infty} \mathbb{P}(X_n = x) = 0$.

Corollary 1.17

An irreducible Markov chain on a **finite** *space E is positive recurrent and thus, ν is its unique invariant probability and Theorem 1.16 applies.*

2 Discrete time martingales

2.1 Definitions and first properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Definition 2.1

Let $(\mathcal{F}_n)_{n\geq 0}$ a filtration of Ω, i.e. an increasing family of sub σ-algebras of $\mathcal{F}$. A sequence $(M_n)_{n\geq 0}$ of random

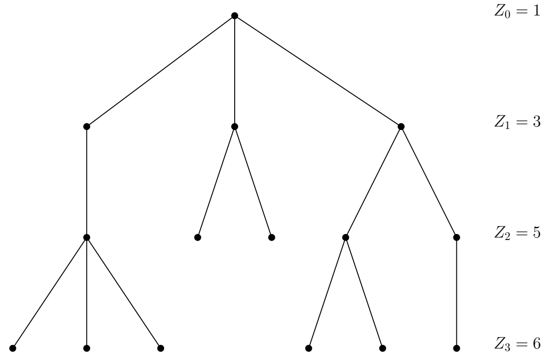


Figure 3 – *A realisation of a Galton-Watson tree.*

*variables is an $\mathcal{F}_n$-**martingale** if, and only if, for all $n \geq 0$,*

(i) M_n is $\mathcal{F}_n$ measurable,

(ii) M_n is integrable, i.e. $\mathbb{E}M_n < +\infty$, and

(iii) $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ almost surely.

In most applications, the considered filtration is $\mathcal{F}_n = \sigma(M_1, \ldots, M_n)$, i.e. contains all the information of the martingale before time n. More generally, given a sequence $(X_n)_{n\geq 0}$ of random variables, we call the filtration $(\mathcal{F}_n = \sigma(X_1, \ldots, X_n))_{n\geq 0}$ its **natural filtration**.

Definition 2.2

If (iii) in Definition 2.1 is replaced by

- *$\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$ a.s., we get the definition of a **super-martingale**.*
- *$\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n$ a.s., we get the definition of a **sub-martingale**.*

Proposition 2.3

Let $(M_n)_{n\geq 0}$ be a $\mathcal{F}_n$-martingale, then, for all $n \geq 0$, $\mathbb{E}M_n = \mathbb{E}M_0$.

What can we say of the sequence $(\mathbb{E}M_n)_{n\geq 0}$ for a super-martingale (resp. sub-martingale)?

Example 2.1: Simple random walk again

Let $(X_n)_{n\geq 0}$ be a sequence of integrable i.i.d. random variables, such that $\mathbb{E}X_1 = 0$. Can you check that $S_n = \sum_{i=1}^n X_i$ is a martingale?

Example 2.2: Galton-Watson tree (cf. Figure 3)

A Galton-Watson tree is described as follows: The first generation is composed of a unique root. Each individual of generation n gives birth to a random number ξ of individuals of generation $n+1$, independently from the rest of the process. We denote by Z_n the number of individuals in generation n: $Z_0 = 1$ and, for all $n \geq 0$,

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)},$$

where the $(\xi_i^{(n)})_{i,n}$ are i.i.d. copies of ξ.

Denote by $m = \mathbb{E}\xi$, then,

$$M_n = m^{-n} Z_n$$

is a martingale.

Example 2.3: The profile of the random Binary Search Tree (cf. Example 1.4)

This exercise is inspired by an article by Chauvin, Klein, Marckert and Rouault (2005): Martingales and Profile of Binary Search Trees, in which martingales are used to get precise information about the shape of the random BST.

Let $\mathcal{T}_n$ be the random BST at time n. For all $n, k \in \mathbb{N}$, let us denote by $N_k(n)$ the number of leaves of $\mathcal{T}_n$ that are at distance k from the root (i.e. at height k in the tree). We denote by $P_n(z)$ the profile polynomial of the BST at time n, given by

$$P_n(z) := \sum_{k\geq 0} N_k(n) z^n.$$

Remark that, if we denote by $|\ell|$ the height of a leaf ℓ of a tree, then

$$P_n(z) = \sum_{\ell\in\mathcal{T}_n} z^{|\ell|}.$$

Can you determine a sequence of rational functions $Z_n(z)$ such that $(M_n := Z_n P_n)_{n\geq 0}$ is a martingale?

Example 2.4: Pólya urn

A Pólya urn is a random process defined by two parameters: an initial composition vector ${}^t(\alpha, \beta)$, and a replacement matrix

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where α, β, a, b, c and d are integers.

We define the sequence of random vectors $(U(n) = (X_n, Y_n))_{n\geq 0}$ representing the composition of a two-colour urn at time t, meaning that the urn contains X_n red balls and Y_n black balls at time n: The urn contains initially α red balls and β black balls. At each step, we pick up uniformly at random a ball in the urn. If the ball is red, we replace it in the urn together with a additional red balls and b black balls. If it is black, we replace it in the urn together with c red balls and d additional black balls.

Let us assume that the urn is balanced, meaning that $a+b = c+d = S$. It implies that the total number of the urn at time n is $X_n + Y_n = \alpha + \beta + nS$. Let

$$Z_n = \left(1 + \frac{A}{\alpha+\beta}\right)^{-1} \dots \left(1 + \frac{A}{\alpha+\beta+(n-1)S}\right)^{-1},$$

where $A =^t R$ (we assume that all the matrices involved in Z_n are indeed invertible). One can then prove that $(M_n := Z_n U(n))_{n\geq 0}$ is a martingale on $\mathbb{R}^2$ for its natural filtration.

2.2 Stopping theorems

Definition 2.4

A **stopping time** *with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$ is a random variable T such that, for all $n \geq 0$, the event $\{T \leq n\}$ is $\mathcal{F}_n$-measurable.*

Example 2.5: Back to Markov chains

Let $(X_n)_{n\geq 0}$ be a Markov chain on a discrete space E. Let $x \in E$, then $\tau_x := \inf\{n \geq 1 \,|\, X_n = x\}$ is a stopping time with respect to the natural filtration of $(X_n)_{n\geq 0}$.

Lemma 2.5

For all martingale $(M_n)_{n\geq 0}$, and for all stopping time T, the **stopped process** *$(M_n^T := M_{n\wedge T})_{n\geq 0}$ is a martingale, (where $\wedge$ denotes the minimum between its two terms).*

This lemma is also true for sub-martingales and super-martingales.

Proof. For all $n \geq 1$,

$$\mathbb{E}[M_{n+1}^T|\mathcal{F}_n] = \mathbb{E}[M_{n+1}\mathbb{1}_{T>n}|\mathcal{F}_n] + \mathbb{E}[M_T\mathbb{1}_{T\leq n}|\mathcal{F}_n].$$

Since $\{T > n\}$ and $\{T \leq n\}$ are both $\mathcal{F}_n$-measurable, we get

$$\mathbb{E}[M_{n+1}^T|\mathcal{F}_n] = \mathbb{E}[M_{n+1}|\mathcal{F}_n]\mathbb{1}_{T>n} + M_T\mathbb{1}_{T\leq n} = M_n\mathbb{1}_{T>n} + M_T\mathbb{1}_{T\leq n} = M_n^T.$$

□

Corollary 2.6

For all martingale $(M_n)_{n\geq 0}$ and for all bounded stopping time T, $\mathbb{E}M_T = \mathbb{E}M_0$.

Definition 2.7

Given a stopping time T, we define its σ-algebra

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid \forall n \geq 0, A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

Of course, one has to check that $\mathcal{F}_T$ is a σ-algebra. We omit this proof.

Proposition 2.8

Let $(M_n)_{n\geq 1}$ be a $\mathcal{F}_n$ martingale and T a finite stopping time. Then M_T is $\mathcal{F}_T$-measurable.

Proposition 2.9

Let T and S two $(\mathcal{F}_n)$-stopping times such that $S \leq T$ almost surely. Then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

Theorem 2.10 (*Doob's stopping theorem*)

Let $(M_n)_{n\geq 0}$ be a martingale, let S and T two bounded stopping times such that, $S \leq T$ almost surely. Then, almost surely,

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S.$$

Proof. It is enough to prove that, for all $A \in \mathcal{F}_S$, $\mathbb{E}[M_T\mathbb{1}_A] = \mathbb{E}[M_S\mathbb{1}_A]$. Let $A \in \mathcal{F}_S$. Define

$$R = S\mathbb{1}_A + T\mathbb{1}_{^cA}.$$

Remark that for all $n \geq 1$,

$$\{R \leq n\} = (A \cap \{S \leq n\}) \cup ({}^cA \cap \{T \leq n\}) \in \mathcal{F}_n.,$$

which implies that R is a bounded stopping time. We thus have $\mathbb{E}M_R = \mathbb{E}M_0 = \mathbb{E}M_T$. Since

$$\mathbb{E}M_T = \mathbb{E}[M_T\mathbb{1}_A + M_T\mathbb{1}_{^cA}]$$

$$\mathbb{E}M_R = \mathbb{E}[M_S\mathbb{1}_A + M_T\mathbb{1}_{^cA}],$$

we get

$$\mathbb{E}[M_T\mathbb{1}_A] = \mathbb{E}[M_S\mathbb{1}_A].$$

□

2.3 Doob's inequalities

Proposition 2.11

Let $(M_n)_{n\geq 0}$ a non-negative sub-martingale such that $\mathbb{E}M_0 < +\infty$. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{i\leq n} M_i \geq \alpha) \leq \frac{\mathbb{E}M_n}{\alpha}$$

Proof. We illustrate the proof by considering the random walk $W = (M_n)_n$ Let us denote $A = \{\max_{i\leq n} M_i \geq \alpha\}$, (so that A is the event "the random walk W went over level α before time n"), and define, for all $k \geq 0$,

$$A_k := \{\max_{i<k} M_i < \alpha \leq M_k\},$$

this last event being "the random walk W went over level α at time k for the first time".

The events A_k are disjoints and we have $A = \bigcup_{k=0}^n A_k$. Therefore

$$\mathbb{E}[M_n \mathbb{1}_A] = \sum_{k=0}^n \mathbb{E}[\mathbb{1}_{A_k} M_n] = \sum_{k=0}^n \mathbb{E}\left[\mathbb{1}_{A_k}\mathbb{E}[M_n|\mathcal{F}_k]\right] = \sum_{k=0}^n \mathbb{E}\left[\mathbb{1}_{A_k} M_k\right] \geq \alpha \sum_{k=0}^n \mathbb{1}_{A_k} = \alpha\mathbb{P}(A).$$

Thus,

$$\mathbb{P}(A) \leq \frac{1}{\alpha}\mathbb{E}[M_n \mathbb{1}_A] \leq \mathbb{E}M_n,$$

since M_n is non-negative. □

The following corollary is a consequence of the following fact: let $(M_n)_{n\geq 0}$ be a martingale and ϕ be a convex function. Then, $(\phi(M_n))_{n\geq 0}$ is a sub-martingale. Apply this property to the convex function $(x \mapsto x^2)$ to get the corollary:

Corollary 2.12

Let $(M_n)_{n\geq 0}$ be a square integrable martingale. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{i\leq n} M_i \geq \alpha) \leq \frac{\mathbb{E}M_n^2}{\alpha^2}.$$

2.4 Convergence of martingales

Definition 2.13

A sequence of random variables $(X_n)_{n\geq 0}$ is **bounded in** L^p *if and only if*

$$\sup \mathbb{E}|X_n|^p < +\infty.$$

The sequence is **uniformly integrable** *if and only if*

$$\lim_{x\to+\infty} \mathbb{E}[X_n \mathbb{1}_{X_n > x}] \to 0,$$

when $x \to +\infty$.

Theorem 2.14

A martingale bounded in L^2 converges in L^2, meaning that there exists a random variable M_∞ such that

$$\lim_{n\to+\infty} \mathbb{E}[|M_n - M_\infty|^2] = 0.$$

Example 2.6: Super-critical Galton-Watson process (cf. Example 2.2).

Let us recall that is Z_n is the number of individuals composing the n^{th} generation in a Galton-Watson process, then $M_n = m^{-n} Z_n$ is a martingale. Let us prove[1] that this martingale is bounded in L^2:

$$\begin{aligned}\mathbb{E}[Z_{n+1}^2|\mathcal{F}_n] &= \mathbb{E}\left[\left(\sum_{i=1}^{Z_n}\xi_n^{(i)}\right)^2\middle|\mathcal{F}_n\right] = \mathbb{E}\left[\sum_{i=1}^{Z_n}\left(\xi_n^{(i)}\right)^2\middle|\mathcal{F}_n\right] + \mathbb{E}\left[\sum_{i\neq j}^{Z_n}\xi_n^{(i)}\xi_n^{(j)}\middle|\mathcal{F}_n\right]\\ &= Z_n \times \mathbb{E}\left[(\xi^{(i)})^2\right] + Z_n(Z_n-1)\times\left(\mathbb{E}\xi^{(i)}\right)^2 = Z_n^2(\mathbb{E}\xi)^2 + Z_n\mathtt{Var}\xi.\end{aligned}$$

This gives

$$\mathbb{E}Z_{n+1}^2 = m^2\mathbb{E}Z_n^2 + m^n\mathtt{Var}\xi, \qquad \text{and thus,} \qquad \mathbb{E}M_{n+1}^2 = \mathbb{E}M_n^2 + m^{-n-2}\mathtt{Var}\xi,$$

which implies that the martingale is bounded in L^2 as soon as $m > 1$, i.e, as soon as the process is super-critical, and assuming that ξ is square-integrable.

Theorem 2.15 (*Doob's Theorem*)

Let $(M_n)_{n\geq 0}$ be a sub-martingale such that

$$\sup_{n\geq 0}\mathbb{E}X_n\mathbb{1}_{X_n\geq 0} < +\infty.$$

Then, M_n converges almost surely to an integrable random variable M_∞.

Corollary 2.16

Any martingale bounded in L^1 converges almost surely to an integrable random variable.

It is very important to note that, in the corollary above, even if the martingale is bounded in L^1 and its almost sure limit is integrable, there is, a priori, no convergence in L^1!

The following corollary is maybe the most useful in practise:

Corollary 2.17

Any non negative super-martingale converges almost surely to an integrable random variable M_∞ and

$$\mathbb{E}M_\infty \leq \liminf_{n\to+\infty}\mathbb{E}M_n.$$

---

[1] We give here an alternate proof using generating functions. We recall that ξ is the law of reproduction of the individuals. Let $\psi(s) = \sum_{i\geq 0}\mathbb{P}(\xi = i)s^i$ be the corresponding probability generating function, and

$$\phi_n(s) = \sum_{k\geq 0}\mathbb{P}(Z_n = k)s^k = \mathbb{E}(s^{Z_n})$$

be the probability generating function of the number of individuals at generation n.

If there are k individuals at generation n, the generating function of individuals at generation $n+1$ is $\psi^k(s)$, by convolution; this corresponds to the substitution $s^k \rightsquigarrow \psi^k(s)$.

Therefore,

$$\phi_{n+1}(s) = \sum_{k\geq 0}\mathbb{P}(Z_n = k)\psi^k(s).$$

By differentiation and evaluation at $s = 1$, we get

$$\begin{aligned}\phi'_{n+1}(1) &= \mathbb{E}(Z_n)\mathbb{E}(\xi),\\ \phi''_{n+1}(1) &= \mathbb{E}(Z_n^2)\mathbb{E}^2(\xi) - \mathbb{E}(Z_n)\mathbb{E}^2(\xi) + \mathbb{E}(Z_n)(\mathbb{E}(\xi^2) - \mathbb{E}(\xi))\end{aligned}$$

But we also have $\phi_{n+1}(s) = \mathbb{E}(s^{Z_{n+1}})$, by construction. Therefore

$$\phi'_{n+1} = \mathbb{E}(Z_{n+1}) \quad \text{and} \quad \phi''_{n+1}(1) = \mathbb{E}(Z_{n+1}^2) - \mathbb{E}(Z_{n+1});$$

moreover $\mathbb{E}(Z_n) = \mathbb{E}(Z_{n-1})\mathbb{E}(\xi) = m\mathbb{E}(Z_{n-1}) = m^n$, which concludes the proof.

Proof. If $(M_n)_{n\geq 0}$ is a super-martingale, then $(-M_n)_{n\geq 0}$ is a sub-martingale. Moreover, it is a non-positive sub-martingale, which implies that

$$\sup_{n\geq 0} \mathbb{E} X_n \mathbb{1}_{X_n \geq 0} = 0 < +\infty.$$

The Doob's Theorem thus applies and $(-M_n)_{n\geq 0}$ converges almost surely to an integrable random variable $-M_\infty$, which concludes the proof. The last inequality is an application of Fatou's lemma. □

Example 2.7: Galton-Watson process (cf. Example 2.2).

Let us recall that is Z_n is the number of individuals composing the n^{th} generation in a Galton-Watson process, then $M_n = m^{-n} Z_n$ is a martingale. It is non-negative and therefore converges almost surely to a random variable M_∞ by Corollary 2.17.

`Exercise: calculate the probability of extinction of a Galton-Watson process.`

Theorem 2.18

Let $(M_n)_{n\geq 0}$ be a martingale. The three following propositions are equivalent:

(i) M_n converges in L^1 to an integrable random variable M_∞;

(ii) $(M_n)_{n\geq 0}$ is bounded in L^1 and there exists a random variable M_∞ such that

$$\mathbb{E}[M_\infty | \mathcal{F}_n] = M_n \qquad (\textit{for all } n \geq 0);$$

(iii) $(M_n)_{n\geq 0}$ is uniformly integrable.

Such a martingale is called **regular**. *It implies in particular that, for all $n \geq 0$, $\mathbb{E} M_n = \mathbb{E} M_\infty$.*

Corollary 2.19

Any martingale bounded in L^p $(p > 1)$ converges almost surely and in L^p.

Proof. Let $(M_n)_{n\geq 0}$ be a martingale bounded in L^p: then, for all $x \geq 0$

$$\mathbb{E}[|M_n|^p] \geq \mathbb{E}[|M_n|^p \mathbb{1}_{M_n \geq x}] + \mathbb{E}[|M_n|^p \mathbb{1}_{M_n < x}] \geq \mathbb{E}[M_n^p \mathbb{1}_{M_n \geq x}] \geq x^{p-1} \mathbb{E}[M_n \mathbb{1}_{M_n \geq x}].$$

Since $(M_n)_{n\geq 0}$ is bounded in L^p, there exists a constant $C > 0$ such that

$$\mathbb{E}[M_n \mathbb{1}_{M_n \geq x}] \leq \frac{C}{x^{p-1}} \to 0$$

when $x \to +\infty$, because $p > 1$. Thus $(M_n)_{n\geq 0}$ is uniformly-integrable and Theorem 2.18 applies: $(M_n)_{n\geq 0}$ is bounded in L^1 and there exists a random variable M_∞ such that

$$\mathbb{E}[M_\infty | \mathcal{F}_n] = M_n \qquad (\text{for all } n \geq 0).$$

By Fatou's lemma,

$$\mathbb{E}[|M_\infty|^p] = \mathbb{E}[\liminf_{n\to+\infty} |M_n|^p] \leq \liminf_{n\to+\infty} \mathbb{E}|M_n|^p \leq K_p,$$

where $K_p < +\infty$ is a constant. Therefore, if we denote by $\|\cdot\|_p$ the L_p-norm ($\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$),

$$\mathbb{E}|M_n - M_\infty|^p \leq (\|M_n\|_p + \|M_\infty\|_p)^p \leq (2K_p^{1/p})^p < +\infty.$$

Therefore, by dominated convergence,

$$\lim_{n\to+\infty} \mathbb{E}|M_n - M_\infty|^p = \mathbb{E} \lim_{n\to+\infty} |M_n - M_\infty|^p = 0,$$

implying that M_n converges to M_∞ in L^p. □

3 Continuous time Markov processes

The aim of this section is not to introduce Markov processes in full generality: we will only focus on jump Markov processes and their main application to queuing theory.

3.1 Definitions

Let E be a discrete state space. Let $(Z_n)_{n\geq 0}$ and $(T_n)_{n\geq 0}$ be two sequences of random variables such that $0 = T_0 \leq T_2 \leq \ldots$, $T_n \to +\infty$ when $n \to +\infty$ and $Z_n \in E$ for all $n \geq 0$.

Definition 3.1

The random function

$$X_t := \sum_{n\geq 0} Z_n \mathbb{1}_{[T_n,T_{n+1}[}(t)$$

is called the **random jump function** *associated to the sequences* $(Z_n)_{n\geq 0}$ *and* $(T_n)_{n\geq 0}$.

Definition 3.2

A random jump function $(X_t)_{t\geq 0}$ *is a* **jump Markov process** *if, for all* $0 < s < t$*, for all* $n \geq 0$*, for all* $t_0 < t_1, \ldots, t_n < s$*, for all* $x_0, x_1, \ldots, x_n, x, y \in E$*,*

$$\mathbb{P}(X_t = y \mid X_{t_0} = x_0, \ldots, X_{t_n} = x_n \text{ and } X_s = x) = \mathbb{P}(X_t = y \mid X_s = x).$$

If, in addition, $P(X_t = y \mid X_s = x)$ *only depends on* x, y *and* $(t - s)$*, then the jump Markov process is called* **homogeneous**.

In the following, we will only consider **homogeneous jump Markov processes**, and we will denote

$$P_{x,y}(t-s) := \mathbb{P}(X_t = y \mid X_s = x).$$

For all $t \geq 0$, the matrix $P(t) = (P_{x,y}(t))_{x,y\in E}$ is the **transition matrix** of the process $(X_t)_{t\geq 0}$ at time t. We denote by $(\mu(t))$ the law of the random variable X_t, for all $t \geq 0$.

Proposition 3.3

Let $(X_t)_{t\geq 0}$ *be a (homogeneous) Markov jump process on* E*, with initial law* $\mu(0) = \mu$ *and transition matrix* $(P(t))_{t\geq 0}$*. Then, for all* $0 < s < t$*,*

(i) $\mu(t) = \mu(t)P(t)$

(ii) $P(s+t) = P(s)P(t)$ *(semi-group condition)*

Example 3.1: Poisson process.
A Poisson process $(N_t)_{t\geq 0}$ is a Markov jump process on $\mathbb{N}$, with transition matrix

$$P_{x,y}(t) = \begin{cases} \dfrac{(\lambda t)^{y-x}}{(y-x)!}\mathrm{e}^{-\lambda t} & \text{if } y \geq x, \\ 0 & \textit{otherwise.} \end{cases}$$

Example 3.2: Let $(T_n)_{n\geq 0}$ be a Poisson point process on $[0, +\infty[$ with intensity λ and let $(Z_n)_{n\geq 0}$ be a discrete time Markov chain on E, of transition matrix P, independent of $(T_n)_{n\geq 0}$. Then, the continuous time process

$$X_t := \sum_{n\geq 0} Z_n \mathbb{1}_{[T_n,T_{n+1}[}$$

is a Markov jump process. Can you determine its transition matrix?

The semi-group property tells us that the transition matrix $(P(t))_{t\geq 0}$ is determined by its values for small $t \geq 0$. Said differently, it is determined by its derivative at 0:

Definition 3.4

Let $(P(t))_{t\geq 0}$ be the transition matrix of a Markov jump process $(X_t)_{t\geq 0}$. Then, there exists $Q = (Q_{x,y})_{x,y\in E}$ called the **generator** *of $(X_t)_{t\geq 0}$, such that*

(i) $Q_{x,y} \geq 0$ *if* $x \neq y$,

(ii) $Q_{x,x} = -\sum_{y\neq x} Q_{x,y} \leq 0$,

(iii) $P_{x,y}(h) = hQ_{x,y} + o(h)$ *when* $h \to 0$, *if* $x \neq y$,

(iv) $P_{x,x}(h) = 1 + hQ_{x,x} + o(h)$ *when* $h \to 0$.

One can see $Q_{x,y}$ as the rate with which the Markov jump process will jump from site x to site y.

Theorem 3.5

Markov property Let $(X_t)_{t\geq 0}$ be a jump Markov process of generator Q. For all real t_0, the process $(X_{t_0+t})_{t\geq 0}$ is a Markov process of initial law $\delta_{X_{t_0}}$.

If we forget time and just focus on the successive positions of the process, we exhibit the underlying Markov chain of the process. Let us denote by τ_n the time of the n^{th} jump of the process: then, the discrete time process $M_n := X_{\tau_n}$ is a Markov chain and its transition matrix $P = (p_{i,j})_{i,j\in E}$ is given by

$$p_{x,y} = \begin{cases} \frac{Q_{x,y}}{q_x} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases},$$

where $q_x := -Q_{x,x}$ for all $x \in E$.

3.2 Ergodicity

A jump Markov process is **irreducible** as soon as its underlying Markov chain is irreducible. It implies that, for all $t > 0$, for all $x, y \in E$, $P_{x,y}(t) > 0$. A state $x \in E$ is recurrent (resp. transient) for the Markov jump process $(X_t)_{t\geq 0}$ if it is recurrent (resp. transient) for its underlying Markov chain.

Theorem 3.6

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and recurrent, with generator $Q = (Q_{x,y})_{x,y\in E}$ and transition matrix $(P(t))_{t\geq 0}$. Then, there exists a unique measure (up to a constant factor) π such that $\pi Q = 0$ and $\pi P(t) = \pi$ for all $t \geq 0$. And this measure π is called an **invariant** *measure of the jump process.*

Definition 3.7

For all $x \in E$, we denote by $\tau_x := \inf\{t > 0 \mid X_t = x\}$. A state $x \in E$ is **positive recurrent** *(resp.* **null recurrent***) for $(X_t)_{t\geq 0}$ if x is recurrent and if $\mathbb{E}_x\tau_x < +\infty$ (resp. $\mathbb{E}_x\tau_x = +\infty$)*

Theorem 3.8

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and recurrent. Then, the following assumptions are equivalent:

(i) $x \in E$ *is positive recurrent,*

(ii) all states are positive recurrent,

(iii) there exists a unique invariant probability distribution π.

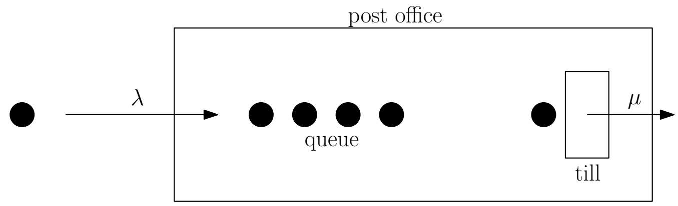


FIGURE 4 – *The $M/M/1$ queue.*

If these assumptions are verified, then, for all $x \in E$,

$$\mathbb{E}_x \tau_x = \frac{1}{\pi_x q_x}.$$

Theorem 3.9

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and positive recurrent. Denote by π its invariant probability. Then, for all bounded function $f : E \to \mathbb{R}$, almost surely, when $t \to +\infty$,

$$\frac{1}{t}\int_0^t f(X_s)ds \to \sum_{x\in E} f(x)\pi_x.$$

Proposition 3.10

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and positive recurrent. Denote by π its invariant probability. Then, for all probability distribution μ on E, for all $x \in E$, asymptotically when $t \to +\infty$,

$$(\mu P(t))_x \to \pi_x.$$

3.3 Queues

The example we will study in the whole section is the queuing theory. It is very important in computer science, since it permits to model routers activity.

The idea is the following: in my post office, there are N tills. People enter the post office according to a Poisson process of rate λ, meaning that the interval between a client and the next one is exponentially distributed with parameter λ, independently from the rest of the process. The time needed to serve a client is exponentially distributed with parameter μ, independently from the rest of the process.

When a client enters the post office: either all tills are occupied and he joins the queue, or one till is free, and he begins to be served as soon as he enters.

This model is usually called $M/M/N$ meaning that the arrivals and service times are exponentially distributed, with respective parameters λ and μ, and that there are N tills.

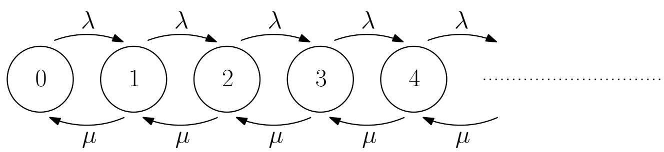
The question is the following: do you need to add more tills so that the length of the queue does not explode? Quite an important question for router, post office or server management.

Example 3.3: The $M/M/1$ queue (cf. Figure 4)

Let us first focus on the case where there is a unique till in the post office. Let X_t be the number of clients inside the post office (queue + till) at time t. Then $(X_t)_{t\geq 0}$ is indeed a Markov process and its generator is the following infinite matrix:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & & \\ \mu & -(\mu+\lambda) & \lambda & 0 & \cdots & \\ 0 & \mu & -(\mu+\lambda) & \lambda & 0 & \cdots \\ & & & \ddots & & \end{pmatrix}.$$

This information can be represented as follows:



It is possible to prove that $\pi_x := \rho^x(1-\rho)$, where $\rho := {}^{\lambda}/_{\mu}$, is an invariant probability of the queue, as soon as $\rho < 1$. If $\rho \geq 1$, then, the queue admits no invariant probability and is thus transient. It means that our queue will explode. Can you calculate the probability that a newly arrived client will have to queue before being served?

Exercise 3.1: Can you give the generator of the queue $M/M/\infty$?

Example 3.4: In the queues described above, the $M/M/N$, the capacity of the queue is infinite, meaning that the queue can become arbitrarily large. One can also describe queues with finite capacity K: the queues $M/M/N/K$. It behaves as the $M/M/N$, except that when the queue is full (i.e. contains K clients), any client arriving to the shop cannot enter the shop and evaporates.

Can you give the generator of such a queue? What is its invariant probability?

4 Continuous time martingales

4.1 Definitions and first properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Definition 4.1

A continuous time process $(M_t)_{t\geq 0}$ is a martingale for the filtration $(\mathcal{F}_t)_{t\geq 0}$ if and only if, for all $t \geq 0$,

(i) M_t is $\mathcal{F}_t$-measurable;

(ii) M_t is integrable; and

(iii) for all $s < t$, $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$.

Definition 4.2

Replacing (iii) in the above definition by

- *for all $s < t$, $\mathbb{E}[M_t|\mathcal{F}_s] \leq M_s$ gives the definition of a* **super-martingale**.
- *for all $s < t$, $\mathbb{E}[M_t|\mathcal{F}_s] \geq M_s$ gives the definition of a* **sub-martingale**.

Example 4.1: The Yule tree (cf. Figure 5)

Let us consider the stochastic process $(Y_t)_{t\geq 0}$ defined as follows. At time zero, there is one particle in the system: $Y_0 = 1$. Each particle dies and gives birth to two new particles after an exponentially distributed random time, independently from the other particles. Let us denote by Y_t the number of particles alive at time t.

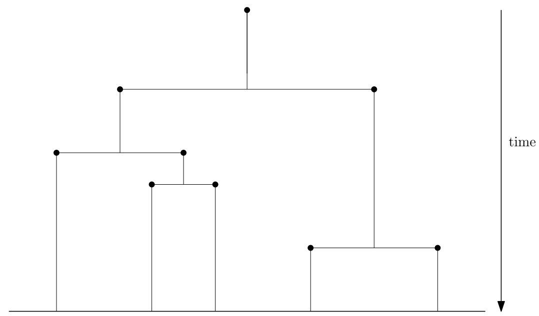


FIGURE 5 – *A realisation of the Yule tree*

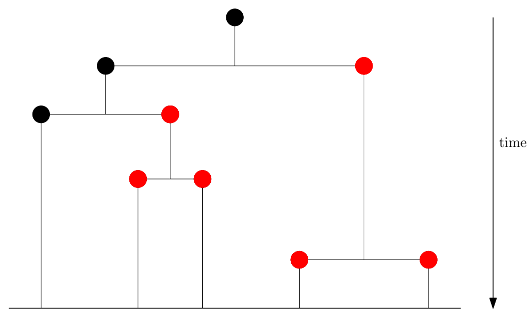


FIGURE 6 – *A realisation of the multi-type branching process defined by the initial composition* ${}^t(0,1)$ *and the replacement matrix* $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Can you find $(m_t)_{t\geq 0}$ a function such that $M_t := m_t^{-1} Y_t$ is a martingale?

Example 4.2: Multi-type branching process (cf. Figure 6)

A multi-type branching process is the embedding in continuous time of a Pólya urn. It is defined by an initial composition $U(0) =^t (\alpha, \beta)$ and a replacement matrix

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The vector composition of the urn at time t is given by $U(t) =^t (X_t, Y_t)$, where X_t is the number of red balls and Y_t the number of black balls at time t in the urn. Each ball in the urn will split after an exponentially distributed random time into

- $a+1$ red balls and b black balls if it is a red ball;
- or c red balls and $d+1$ black balls if it is a black ball,

independently for the other balls.

Assume that the replacement matrix is balanced: $a+b=c+d=S$. What can you say about the total number of balls in the urn at time t? Can you prove that $M_t := \mathsf{e}^{-tA} U(t)$ is a vector valued martingale, where $A =^t R$?

4.2 Stopping times

Definition 4.3

A random variable T is a stopping time for the filtration $(\mathcal{F}_t)_{t\geq 0}$ if and only if, for all $t \geq 0$, the event $\{T \leq t\}$ is $\mathcal{F}_t$-measurable.

Lemma 4.4

For all martingale $(M_t)_{t\geq 0}$, and for all stopping time T, the **stopped process** $(M_t^T := M_{t\wedge T})_{t\geq 0}$ *is a martingale, (where $\wedge$ denotes the minimum between its two terms).*

Theorem 4.5

Stopping theorem Let $(M_t)_{t\geq 0}$ be a martingale, let S and T two bounded stopping times such that, $S \leq T$ almost surely. Then, almost surely,

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S.$$

4.3 Doob's inequalities

Proposition 4.6

Let $(M_t)_{t\geq 0}$ a non-negative sub-martingale such that $\mathbb{E}M_0 < +\infty$. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{s\leq t} M_s \geq \alpha) \leq \frac{\mathbb{E}M_t}{\alpha}$$

Corollary 4.7

Let $(M_t)_{t\geq 0}$ be a square integrable martingale. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{s\leq t} M_s \geq \alpha) \leq \frac{\mathbb{E}M_t^2}{\alpha^2}.$$

4.4 Convergence of continuous time martingales

Definition 4.8

A sequence of random variables $(X_t)_{n\geq 0}$ is **bounded in** L^p *if and only if*

$$\sup_{t\geq 0} \mathbb{E}|X_t|^p < +\infty.$$

The sequence is uniformly integrable if and only if

$$\lim_{x\to+\infty} \sup_{t\geq 0} \mathbb{E}[X_t \mathbb{1}_{X_t > x}] \to 0,$$

when $x \to +\infty$.

Theorem 4.9

A martingale bounded in L^2 converges in L^2, meaning that there exists a random variable M_∞ such that

$$\lim_{t\to+\infty} \mathbb{E}[|M_t - M_\infty|^2] = 0.$$

Theorem 4.10 (*Doob's Theorem*)

Let $(M_t)_{t\geq 0}$ be a sub-martingale such that

$$\sup_{t\geq 0} \mathbb{E}X_t \mathbb{1}_{X_t\geq 0} < +\infty.$$

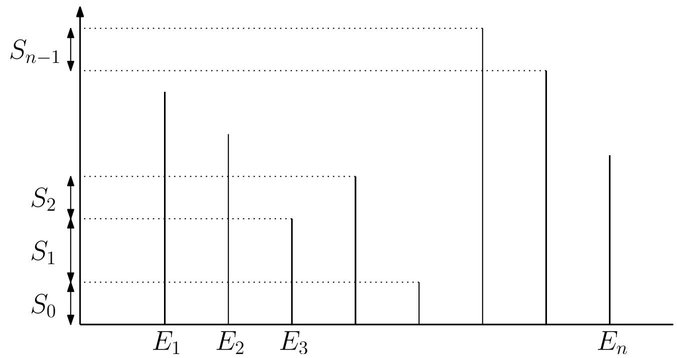


Figure 7 – *The random variables $E_1, \dots, E_n$ are i.i.d. exponentially distributed of parameter 1 and represented by the length of the vertical sticks. Sorting them by increasing order, we get a sequence of random variables $E_n^{(n)}, \dots, E_n^{(1)}$. The S_i verify $S_i = E_n^{(i+1)} - E_n^{(i)}$; they are independent random variables exponentially distributed, of respective parameters $n-i$.*

Then, M_t converges almost surely to an integrable random variable M_∞.

Corollary 4.11

All non negative super-martingale $(M_t)_{t\geq 0}$ converges almost surely to an integrable random variable M_∞ and

$$\mathbb{E}M_\infty \leq \liminf_{t\to+\infty} \mathbb{E}M_t.$$

Example 4.3: The Yule tree martingale (cf. Example 4.1)

The process $(M_t) := (\mathsf{e}^{-t} Y_t)$ is a non negative martingale and thus converges almost surely to a limit random variable W. Let us prove that this random variable is exponentially distributed.

For all $t \geq 0$, $\mathbb{P}(Y_t \geq n) = \mathbb{P}(\tau_n \leq t)$ where τ_n is the time of the n^{th} split in the Yule process. Remark that, by definition, $\tau_n = \sum_{i=1}^n T_i$ where T_i is exponentially distributed of parameter i and the $(T_i)_{i=1..n}$ are independent of each other.

Let us consider $E_1, \dots, E_n$ being n i.i.d. random variables exponentially distributed of parameter 1 (see Figure 7). We can look backwards to the split times and consider that E_n, the largest E_i, corresponds to the time of the last split; similarly then, E_{i-1} may be seen at the precedent split, and this until the first split. Considering the variables $E_n^{(1)}, \dots, E_n^{(n)}$ defined in Figure 7, we have $S_i = E_n^{(i+1)} - E_n^{(i)}$ (with $E_n^{(0)} = 0$); therefore S_i is distributed as the time separating the ith split from the $(i-1)$th split and is $\text{Exp}(n-i)$, this for i from 0 to $n-1$. Moreover the $(S_i)_{i=1..n}$ are independent of each other . Let us denote by m_n the maximum of the E_i. Remark that $m_n = \sum_{i=0}^{n-1} S_i$.

Thus,

$$\mathbb{P}(\tau_n \leq t) = \mathbb{P}(m_n \leq t) = \mathbb{P}(E_i \leq t; \forall 1 \leq i \leq n) = (1 - \mathsf{e}^{-t})^n,$$

since $m_n \leq t$ implies that $E_i \leq t$ for all i from 1 to n. We then get, for all $t \geq 0$, for all $x \geq 0$,

$$\mathbb{P}(M_t \geq x) = \mathbb{P}(Y_t \geq x\mathsf{e}^t) = (1 - \mathsf{e}^{-t})^{x\mathsf{e}^t} \to \mathsf{e}^{-x},$$

when $t \to +\infty$. Thus, for all $x \geq 0$,

$$\mathbb{P}(W \geq x) = \mathsf{e}^{-x},$$

and W is exponentially distributed of parameter 1.

Theorem 4.12

Let $(M_t)_{t\geq 0}$ be a martingale. The three following propositions are equivalent:

(i) M_t converges in L^1 to an integrable random variable M_∞;

(ii) $(M_t)_{t\geq 0}$ is bounded in L^1 and there exists a random variable M_∞ such that

$$\mathbb{E}[M_\infty|\mathcal{F}_t] = M_t \qquad (\text{for all } t \geq 0);$$

(iii) $(M_t)_{t\geq 0}$ is uniformly integrable.

Such a martingale is called **regular**. *It implies in particular that, for all $t \geq 0$, $\mathbb{E}M_t = \mathbb{E}M_\infty$.*

Corollary 4.13

Any martingale bounded in L^p $(p > 1)$ converges in L^p.

5 Exercises

Exercise 5.1: Simple random walk

Let us consider the biased random walk on $\mathbb{Z}$ defined as follows: choose $p \in (0,1)$ and denote $q = 1-p$; when the walker is in state x, it jumps to $x+1$ with probability p and to $x-1$ with probability q.

(1) Prove that the unbiased random walk on $\mathbb{Z}$ is recurrent but has no invariant probability: it is thus null recurrent.

(2) A gambler enters a casino with a GBP (British Pound) and begins to play *heads or tails* with the casino. The casino has b GBP when the gambler begins to play. The coin is biased and gives *heads* with probability p and *tails* with probability q. The gambler gives one pound to the casino when it's *heads* and the casino gives him one pound when it's *tails*. The game ends when either the gambler or the casino is ruined. What is the probability that the gambler gets ruined?

Hint: Denote by X_n the wealth of the gambler at time n, $\tau_0 := \inf\{s \geq 0 \mid X_s = 0\}$ and $\tau_{a+b} := \inf\{s \geq 0 \mid X_s = a+b\}$. It is a good idea to define $u_x := \mathbb{P}(\tau_0 < \tau_{a+b} \mid X_0 = x)$, for all $x \in \mathbb{Z}$.

Solution. (1) Let us first stay in the general case $p \in (0,1)$ before reducing ourselves to the unbiased case $p = 1/2$. Let us calculate the probability, starting from 0, to be in state 0 at time n. This probability is zero for all odd n. We therefore focus on even values of n. Let m be an integer: the only possibility for a walker, starting from state 0, to be in state 0 after $2m$ steps is having done exactly m steps to the right and m steps to the left. Therefore,

$$\mathbb{P}(X_{2m=0} \mid X_0 = 0) = \binom{2m}{m} p^m q^m.$$

If we denote $p_{0,0}^{(n)}$ the probability, starting from 0, to be in 0 at time n, we have:

$$\sum_{n\geq 1} p_{0,0}^{(n)} = \sum_{m\geq 1} p_{0,0}^{(2m)} = \sum_{m\geq 1} \binom{2m}{m} p^m q^m.$$

In the case of the unbiased random walk, we have $p = q = 1/2$, which implies

$$\sum_{n\geq 1} p_{0,0}^{(n)} = \sum_{m\geq 1} \binom{2m}{m} \frac{1}{2^{2m}},$$

and since, in view of Stirling's formula,

$$\binom{2m}{m} \frac{1}{2^{2m}} \sim \frac{1}{\sqrt{m}},$$

we get

$$\sum_{n\geq 1} p_{0,0}^{(n)} = +\infty,$$

implying that the symmetric random walk is recurrent (see Proposition 1.11).

Now assume that $\pi = (\pi_x)_{x\in\mathbb{Z}}$ is an invariant probability measure of the symmetric random walk. Then, for all $x \in \mathbb{Z}$, (see Definition 1.5)

$$\frac{1}{2}\pi_{x-1} + \frac{1}{2}\pi_{x+1} = \pi_x,$$

implying that $\pi_x = \pi_0$ for all integer x, which is impossible since $\sum_{x\in\mathbb{Z}} \pi_x = 1$. Therefore, the symmetric random walk is null recurrent.

(2) Let us use the notations proposed in the "hint". Our aim is thus to calculate $\mathbb{P}(\tau_0 < \tau_{a+b})$. Note also that

$$\begin{cases} u_0 & = 1 \\ u_x & = pu_{x+1} + qu_{x-1} \text{ for all } 1 \leq x \leq a+b-1 \\ u_{a+b} & = 0 \end{cases}$$

Note that $u_x = pu_{x+1} + qu_{x-1}$ is equivalent to $p(u_{x+1} - u_x) = q(u_x - u_{x-1})$, implying that

$$u_{x+1} = \left[\sum_{i=0}^{x} \left(\frac{1-p}{p}\right)^i\right](u_1 - u_0) + u_0.$$

Taking $x = a+b-1$, we get (after some simplification):

$$u_{a+b} = 1 - p\left(1 - \left(\frac{1-p}{p}\right)^{a+b}\right)(u_1 - 1).$$

but we also know that $u_{a+b} = 0$, which gives

$$1 - u_1 = \frac{1}{p\left(\left(\frac{1-p}{p}\right)^{a+b} - 1\right)}.$$

Therefore,

$$u_x = 1 - \frac{1 - \left(\frac{1-p}{p}\right)^x}{1 - \left(\frac{1-p}{p}\right)^{a+b}},$$

and

$$\mathbb{P}(\tau_0 < \tau_{a+b} \mid X_0 = a) = 1 - \frac{1 - \left(\frac{1-p}{p}\right)^a}{1 - \left(\frac{1-p}{p}\right)^{a+b}}$$

is the probability that the gambler gets ruined. □

Exercise 5.2: The original Pólya urns

Consider the Pólya urn with initial composition vector ${}^t(1,1)$ and replacement matrix I_2. Let us denote by ${}^t(X_n, Y_n)$ the composition vector of the urn process at time n.

(1) Prove that X_n is a Markov chain and give its transition probabilities.

(2) Let $\overline{X}_n = \frac{X_n}{X_n+Y_n} = \frac{X_n}{n+2}$ be the proportion of balls of type 1 in the urn at time n. Prove that $(\overline{X}_n)_{n\geq 0}$ is a martingale.

(3) Prove that $(\overline{X}_n)_{n\geq 0}$ converges almost surely and in L^1 to a limit X_∞.

(4) Let

$$Z_n^{(k)} := \frac{X_n(X_n+1)\cdots(X_n+k-1)}{(n+2)(n+3)\cdots(n+k+1)}.$$

Prove that $(Z_n^{(k)})_{n\geq 0}$ is a martingale for all $k \geq 1$.

(5) Prove that, for all $k \geq 1$, $\mathbb{E}X_\infty^k = \mathbb{E}Z_0^{(k)} = \frac{1}{k+1}$ and deduce from it that X_∞ has uniform law on $[0,1]$.

Solution. (1) First note that at time n, there are $n+2$ balls in the urn (white and blacks). Thus, for all $n \geq 0$, for all $x \geq 0$,

$$\mathbb{P}(X_{n+1} = x+1 \mid X_n = x) = \frac{x}{n+2}$$

$$\mathbb{P}(X_{n+1} = x \mid X_n = x) = \frac{n+2-x}{n+2}$$

(2) For all $n \geq 1$

$$\mathbb{E}[\overline{X}_{n+1}|\mathcal{F}_n] = \frac{X_n}{n+2}\frac{X_n+1}{n+3} + \frac{n+2-X_n}{n+2}\frac{X_n}{n+3} = \frac{X_n}{n+2} = \overline{X}_n.$$

Therefore, $(\overline{X}_n)_{n\geq 1}$ is a martingale.

(3) Note that for all $n \geq 1$, $\overline{X}_n \in [0,1]$, therefore, $(\overline{X}_n)_{n\geq 1}$ is a non-negative, bounded martingale. It is therefore almost surely convergent (see Corollary 2.17), and uniformly integrable implying convergent in L^1 (see Theorem 2.18).

(4) For all $n \geq 0$, for all $k \geq 1$,

$$\mathbb{E}[Z_{n+1}^{(k)}|\mathcal{F}_n] = \overline{X}_n \; \frac{(X_n+1)(X_n+2)\cdots(X_n+k)}{(n+3)(n+4)\cdots(n+k+2)} + (1-\overline{X}_n)\; \frac{X_n(X_n+1)\cdots(X_n+k-1)}{(n+3)(n+4)\cdots(n+k+2)} = Z_n^{(k)},$$

after simplifications, implying that $(Z_n^{(k)})_{n\geq 1}$ is a martingale for all $k \geq 1$.

(5) For all $k \geq 1$, $(Z_n^{(k)})_{n\geq 1}$ is a non-negative, bounded martingale. It is thus almost surely convergent and convergent in L^1 to a random variable $Z_\infty^{(k)}$. Moreover (see Theorem 2.18),

$$\mathbb{E}Z_\infty^{(k)} = \mathbb{E}Z_0^{(k)} = \frac{1}{k+1}.$$

In addition, we know that $\overline{X}_n$ converges almost surely to $\overline{X}_\infty$, implying that $Z_n^{(k)}$ converges almost surely to $\overline{X}_\infty^k$. Therefore, $Z_\infty^{(k)} = \overline{X}_\infty^k$, which concludes the proof, because the uniform law on $(0,1)$ has the same sequence of moments and is determined by them. □

Exercise 5.3: Queue with finite capacity

Let us study the queue $M/M/1/K$, corresponding to a queue with arrivals of rate λ, service times of rate μ, with 1 tills and K maximum places in the queue. The number of customers in the post office is a Markov jump process on $\{0,\dots,K\}$:

(1) write its generator Q and its transition matrix $(P(t))_{t\geq 0}$;

(2) convince yourself that the process is irreducible, and calculate its invariant probability;

(3) what is the average number of customers in the system?

Solution. (1) The generator is given by the following $(K+1)\times(K+1)$ matrix:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & & & \\ \mu & -(\mu+\lambda) & \lambda & 0 & \cdots & & \\ 0 & \mu & -(\mu+\lambda) & \lambda & 0 & \cdots & \\ & & \ddots & \ddots & \ddots & & \\ & & & & \mu & -(\mu+\lambda) & \lambda \\ & & & & & \mu & -\mu \end{pmatrix},$$

and the transition matrix is given by

$$P(t) = \begin{pmatrix} 0 & \lambda & 0 & \cdots & & & \\ \mu & 0 & \lambda & 0 & \cdots & & \\ 0 & \mu & 0 & \lambda & 0 & \cdots & \\ & & \ddots & \ddots & \ddots & & \\ & & & & \mu & 0 & \lambda \\ & & & & & \mu & 0 \end{pmatrix} t.$$

(2) If $\pi = (\pi_x)_{0\leq x\leq K}$ is an invariant probability, then $\sum_{x=0}^{K} \pi_x = 1$ and $\pi Q = 0$ (see Theorem 3.6), i.e.

$$\begin{cases} -\lambda\pi_0 + \mu\pi_1 = 0 \\ \lambda\pi_{x-1} - (\lambda+\mu)\pi_x + \mu\pi_{x+1} = 0 \text{ for all } 1 \leq x \leq K-1 \\ \lambda\pi_{K-1} - \mu\pi_K = 0 \end{cases}$$

Therefore, for all $1 \leq x \leq K-1$,

$$\pi_{x+1} - \pi_x = \frac{\lambda}{\mu}(\pi_x - \pi_{x-1}),$$

which implies

$$\pi_{x+1} = \frac{1-(\lambda/\mu)^{x+1}}{1-\lambda/\mu}\,(\pi_1 - \pi_0) + \pi_0.$$

Recall that $\pi_1 = \lambda/\mu\,\pi_0$, which finally gives

$$\pi_x = (\lambda/\mu)^x \pi_0.$$

Using $\sum_{x=0}^{K} \pi_x = 1$ gives

$$\pi_0 = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}},$$

which concludes the proof.

(3) Therefore, the average number of customers in the post office in the stationary regime is given by

$$\sum_{x=0}^{K} x\pi_x = \sum_{x=0}^{K} x(\lambda/\mu)^x \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}}.$$

We let the simplification exercise to the reader. □

CIMPA Summer School 2014
Random structures, analytic and probabilistic approaches
University An Najah, Nablus (Palestine),
Nicolas Pouyanne

Pólya urn models

— Lecture notes —

Contents

1 Pólya urn: first steps **65**

2 The approach in analytic combinatorics **67**

3 The probabilistic approach **72**
3.1 Introduction: an experimental computational approach 73
3.1.1 Distributions 73
3.1.2 Simulations of trajectories 73
3.1.3 Three urns 74
3.2 Asymptotics of the composition vector, phase transition, figures 77
3.3 Hint of proof 83

1 Pólya urn: first steps

Let R be a 2-dimensional square matrix having integral entries and U_0 a nonzero 2-dimensional (column) vector with nonnegative integral entries:

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad U_0 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

The *Pólya urn process* $(U_n)_{n\in\mathbb{N}}$ with *replacement matrix* R and initial *composition vector* U_0 is in an imaging way defined as follows. An urn contains red and black balls. At time 0, it contains α red balls and β black ones. A ball is drawn uniformly at random from the urn and its colour is checked. If the drawn ball is red, it is replaced into the urn together with a red balls and b black ones; if the drawn ball is black, it is replaced into the urn as well, together with c red balls and d black ones. One get in this way a new composition vector U_1. The random process $(U_n)_{n\in\mathbb{N}}$ is recursively defined by iterating this mechanism.

In this lecture, the following assumptions on R and U_0 are made:
(i) R est *balanced*, *i.e.* $a + b = c + d \geq 1$;

(ii) R is *"tenable"*, *i.e.* $\Big(b,c\geq 0\Big)$ and $\Big(a\leq -1\Longrightarrow a|c \text{ and } a|\alpha\Big)$ and $\Big(d\leq -1\Longrightarrow d|b \text{ and } d|\beta\Big)$.
The balance hypothesis guarantees that the same number of balls $S=a+b=c+d\geq 1$ is added at any step of time. Thanks to the tenability assumption, the process can never extinguish, which means that if a or d is negative, one can always respectively subtract $-a$ or $-d$ balls from the urn.

• In more rigorous terms,

$$(U_n)_{n\in\mathbb{N}}=\left(\begin{array}{c}U_n^{(1)}\\U_n^{(2)}\end{array}\right)_{n\in\mathbb{N}}$$

is the $\mathbb{N}^2\setminus\{0\}$-valued discrete time Markov chain defined by the transition conditional probabilities

$$\left\{\begin{array}{l}\mathbf{P}\left(U_{n+1}=U_n+\binom{a}{b}\Big|U_n\right)=\dfrac{U_n^{(1)}}{U_n^{(1)}+U_n^{(2)}}\ ;\\[2ex]\mathbf{P}\left(U_{n+1}=U_n+\binom{c}{d}\Big|U_n\right)=\dfrac{U_n^{(2)}}{U_n^{(1)}+U_n^{(2)}}.\end{array}\right.\tag{1}$$

The balance assumption implies that $U_n^{(1)}+U_n^{(2)}=\alpha+\beta+nS$ for any n: at any time n, the composition of the urn is random but the total number of balls is deterministic.

• A complete definition of the Pólya urn process as a Markov chain is given by the family

$$\left(\mu_{\binom{x}{y}}\right)_{\binom{x}{y}\in\mathbb{N}^2\setminus\{0\}}$$

of probability measures on $\mathbb{N}^2\setminus\{0\}$ defined by:

$$\forall\binom{x}{y}\in\mathbb{N}^2\setminus\{0\},\qquad \mu_{\binom{x}{y}}=\frac{x}{x+y}\delta_{\binom{x}{y}+\binom{a}{b}}+\frac{y}{x+y}\delta_{\binom{x}{y}+\binom{c}{d}},$$

where δ_P denotes the Dirac measure at P. Notice that the tenability assumption guarantees that $\binom{x}{y}+\binom{a}{b}$ and $\binom{x}{y}+\binom{c}{d}$ belong to $\mathbb{N}^2\setminus\{0\}$ as soon as $\binom{x}{y}$ does.

[Generalisation to any finite number of colour, to random replacement matrices. In the present lecture, we will restrict ourselves to non random replacement matrices.]

Notations (spectral decomposition of R)
Thanks to the balance assumption, S is an eigenvalue of ${}^tR$. By elementary considerations à la Perron-Frobenius, the second eigenvalue $m:=a-c=d-b$ of ${}^tR$ is less than or equal to S. We denote

$$\sigma=m/S\leq 1$$

(note that σ may be negative).
When $(b,c) \neq (0,0)$, let

$$v_1 = \frac{S}{b+c}\begin{pmatrix} c \\ b \end{pmatrix} \quad \text{and} \quad v_2 = \frac{S}{b+c}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The vectors v_1 and v_2 are eigenvectors of ${}^tR$, respectively associated with the eigenvalues S and m. The dual basis (u_1, u_2) of linear eigenforms is given by the formulae

$$u_1(x,y) = \frac{1}{S}(x+y) \quad \text{and} \quad u_2(x,y) = \frac{1}{S}(bx - cy).$$

These vectors and linear forms will be useful later on in the lecture.

Note that in dimension larger than 3, the matrix R is not necessarily diagonalizable, even on $\mathbb{C}$. This fact leads to some intricacy in the statement of the results but in a first approach, one can assume that R is diagonalizable.

2 The approach in analytic combinatorics

The approach by analytic combinatorics is due to Philippe Flajolet and his co-authors Philippe Dumas, Joaquim Gabarró, Helmut Pekari and Vincent Puyhaubert in the 2000's. There are two founding articles, namely [4] et [3].

The very first idea consists in coding the urn composition by a sequence $(W_n)_{n\in\mathbb{N}}$ of finite words written in the 2-letter alphabet $\{\mathtt{r}, \mathtt{b}\}$ ($\mathtt{r}$ for *red*, $\mathtt{b}$ for *black*). The initial composition is coded by

$$W_0 = \mathtt{rr}\ldots\mathtt{rbb}\ldots\mathtt{b} = \mathtt{r}^\alpha \mathtt{b}^\beta.$$

Drawing a ball in the urn amounts to choosing a letter in the word uniformly at random. When the chosen letter is an $\mathtt{r}$, it is replaced in the world by the subword $\mathtt{r}^{a+1}\mathtt{b}^b$; when the chosen letter is a $\mathtt{b}$, it is replaced by $\mathtt{r}^c\mathtt{b}^{d+1}$. Thus, the successive drawings give rise to a sequence of random words

$$W_0, W_1, W_2 \ldots$$

Of course, at any time n, the composition vector U_n can be recovered by counting the number of $\mathtt{r}$'s and the number of $\mathtt{b}$'s in the word W_n.

Definition 1 (Histories of the process)
When n is a natural number, when $\begin{pmatrix} u_0 \\ v_0 \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{N}^2 \setminus \{0\}$, *a* history of length n leading from $\begin{pmatrix} u_0 \\ v_0 \end{pmatrix}$ to $\begin{pmatrix} u \\ v \end{pmatrix}$ *is a sequence of words $W_0 = \mathtt{r}^{u_0}\mathtt{b}^{v_0}, W_1, W_2, \ldots, W_n$ produced in that way, for which W_n contains exactly u letters $\mathtt{r}$ et v letters $\mathtt{b}$.*

Of course, with this coding, because of the balance hypothesis, the word W_n always contains u_0+v_0+nS letters, whatever its history is. The key object of Flajolet's method is the number of these histories: denote by

$$\boxed{H_n\begin{pmatrix} u_0 & u \\ v_0 & v \end{pmatrix}}$$

the number of histories of length n leading from $\binom{u_0}{v_0}$ to $\binom{u}{v}$.

Exercise 1. When $R = \begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix}$, code and count all histories of length 2 leading from $\binom{2}{0}$ to $\binom{4}{4}$.

[One possible solution: start from $W_0 = \mathtt{r}^2$. One can draw a tree of all possibilities: $W_1 \in \{\mathtt{rb}^3\mathtt{r}, \mathtt{r}^2\mathtt{b}^3\}$, then $W_2 \in \{\mathtt{rb}^6\mathtt{r}, \mathtt{r}^3\mathtt{b}^4\mathtt{r}, \mathtt{rbr}^2\mathtt{b}^3\mathtt{r}, \mathtt{rb}^2\mathtt{r}^2\mathtt{b}^2\mathtt{r}, \mathtt{rb}^3\mathtt{rb}^3\}$ or $W_2 \in \{\mathtt{rb}^3\mathtt{rb}^3, \mathtt{r}^2\mathtt{b}^6, \mathtt{r}^4\mathtt{b}^4, \mathtt{r}^2\mathtt{br}^2\mathtt{b}^3, \mathtt{r}^2\mathtt{b}^2\mathtt{r}^2\mathtt{b}^2\}$. Amongst the ten histories of length 2, six of them lead to $\binom{4}{4}$ and four lead to $\binom{2}{6}$: starting from two red balls, the probability that the urn contains four red balls and four black ones after two drawings is $3/5$.
Beware: in the example, the configuration $\mathtt{rb}^3\mathtt{rb}^3$ is reached by two different histories. We count histories, not the different word that are potentially obtained.]

Exercise 2 (this urn is Pólya's original one in his article published in 1930)**.** Whenever $R = SI_2$, compute all numbers H_n, $n \geq 0$.
[This is elementary enumerative combinatorics. Make the picture of a path in $\mathbb{N}^2$ and count the histories that follow each of these paths. For any $(p, q) \in \mathbb{N}^2$ such that $p + q = n$, one gets

$$\begin{aligned} H_n\begin{pmatrix} \alpha & \alpha + pS \\ \beta & \beta + qS \end{pmatrix} &= \binom{n}{p} \alpha\big(\alpha + S\big) \dots \big(\alpha + (p-1)S\big)\beta\big(\beta + S\big) \dots \big(\beta + (q-1)S\big) \\ &= n! S^n \binom{\frac{\alpha}{S} + p - 1}{p} \binom{\frac{\beta}{S} + q - 1}{q} \ ; \end{aligned}$$

all others H_n vanish.]

Exercise 3. For any urn, if $N = \alpha + \beta$, show that the total number of histories of length n starting from $\binom{\alpha}{\beta}$ equals $N\big(N + S\big)\big(N + 2S\big) \dots \big(N + (n-1)S\big) = n! S^n \binom{\frac{N}{S} + n - 1}{n}$.

Generating series (or functions) are central tools in analytic combinatorics. In the case of 2-colour urns, the relevant one is the trivariate generating series of histories: the variable x counts the final number of red balls, the variable y counts the final number of black ones while the variable z counts the length of the history. Thus, the replacement matrix R being given, denote

$$\boxed{H\left(x, y, z \,\middle|\, \begin{matrix} u_0 \\ v_0 \end{matrix}\right) = \sum_{u,v,n \in \mathbb{N}} H_n \begin{pmatrix} u_0 & u \\ v_0 & v \end{pmatrix} x^u y^v \frac{z^n}{n!}.}$$

Exercise 4. For any urn (*i.e.* for any R), $H\left(1, 1, z \,\middle|\, \begin{matrix} u_0 \\ v_0 \end{matrix}\right) = \left(\dfrac{1}{1 - Sz}\right)^{\frac{u_0+v_0}{S}}$.

Exercise 5. For the original urn ($R = SI_2$),

$$H\left(x, y, z \,\middle|\, \begin{matrix} u_0 \\ v_0 \end{matrix}\right) = \frac{x^{u_0} y^{v_0}}{\left(1 - Szx^S\right)^{\frac{u_0}{S}} \left(1 - Szy^S\right)^{\frac{v_0}{S}}}.$$

[Computations on multivariate power series, based on the formula $\frac{1}{(1-X)^N} = \sum_{n\geq 0} \binom{N+n-1}{n} X^n$.]

[Commentary on papers by P. Flajolet *et al.*: pointing an object amounts to make a partial derivative on the generating series; proceeding to a replacement amounts to multiply the series by some appropriate monomial. Such considerations lead to the following "Basic isomorphism", stated and proven in [3].]

Theorem 1 (Flajolet, Dumas, Puyhaubert, 2006)
Let x and y be complex numberes such that $xy \neq 0$. Let $X(t)$ and $Y(t)$ be the solutions of the Cauchy Problem (formal version or analytic version)

$$\begin{cases} \dfrac{dX}{dt} = X^{a+1}Y^b \\ \\ \dfrac{dY}{dt} = X^cY^{d+1} \\ \\ X(0) = x, Y(0) = y. \end{cases} \tag{2}$$

Then, for any initial composition (u_0, v_0), for any z in some small enough neighbour of the origin (analytic version),

$$H\left(x, y, z \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) = X(z)^{u_0}Y(z)^{v_0}.$$

Example 1. Back to the original Pólya urn for which $R = SI_2$: the differential system writes $X' = X^{S+1}, Y' = Y^{S+1}$and can be solved. The solution of the Cauchy Problem is $X(t) = x(1 - Stx^S)^{-1/S}$, $Y(t) = y(1 - Sty^S)^{-1/S}$. Theorem 1 provides a second proof of exercice 5.

Proof of Theorem 1. Consider the following differential operator on 2-variable functions:

$$\mathcal{D} = x^{a+1}y^b\frac{\partial}{\partial x} + x^cy^{d+1}\frac{\partial}{\partial y}.$$

The action of $\mathcal{D}$ on monomials is related to urn histories *via* the formula

$$\begin{aligned} \mathcal{D}\left(x^{u_0}y^{v_0}\right) &= u_0x^{a+u_0}y^{b+v_0} + v_0x^{c+u_0}y^{d+v_0} \\ &= H_1\begin{pmatrix} u_0 & u_0+a \\ v_0 & v_0+b \end{pmatrix} x^{a+u_0}y^{b+v_0} + H_1\begin{pmatrix} u_0 & u_0+c \\ v_0 & v_0+d \end{pmatrix} x^{c+u_0}y^{d+v_0} \end{aligned}$$

which can also be written

$$\mathcal{D}\left(x^{u_0}y^{v_0}\right) = \sum_{u,v\geq 0} H_1\begin{pmatrix} u_0 & u \\ v_0 & v \end{pmatrix} x^uy^v$$

where only two terms of the infinite sum are nonzero. This implies by induction that for any $n \in \mathbb{N}$,

$$\mathcal{D}^n\left(x^{u_0}y^{v_0}\right) = \sum_{u,v\geq 0} H_n\begin{pmatrix} u_0 & u \\ v_0 & v \end{pmatrix} x^uy^v. \tag{3}$$

[Notice that the Markov property of the urn process is expressed in this induction.] Besides, if (X, Y) is a solution of the differential system $X' = X^{a+1}Y^b$, $Y' = X^cY^{d+1}$, then

$$\begin{aligned}\frac{d}{dt}\left(X(t)^{u_0}Y(t)^{v_0}\right) &= u_0X(t)^{a+u_0}Y(t)^{b+v_0} + v_0X(t)^{c+u_0}Y(t)^{d+v_0}\\ &= \mathcal{D}\left(x^{u_0}y^{v_0}\right)\Big|_{\substack{x=X(t)\\ y=Y(t)}}\end{aligned}$$

which extends to an analogous formula for the n-th derivative. Gathering these results leads successively to

$$\begin{aligned}H\left(X(t), Y(t), z \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) &= \sum_{n\geq 0} \mathcal{D}^n\left(x^{u_0}y^{v_0}\right)\Big|_{\substack{x=X(t)\\ y=Y(t)}} \frac{z^n}{n!}\\ &= \sum_{n\geq 0} \frac{d^n}{dt^n}\left(X(t)^{u_0}Y(t)^{v_0}\right)\frac{z^n}{n!}.\end{aligned}$$

Thanks to Taylor Formula at the origin (analytic or formal version), one concludes by

$$H\left(X(t), Y(t), z \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) = X(t+z)^{u_0}Y(t+z)^{v_0}.$$

The final result follows taking the value at the origin ($t = 0$). ∎

When the differential system can be solved, applying Theorem 1 leads to a close form of the H function. When this is possible, one gets very accurate probabilistic consequences on the distribution of the composition of the urn at finite time, or on the asymptotics of the process as well. We give hereunder a couple of examples, essentially drawn from [4] and [3].

Remark. 1- One gets immediately from Theorem 1 that

$$H\left(x, y, z \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) = H\left(x, y, z \,\middle|\, \begin{array}{c} 1 \\ 0 \end{array}\right)^{u_0} H\left(x, y, z \,\middle|\, \begin{array}{c} 0 \\ 1 \end{array}\right)^{v_0}.$$

This formula evokes some (combinatoric) convolution property. It has to be related to the branching property of the continuous time corresponding urn process, that leads to a similar equation on the Fourier transforms of large urns limit laws. See [2]. A direct link between both properties remains an open question.

Example 2. Take the urn having $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ as replacement matrix.. [Friedmann's urn. Talk about the propaganda campaign used by P. Flajolet.] The Cauchy Problem writes

$$\begin{cases} X' = XY \\ Y' = XY \\ X(0) = x, Y(0) = y \end{cases}$$

and can be easily solved. One finds

$$H\left(x, y, z \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) = \left(\frac{x(x-y)}{x - ye^{z(x-y)}}\right)^{u_0} \left(\frac{y(y-x)}{y - xe^{z(y-x)}}\right)^{v_0}.$$

For example, when one starts with a sole red ball, the probability generating function of the number of red balls is

$$\mathbf{E}\left(x^{U_n^{(1)}}\right) = \left[\frac{z^n}{n!}\right] \sum_{n,k} \mathbf{P}(U_n^{(1)} = k) x^k \frac{z^n}{n!} = [z^n]\, H\left(x, 1, z \,\middle|\, \begin{matrix} 1 \\ 0 \end{matrix}\right),$$

since the total number of histories of length n starting from one red ball is $n!$ (see Exercise 3). Using the explicit expression of H, one gets

$$\mathbf{E}\left(x^{U_n^{(1)}}\right) = [z^n]\, \frac{x(x-1)}{x - e^{z(x-1)}}.$$

This function of the z-variable has a simple pôle at $z = \frac{\log x}{x-1}$ as unique singularity. Since this function of the x-variable is analytic at 1, singularity analysis shows that one can apply Hwang's Quasi-power Theorem: the mean and the variance of $U_n^{(1)}$ are both asymptotically proportional to n, and the number of red balls at time n (*i.e.* the random variable $U_n^{(1)}$) satisfies a Law of Large Numbers and a Central Limit Theorem as well (Gaussian distribution).

Exemple 3. This example is the central one in [4]. It deals with the urn process that models the leaves of a 2-3–tree, which is an important search tree algorithm. Its replacement matrix is $\begin{pmatrix} -2 & 3 \\ 4 & -3 \end{pmatrix}$. Here, the Cauchy Problem writes

$$\begin{cases} X' = X^{-1}Y^3 \\ Y' = X^4 Y^{-2} \\ X(0) = x, Y(0) = y. \end{cases} \tag{4}$$

Pose $Z = X^2$; one gets successively $Z' = 2Y^3$ and $Z'' = 6Z^2$. Multiply first the latter equation by Z' then integrate. This leads to show that Z is necessarily a solution of the Cauchy Problem

$$\begin{cases} Z'^2 = 4Z^3 - g_3 \\ Z(0) = x^2 \\ Z'(0) = 2y^3 \end{cases} \tag{5}$$

where $g_3 = 4(x^6 - y^6)$. This equation is solved using the famous and beautiful theory of elliptic functions. Quickly said, let $\wp(z) = \wp(z; 0, -4)$ be the elliptic Weiestrass function, associated to the (so-called) invariants $g_2 = 0$ et $g_3 = -4$: if one denotes

$$\omega = \frac{1}{2} B\left(\frac{1}{6}, \frac{1}{3}\right)$$

(Euler Beta function) and if Λ denotes the hexagonal lattice

$$\Lambda = \omega\left(e^{i\pi/6}\mathbb{Z} + e^{-i\pi/6}\mathbb{Z}\right),$$

then $\wp$ is the meromorphic function of the complex plane defined on the complementary of the lattice Λ by

$$\wp(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda \setminus (0)} \left[\frac{1}{(z+\lambda)^2} - \frac{1}{\lambda^2}\right].$$

The function $\wp$ has a double pôle at any point of Λ and is Λ-periodic (such complex functions are called *doubly periodic*). Modulo Λ, the zeroes of $\wp$ are exactly $\omega/3$ and $2\omega/3$. The theory of holomorphic functions shows that $\wp$ is a solution of (5). There is another way to describe this famous $\wp$: it is the inverse of the elliptic integral that underpins Equation (5). More precisely, if z and w are complex numbers one gets the equivalence

$$\wp(z) = w \Longleftrightarrow z = \int_{[w,\infty]} \frac{d\zeta}{2\sqrt{\zeta^3+1}},$$

where the symbol $[w,\infty]$ denotes any half-line having w as origin, and that do not contain any root of the polynomial ζ^3+1 (the square root denotes here the determination defined by the split plane associated to this half-line). Note for example that the Weierstraß functions, even if they have been defined in the 1860's, are objects of recent interest because they give parametrizations of smooth plane cubics that are central in modern cryptography; here, the pair $(\wp, \wp')$ is a parametrization of the curve $Y^2 = 4X^3 + 4$.

Thus, the solutions of the differential system (4) can be expressed by means of elliptic functions on the hexagonal lattice. Take for instance an urn containing initially 2 red balls and no black ones. Let p_n be the probability that all balls are black at time n. In terms of H functions, this number writes

$$p_n = \frac{1}{n+1}[z^n]H\left(0,1,z \,\middle|\, \begin{matrix} 2 \\ 0 \end{matrix}\right).$$

By solving the Cauchy Problem, one shows that

$$H\left(0,1,z \,\middle|\, \begin{matrix} 2 \\ 0 \end{matrix}\right) = \wp\left(z - \frac{\omega}{3}\right).$$

One concludes by means of singularity analysis: check the pôles of $\wp$ and give an asymptotics of p_n as powers of $3/w \sim 0,7132$.

Remarks. 1- The monomial differential system (2) has a simple first integral: if X and Y are solutions, then $1/X^m - 1/Y^m$ is a (locally) constant function. Writing by this means Y as a function of X and reporting in the system, one gets the inverse abelian integrals described above. All "elliptic urns", *i.e.* all urns for which these abelian integrals are related to curves of genus 1 (*elliptic* curves) are classified in [4].

2- In the case of more than 3 colours, Theorem 1 remains valid. Nevertheless, the efficiency and the preciseness of the beautiful analytic method for urns is darkened by a theoretical obstruction: the monomial differential system is, in general, not integrable in dimension more than 3 (this is a difficult result of differential algebra and algebraic geometry, see final comments and note 11 in [3]).

3 The probabilistic approach

We first adopt two experimental approaches, where the effect of the famous phase transition on urns appears. Then, the results on urns asymptotics are stated. Finally, the methods of proving these asymptotics are evoked.

3.1 Introduction: an experimental computational approach

3.1.1 Distributions

As a first approach, for any urn, consider the probability generating function of the number of (say) red balls at time n, starting from the initial composition $\binom{u_0}{v_0}$:

$$p_n\left(x \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) := \sum_{u\geq 0} \mathbf{P}_{(u_0,v_0)}\left(U_n^{(1)} = u\right) x^u = \mathbf{E}_{(u_0,v_0)}\left(x^{U_n^{(1)}}\right).$$

Since the total number af balls at time n is deterministic, this probability generating function describes the whole distribution of the urn composition at time n. This probability generating function can be expressed by means of H functions: denote by

$$H_n\left(x, y \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) := \sum_{u,v\geq 0} H_n\left(\begin{array}{cc} u_0 & u \\ v_0 & v \end{array}\right) x^u y^v = n!\,[z^n]\,H\left(x, y, z \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)$$

the generating series (it is a 2-variable polynomial) of histories of length n starting from $\binom{u_0}{v_0}$. Then, $H_n\left(1, 1 \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)$ is the total number of histories of length n starting from $\binom{u_0}{v_0}$ (see Exercise 3) and

$$p_n\left(x \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) = \frac{H_n\left(x, 1 \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)}{H_n\left(1, 1 \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)}.$$

Thus, it suffices to compute $H_n\left(x, y \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)$, or even $H_n\left(x, 1 \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)$ to get p_n. But, as shown in the proof of Theorem 1, the bivariate function $H_n\left(x, y \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)$ satisfies Equation (3), namely

$$H_n\left(x, y \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right) = \mathcal{D}^n\left(x^{u_0} y^{v_0}\right).$$

As a matter of consequence, by means of computer algebra, starting from the monomial $x^{u_0}y^{v_0}$, it suffices to make an iteration of the operator $\mathcal{D}$ to get a symbolic expression of the entire function $H_n\left(x, y \,\middle|\, \begin{array}{c} u_0 \\ v_0 \end{array}\right)$. The probability generating function p_n is then extracted by substitutions ($y = 1$ and $x = 1$). By this means, the distribution of red balls at given times can be graphically represented. This is done below for three particular urns and initial compositions.

3.1.2 Simulations of trajectories

Another approach consists in simulating the random successive compositions of an urn. One can by this means have a representation of *trajectories* of the composition vector, namely $\{(n, U_n),\ n = 0, 1, 2\ldots\}$

for different random drawings. Taking only the first coordinate of U_n leads to trajectories of the number of red balls, namely

$$\left\{\left(n, U_n^{(1)}\right),\ n = 0, 1, 2 \ldots\right\}.$$

This is done below for three particular urns and initial compositions.

3.1.3 Three urns

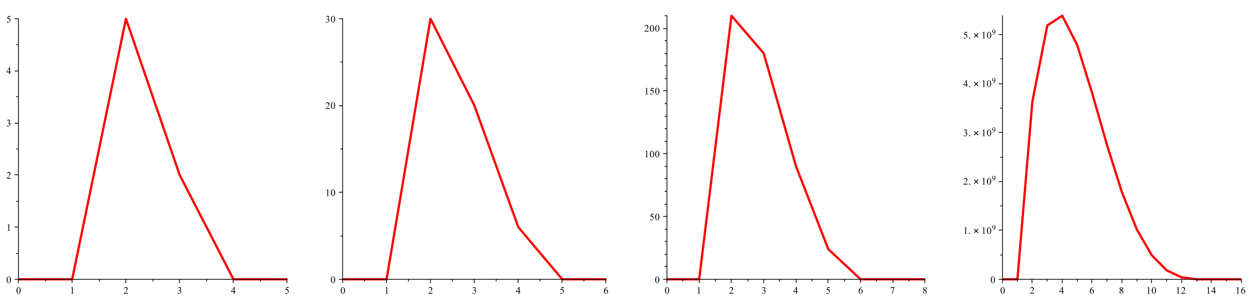
Consider the urn processes having respectively

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R_1 = \begin{pmatrix} 1 & 12 \\ 11 & 2 \end{pmatrix} \text{ and } \quad R_2 = \begin{pmatrix} 12 & 1 \\ 2 & 11 \end{pmatrix}$$

as matrix transitions. The drawings presented hereunder are made taking respectively $\begin{pmatrix} 2 \\ 5 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ as initial composition. All graphics are different representations of the number of red balls contained in the urn.

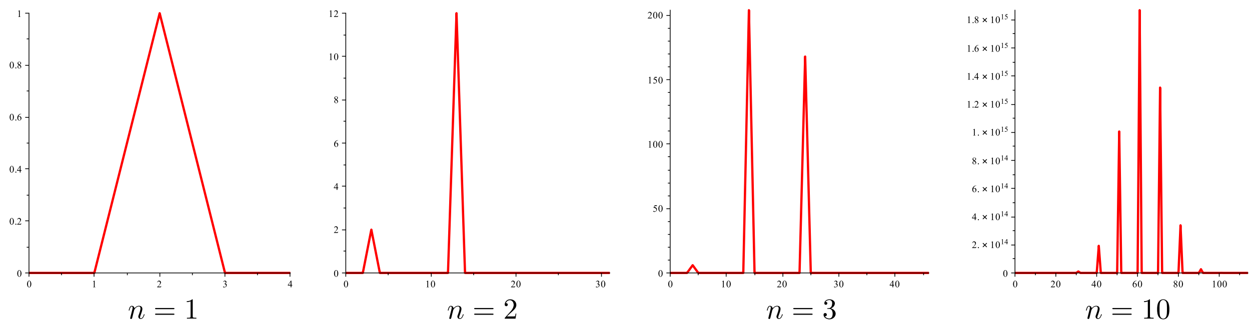
1- Very first histograms

Any picture is made for a given number n of drawings in the urn. On the x-axis, the number of red balls in the urn after n drawings. On the y-axis, the number of histories of length n starting from the initial composition. Points at integer abscissae are related by line segments.

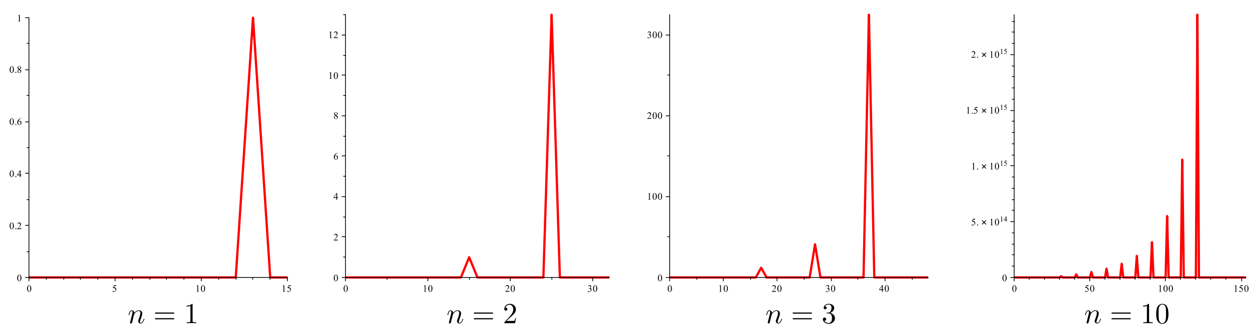


$n = 1$ $\quad$ $n = 2$ $\quad$ $n = 3$ $\quad$ $n = 10$

Original Pólya urn I_2 after n drawings with initial composition $(2, 5)$:
number of histories with respect to the number of red balls.



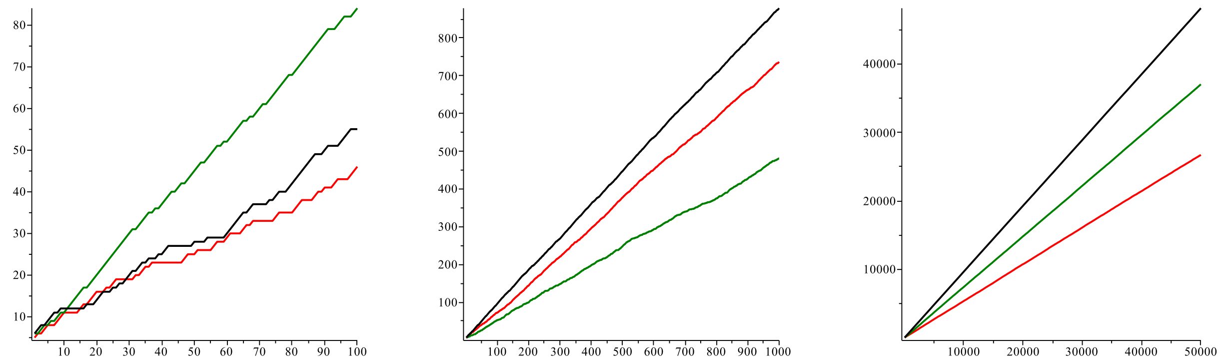
Urn R_1 after n drawings with initial composition $(1, 0)$:
number of histories with respect to the number of red balls.



Urn R_2 after n drawings with initial composition $(1, 0)$:
number of histories with respect to the number of red balls.

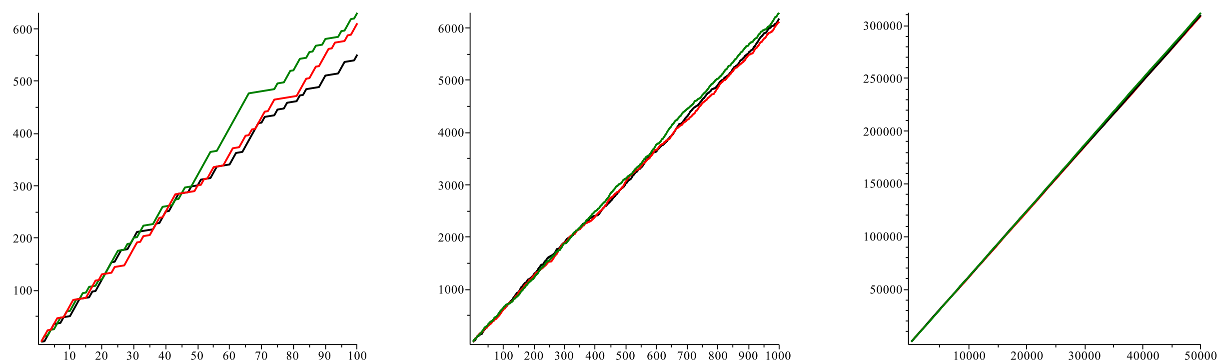
2- Very first trajectories

For a given urn, we draw three different trajectories, corresponding to three random sequences of drawings in the urn. On the x-axis, the number of drawings (discrete time); the maximal number of drawings is successively $N = 100,\ 1000,\ 50000$. On the y-axis, the number of red balls in the urn.



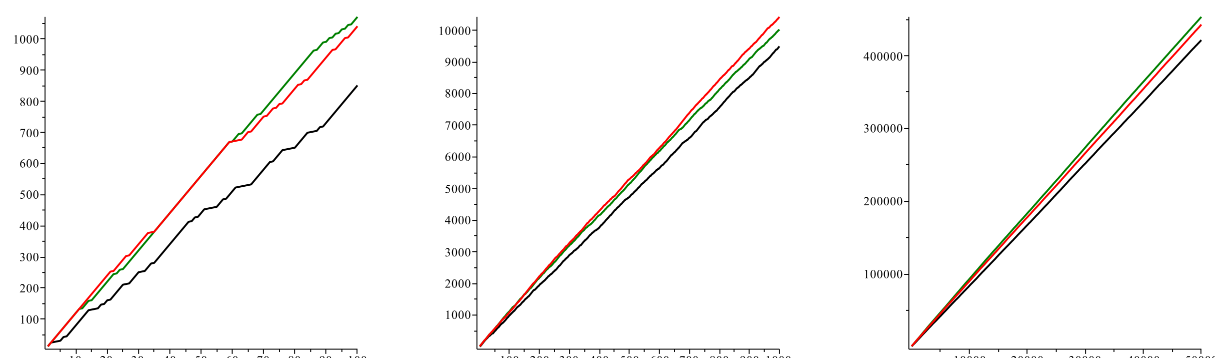
$N = 100$ $N = 1000$ $N = 50000$

Red balls in three sequences of N drawings in an original Pólya urn I_2, initial composition $(2, 5)$



$N = 100$ $N = 1000$ $N = 50000$

Red balls in three sequences of N drawings in an urn R_1, initial composition $(1, 0)$



$N = 100$ $N = 1000$ $N = 50000$

Red balls in three sequences of N drawings in an urn R_2, initial composition $(1, 0)$

3.2 Asymptotics of the composition vector, phase transition, figures

The composition vector U_n of a Pólya urn process has different asymptotics régimes when n tends to infinity, depending on the spectral decomposition of the replacement matrix R. In this section, we state, comment and illustrate these asymptotic results. All of them can be extended in higher dimension (any finite number of colours). Methods of proofs are introduced in Section 3.3.

Take a two-colour Pólya urn with replacement matrix $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and initial composition vector $U_0 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. We adopt the notations of Section 1, especially the *balance* $S = a + b = c + d$, the second eigenvalue $m = a - c = d - b$, the ${}^tR$-eigenvectors v_1, v_2 and, above all, the ratio

$$\sigma = m/S.$$

The original Pólya urn holds a particular place; its asymptotics is described in Theorem 2. The famous phase transition occurs at $\sigma = 1/2$. When $\sigma \leq 1/2$, the urn is said *small* and its composition vector satisfies a central limit theorem as stated in Theorem 3. When $\sigma \in]\frac{1}{2}, 1[$, the urn is said *large* and the centered composition vector admits, after a suitable normalisation, an almost sure random limit; this result is made precise in Theorem 4.

Theorem 2 (Pólya original urn)
Suppose that the urn is Pólya's original one, i.e. *that $R = I_2$. Then, as n tends to infinity,*

$$\frac{U_n}{Sn} \underset{n\to\infty}{\longrightarrow} D$$

almost surely and in any L^p, $p \geq 1$, where D is a Dirichlet distributed 2-dimensional random vector with parameter $\left(\frac{\alpha}{S}, \frac{\beta}{S}\right)$.

If u and v are two positive real numbers, a 2-dimensional Dirichlet distribution with parameter (u, v) is the measure on the simplex $\Sigma = \left\{(x, y) \in [0, 1]^2,\ x + y = 1\right\}$ that admits the function

$$(x, y) \mapsto \frac{\Gamma(u + v)}{\Gamma(u)\Gamma(v)} x^{u-1} y^{v-1}$$

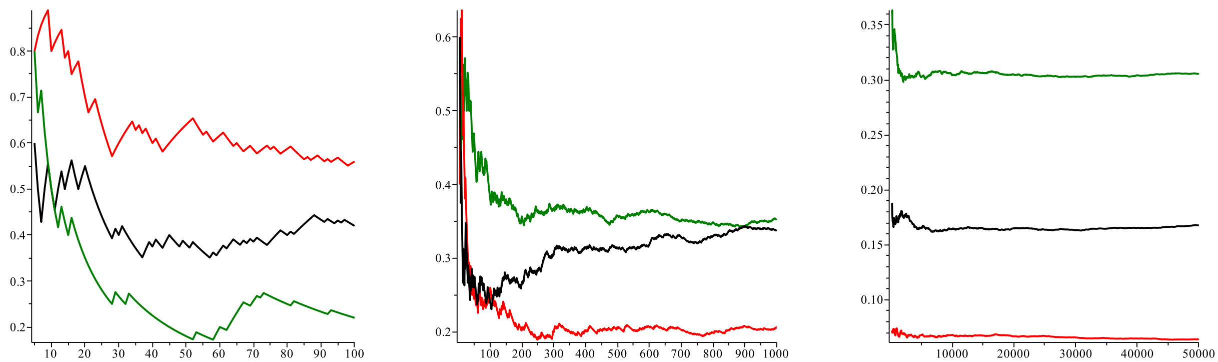
as density with regard to Lebesgue measure on Σ. In other words, if D is a Dirichlet distributed 2-dimensional random vector with parameter (u, v), then for any continuous function f on Σ,

$$\mathbf{E}\left(f(D)\right) = \frac{\Gamma(u + v)}{\Gamma(u)\Gamma(v)} \int_0^1 f(x, 1 - x) x^{u-1} (1 - x)^{v-1} \, dx.$$

In particular, if $D = (X, Y)$, then the *marginals* X and Y are (mutually dependent) Beta distributed random variables, X having parameter (u, v) and Y having parameter (v, u).

Firstly, the convergence is almost sure, which means that, with probability 1, a sequence of random drawings leads to the convergence of the vector U_n/Sn to some vector in the simplex Σ. Secondly,

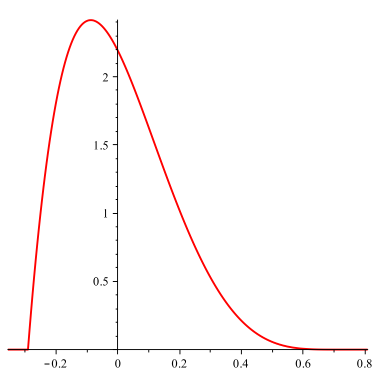
the limit D is random, which means that two different sequences of random drawings converge with probability 1 to two different vectors of Σ.
This almost sure random limit can be visualised on the above simulations: any trajectory gives rise to a (trembled) line, but the three slopes are different. We give hereunder new figures, where three normalised trajectories are represented, showing three different limits: on the x-axis, the number n of drawings up to $N = 100$, 1000 or 50000. On the y-axis, the normalised number of red balls $\frac{1}{n}U_n^{(1)}$.



$N = 100$ $N = 1000$ $N = 50000$

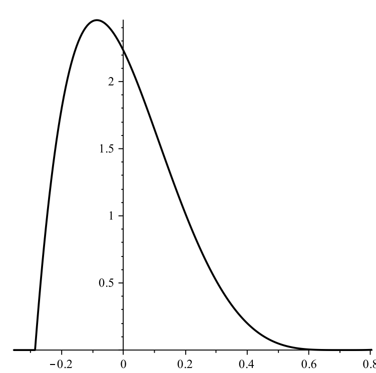
$\frac{1}{n}U_n^{(1)}$ in three sequences of N drawings in an original Pólya urn I_2, initial composition $(2, 5)$

One can also visualise the Beta distributed limit of the normalised number of red balls. Hereunder, the figure on the left represent the (exact) distribution of the normalised number of red balls in the urn after $n = 200$ drawings. On the x-axis, $\frac{1}{n}\left(U_n^{(1)} - \mathbf{E}U_n^{(1)}\right)$. On the y-axis, the probability; it has been computed from the probability generating function p_n introduced above. The figure on the right represents the graph of the density of the centered Beta distribution with parameter $(2, 5)$, namely the function $x \mapsto \frac{1}{B(2,5)}\,(x-\mu)^1\,(1-x+\mu)^4$ where $\mu = B(3,5)/B(2,5) = 2/7$ is the expectation.



$n = 200$

Normalised distribution of the number of red balls in an original Pólya urn I_2, initial composition $(2, 5)$



Density of a centered Beta $(2, 5)$ distribution

Theorem 3 (Small urns)
Suppose that the urn is small, *which means that $\sigma < 1/2$. Then as n tends to infinity,*

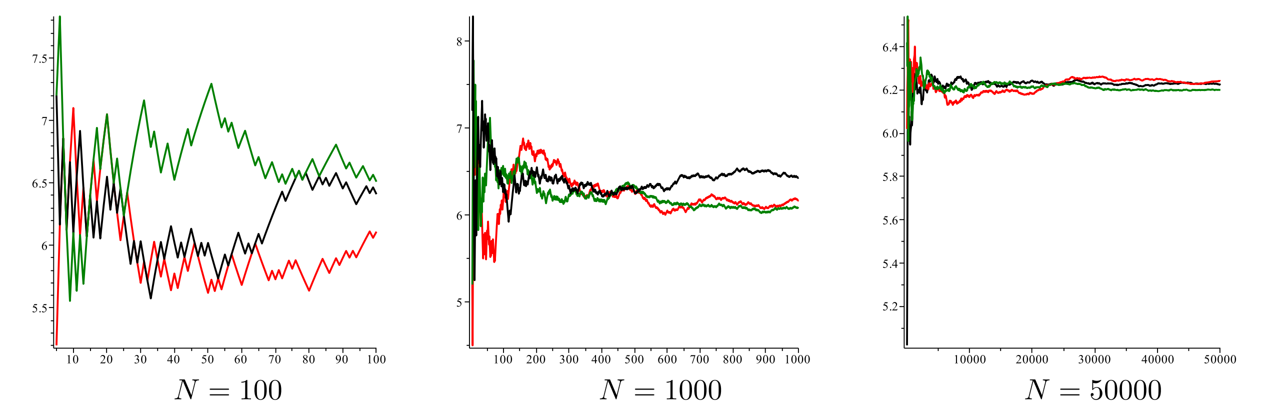
(i) $\dfrac{U_n}{n}$ converges to v_1, almost surely and in any L^p, $p \geq 1$;

(ii) assume further that R is not triangular, i.e. *that $bc \neq 0$. Then, $\dfrac{U_n - nv_1}{\sqrt{n}}$ converges in distribution to a centered gaussian vector with covariance matrix*

$$\frac{1}{1-2\sigma}\frac{bcm^2}{(b+c)^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

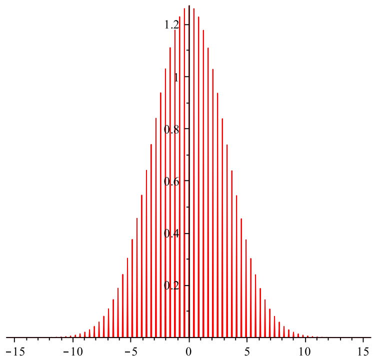
[When $\sigma = 1/2$, one says also that the urn is small. In this case, assertion (i) holds as well whereas, when R is not triangular, assertion (ii) must be replaced by: $\frac{U_n - nv_1}{\sqrt{n\log n}}$ *converges in distribution to a centered Gaussian vector with covariance matrix* $\frac{1}{4}bc\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.]

Here, the convergence of U_n/n is almost sure again, but the limit is deterministic: with probability 1, a sequence of random drawings leads to the convergence of the vector U_n/n, but the limit is now always the same one (namely, v_1). This phenomenon can be visualised on the trajectories for the urn R_1: the three asymptotic slopes are identical. When the normalised trajectories are drawn, one gets the following pictures. Here again, on the x-axis, the number n of drawings up to $N = 100$, 1000 or 50000; on the y-axis, the normalised number of red balls $\frac{1}{n}U_n^{(1)}$.



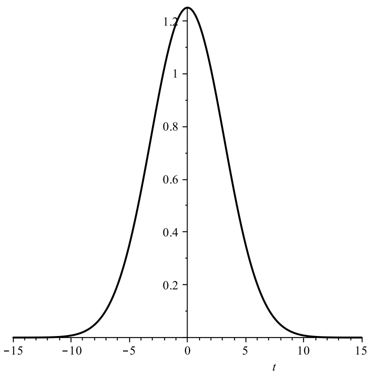
$\frac{1}{n}U_n^{(1)}$ in three sequences of N drawings in a small urn R_1, initial composition $(1, 0)$

The convergence in distribution stated in (ii) is of a radically different nature. It means that the distribution at finite time n converges to some given distribution when n tends to infinity. The limit distribution is here normal. As before, for the R_1-urn, with the help of the probability generating function, the (exact) distribution of the number $\frac{1}{\sqrt{n}}\left(U_n^{(1)} - \mathbf{E}U_n^{(1)}\right)$ is drawn on the leftside figure for $n = 600$. On the right, the graph of the density of the centered normal distribution with variance $\frac{1}{1-2\sigma}\frac{bcm^2}{(b+c)^2} = \frac{5200}{529}$.



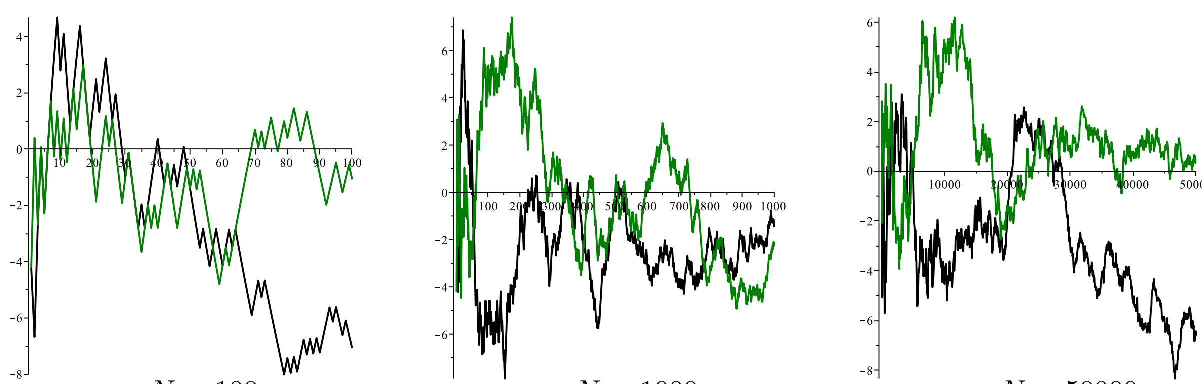
$n = 600$

Normalised distribution of the number of red balls in an small urn R_1, initial composition $(1, 0)$



Density of a centered normal distribution with variance $\frac{5200}{529}$

The difference with almost sure convergence can be visualised on the following trajectory graphs. Even if the distribution at time n converges to a normal distribution, for a given sequence of random drawings, the number $\frac{1}{\sqrt{n}}\left(U_n^{(1)} - \mathbf{E}U_n^{(1)}\right)$ does not converge to a real number. The trajectory is erratic and looks like a brownian motion. On the figure hereunder, two different trajectories of the (completely) normalised number of red balls in a R_1-urn. On the x-axis, the number n of drawings; on the y-axis, $\frac{1}{\sqrt{n}}\left(U_n^{(1)} - nv_1^{(1)}\right)$, where $v_1^{(1)}$ is v_1 first coordinate.



$N = 100$ $\qquad$ $N = 1000$ $\qquad$ $N = 50000$

$\frac{1}{\sqrt{n}}\left(U_n^{(1)} - nv_1^{(1)}\right)$ in two sequences of N drawings in a small urn R_1, initial composition $(1, 0)$

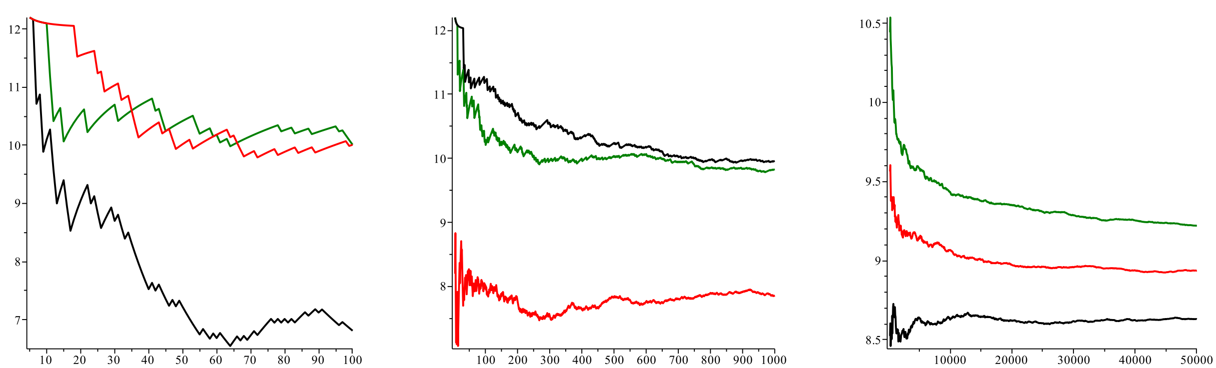
Theorem 4 (Large urns)
Suppose that the urn is large, *which means that* $1/2 < \sigma < 1$. *Then as n tends to infinity,*

(i) $\dfrac{U_n}{n}$ *converges to* v_1, *almost surely and in any* L^p, $p \geq 1$;

(ii) $\dfrac{U_n - nv_1}{n^\sigma}$ *converges almost surely and in any* L^p*,* $p \geq 1$ *to* Wv_2 *where* v_2 *is the (deterministic) eigenvector of* ${}^tR$ *defined in Section 1 and* W *is a real-valued random variable which admits a density and is supported by the whole real line. Besides, with the notations of Section 1,*

$$\mathbf{E}W = \frac{\Gamma\left(\frac{\alpha+\beta}{S}\right)}{\Gamma\left(\frac{\alpha+\beta}{S}+\sigma\right)} \frac{b\alpha - c\beta}{S}.$$

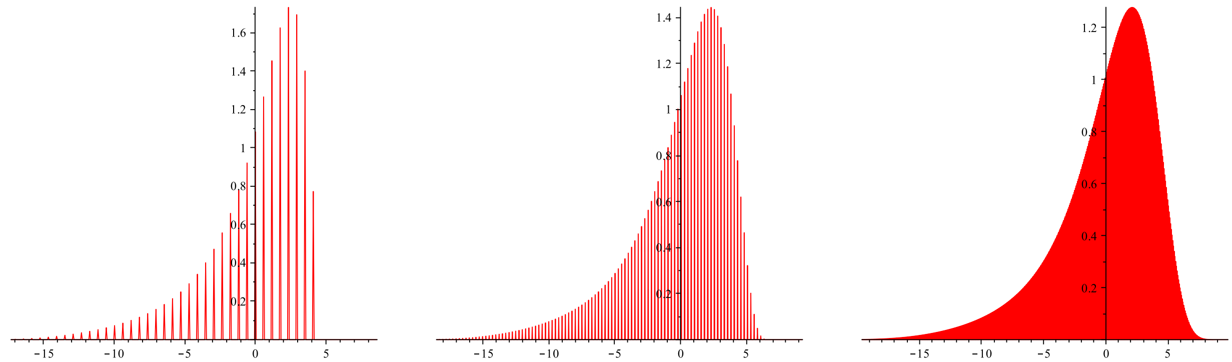
Assertion (i) is the same one as in the case of small urns. We make the same simulations as before for the urn R_2. The convergence to the (same) limit is visibly much slower, due to the second order term which grows like n^σ with $\sigma \simeq 0.77$ (instead of $\sqrt{n}$ for small urns). This second order term was already seeable on the trajectories of the number of red balls: the three slopes do not look not as similar as in the case of the small urn R_1 (but they really tend to a same one as N tends to infinity). Hereunder, again, on the x-axis, the number n of drawings up to $N = 100$, 1000 or 50000; on the y-axis, the normalised number of red balls $\frac{1}{n}U_n^{(1)}$.



$N = 100$ $N = 1000$ $N = 50000$

$\frac{1}{n}U_n^{(1)}$ in three sequences of N drawings in a large urn R_2, initial composition $(1, 0)$

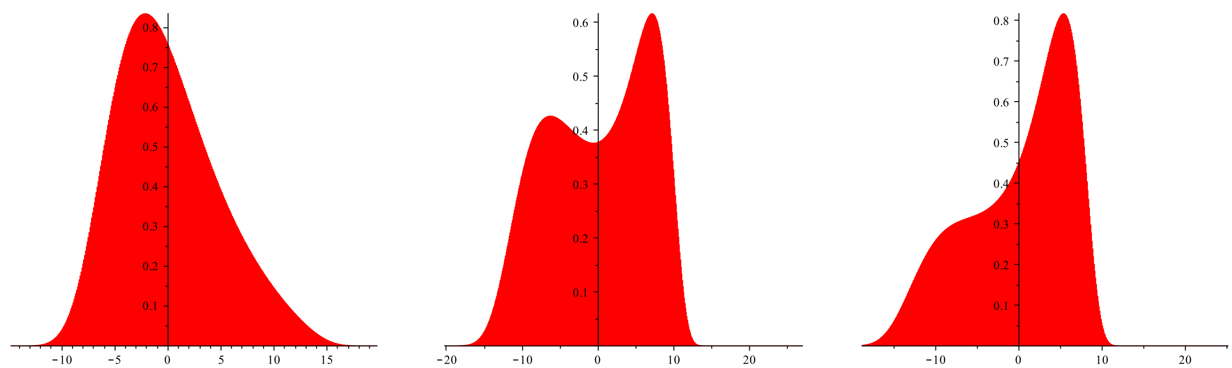
Almost sure convergence implies convergence in distribution. In particular, by formal computation of the probability generating function of red balls, the shape of W's density can be approached as already done (Beta function for the original Pólya urn, Gauss function for a small urn). The Fourier transform of W can be expressed in terms of the inverse of some suitable abelian integral (see [2]). Despite of this, very few is known about its density. The figure hereunder shows the graph of the density of $W - \mathbf{E}W$, approached by the (exact) distribution of $\frac{1}{n^\sigma}\left(U_n^{(1)} - \mathbf{E}U_n^{(1)}\right)$ for $n = 40$, 120 and 800.



$n = 40$ $\qquad$ $n = 120$ $\qquad$ $n = 800$

Normalised distribution of the number of red balls
in a large urn R_2 after n drawings, initial composition $(1, 0)$

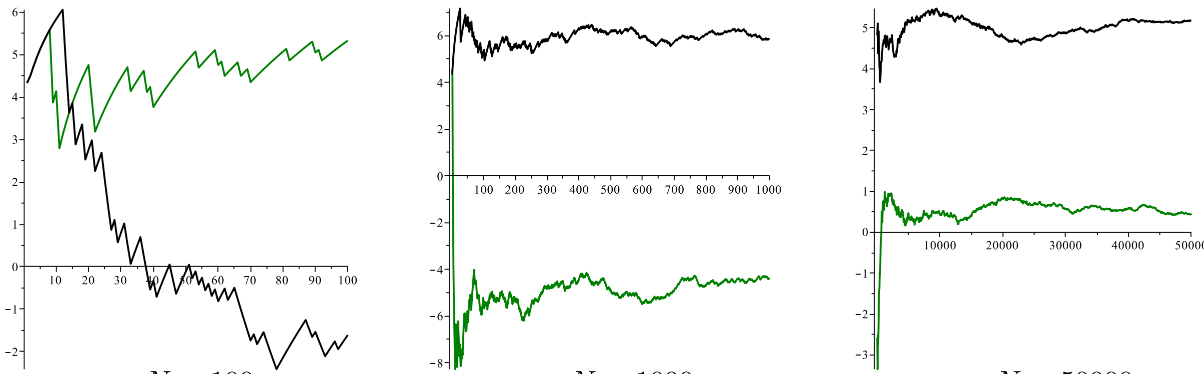
A remarkable fact: the distribution W depends on the initial composition of the urn, which does not happen for small urns. The graphs hereunder illustrate this property, representing $W - \mathbf{E}W$'s density for the large urn R_2 starting with respectively $(1, 0)$, $(1, 1)$ and $(2, 1)$ as initial composition vector.



$(\alpha, \beta) = (1, 0)$ $\qquad$ $(\alpha, \beta) = (1, 1)$ $\qquad$ $(\alpha, \beta) = (2, 1)$

Normalised distribution of the number of red balls in a large urn R_2
after 500 drawings, initial composition (α, β)

The last illustration concerns the second term order which has a random asymptotics. Two normalised trajectories of the number of red balls in an R_2-urn up to time $N = 100$, 1000 and 50000 are plotted. The convergence of $\frac{1}{n^\sigma}\left(U_n^{(1)} - nv_1^{(1)}\right)$ is here almost sure: for (almost) any sequence of random drawings in the large urn, this random variable converges to a (random) limit. The situation is very different from the small urn case, where a given trajectory do not give rise to the convergence of the second order normalised number of red balls. Here again, on the x-axis, the number n of drawings up to N; on the y-axis, the second order normalised number of red balls $\frac{1}{n^\sigma}\left(U_n^{(1)} - nv_1^{(1)}\right)$. Here again, $v_1^{(1)}$ denotes v_1 first coordinate



$N = 100$ $\quad$ $N = 1000$ $\quad$ $N = 50000$

$\frac{1}{n^\sigma}\left(U_n^{(1)} - nv_1^{(1)}\right)$ in three sequences of N drawings in a large urn R_2, initial composition $(1,0)$

3.3 Hint of proof

All the proofs of these asymptotic results rely on martingale theory.
Historically, the first approach was made in the 70's by Athreya and Karlin who considered the composition vector process of an urn as a multitype branching process. They first embed the urn process into continuous time and make its study as a continuous-time branching process [1]. In his seminal article [5], Janson adapts the method in a complete study of an urn process under an irreducibility assumption. A direct discrete time approach based on moments is made in [6]. The arguments presented hereunder rely essentially on this latter approach.

The vector-valued Markov process $(U_n)_{n\in\mathbb{N}}$ is defined by the probability transitions (1) and the initial composition vector $U_0 = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$. In particular, if $f : \mathbb{R}^2 \to V$ is any function that takes its value in any real vector space V, the conditional expectation of U_{n+1} writes

$$\mathbf{E}\left(f\left(U_{n+1}\right)\Big|U_n\right) = \frac{U_n^{(1)}}{nS+\alpha+\beta}\, f\left(U_n + \begin{pmatrix}a\\ b\end{pmatrix}\right) + \frac{U_n^{(2)}}{nS+\alpha+\beta}\, f\left(U_n + \begin{pmatrix}c\\ d\end{pmatrix}\right).$$

Thanks to the deterministic relation $U_n^{(1)}+U_n^{(2)} = nS+\alpha+\beta$, this formula can be written the following way:

$$\mathbf{E}\left(f\left(U_{n+1}\right)\Big|U_n\right) = \left(\mathrm{Id} + \frac{\Phi}{nS+\alpha+\beta}\right)(f)\left(U_n\right) \tag{6}$$

where Φ denotes the operator defined, for any function f as above and any vector $v = \begin{pmatrix}v^{(1)}\\ v^{(2)}\end{pmatrix} \in \mathbb{R}^2$, by

$$\Phi\left(f\right)\left(v\right) = v^{(1)}\left[f\left(v + \begin{pmatrix}a\\ b\end{pmatrix}\right) - f\left(v\right)\right] + v^{(2)}\left[f\left(v + \begin{pmatrix}c\\ d\end{pmatrix}\right) - f\left(v\right)\right]. \tag{7}$$

A first consequence is the expectation of $f\left(U_n\right)$, obtained by recursion from Formula (6): if $f : \mathbb{R}^2 \to V$ is any function,

$$\mathbf{E}f\left(U_n\right) = \gamma_{n,\alpha+\beta}\left(\Phi\right)\left(f\right)\left(U_0\right) \tag{8}$$

where $\gamma_{n,\tau}$ is the real polynomial defined by

$$\gamma_{n,\tau}(X) = \prod_{k=0}^{n-1}\left(1+\frac{X}{kS+\tau}\right)$$

(τ is a non zero real number; if $n=0$, this empty product equals 1). Notice that, thanks to Stirling Formula, when z is any complex number, one gets the asymptotics

$$\gamma_{n,\tau}(z) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau+z}{S}\right)} n^{\frac{z}{S}}\left(1+O\left(\frac{1}{n}\right)\right) \tag{9}$$

where Γ denotes Euler Gamma function. Formulae (6) and (8) are basic tools for the present proof. When $f \neq 0$ is an eigenvector of Φ related to the eigenvalue λ, *i.e.* when $\Phi(f) = \lambda f$, then $\gamma_{n,\tau}(\Phi)(f)(v) = \gamma_{n,\tau}(\lambda) \times f(v)$ so that Formula (9) gives immediately the asymptotics of $\mathbf{E}f(U_n)$ when n tends to infinity. With this elementary remark, one can evaluate the asymptotic joint moments of U_n's coordinates, leading to the proof of Theorem 4. Theorem 2 can also be proven with such tools. Classically, the proof of the small irreducible case (Theorem 3) is made by embedding the process into continuous time, and coming back to discrete time using some suitable random stopping-time. See [5] for a complete proof.

Exercise 6.
6.1- (Linear functions)
Show that if V is a real vector space and if $f : \mathbb{R}^2 \to V$ is linear, then

$$\Phi(f) = f \circ A$$

where $A : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $A(v) = A\begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} := {}^tR\begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix}$.

6.2- (Vector-valued martingale)
Denote $\tau := \alpha+\beta$. Show that the process $\left(\gamma_{n,\tau}\left({}^tR\right)^{-1}(U_n)\right)_n$ is a martingale (with regard to the natural filtration) as soon as it is defined, *i.e.* as soon as all matrices $I_2 + \frac{1}{kS+\tau}R$, $k \in \mathbb{N}$ are invertible. Show that this martingale is not defined if, and only if $m \leq -1$ and S divides $m+\alpha+\beta$.

[Apply **6.1-** to $f = \mathrm{Id}$. This leads to $\mathbf{E}\left(U_{n+1}\middle|U_n\right) = \left(I_2 + \frac{1}{nS+\tau}A\right)(U_n)$. This implies all the answers, because A is diagonalizable, with eigenvalues S and m. For the martingale assertion, one can also refer to Brigitte Chauvin's course *Random trees and probability*, Proposition 3.7, where a similar argument is given.]

6.3- (Expectation)
Let u_1 and u_2 be the eigenforms defined in Section 1. Verify (or remember!) that $u_1 \circ A = Su_1$ and $u_2 \circ A = mu_2$. Show that for any $n \in \mathbb{N}$,

$$\mathbf{E}u_1(U_n) = n + \frac{\tau}{S}$$

and, when n tends to infinity,

$$\mathbf{E}u_2(U_n) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S}+\sigma\right)}\frac{b\alpha - c\beta}{S} n^{\sigma}\left(1+O\left(\frac{1}{n}\right)\right).$$

When $R \neq SI_2$, using that $v = u_1(v)v_1 + u_2(v)v_2$ for any vector $v \in \mathbb{R}^2$, show that, when n tends to infinity,

$$\mathbf{E}U_n \sim nv_1$$

[An induction using **6.1-** leads to $\mathbf{E}u_1\left(U_n\right) = \gamma_{n,\tau}(S) \times u_1\left(U_0\right) = \frac{nS+\tau}{\tau} \times \frac{\tau}{S}$. For u_2, apply Formula (9) to $\mathbf{E}u_2\left(U_n\right) = \gamma_{n,\tau}(m) \times u_2\left(U_0\right)$ with a O-remainder. The third assertion is obtained by addition of asymptotic developments.]

6.4- (Real-valued projected martingales)
Show that

$$\left(\frac{u_1\left(U_n\right)}{nS+\tau}\right)_n$$

is an almost surely bounded (thus convergent) martingale and compute its expectation. Show that

$$\left(\frac{u_2\left(U_n\right)}{\gamma_{n,\tau}\left(m\right)}\right)_n$$

is a martingale as well, as soon as $m \geq 0$ or $m+\tau$ is not a multiple of S.
[Using **6.1-** again, one gets $\mathbf{E}\left(u_1\left(U_{n+1}\right)\big|U_n\right) = \left(1+\frac{S}{nS+\tau}\right) \times u_1\left(U_n\right)$, so that $\mathbf{E}\left(\frac{u_1(U_{n+1})}{(n+1)S+\tau}\big|U_n\right) = \frac{u_1(U_n)}{nS+\tau}$, proving the martingale property. Same argument from $\mathbf{E}\left(u_2\left(U_{n+1}\right)\big|U_n\right) = \left(1+\frac{m}{nS+\tau}\right) \times u_2\left(U_n\right)$.]

6.5- (Second moments)
Denote by P and Q the 2-variable polynomials defined by

$$P(x,y) = u_1(x,y)\Big(u_1(x,y)+1\Big) \quad \text{and} \quad Q(x,y) = \Big(u_1(x,y)+\sigma\Big)u_2(x,y).$$

Show that $\Phi(P) = 2SP$ and $\Phi(Q) = (S+m)Q$ and prove the asymptotics when n tends to infinity

$$\mathbf{E}P\left(U_n\right) = n^2\left(1+O\left(\frac{1}{n}\right)\right)$$

and

$$\mathbf{E}Q\left(U_n\right) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S}+\sigma\right)}\frac{b\alpha - c\beta}{S}n^{1+\sigma}\left(1+O\left(\frac{1}{n}\right)\right)$$

(if one feels depressed, one can just show that $Q\left(U_n\right) \in O\left(n^{1+\sigma}\right)$, :-)).

Suppose that $\sigma \neq 1/2$ and denote

$$R = u_2^2 - \frac{bc\sigma^2}{1-2\sigma}u_1 + (b-c)\,\sigma u_2.$$

Using (7), show that, in this case, $\Phi\left(R\right) = 2mR$ and that, when n tends to infinity,

$$\mathbf{E}R\left(U_n\right) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S}+2\sigma\right)}R\left(\alpha,\beta\right)n^{2\sigma}\left(1+O\left(\frac{1}{n}\right)\right).$$

Show that $(1, u_1, u_2, P, Q, R)$ is a basis of the vector space $\mathbb{R}_2[x, y]$ of polynomials of degree less than or equal to 2. Write x^2, xy and y^2 in this basis and compute the asymptotics of the co-moment matrix $\mathbf{E}\left[U_n{}^t U_n\right]$ and of the covariance matrix $\mathbf{E}\left[(U_n - \mathbf{E}U_n)^t (U_n - \mathbf{E}U_n)\right]$ (one has to discuss whether $\sigma < 1/2$ or $\sigma > 1/2$).

Check what happens when $\sigma = 1/2$ and do the same job using $T = u_2^2 + \frac{2b-m}{2} u_2$ instead of R.

[One gets $\Phi(P)$, $\Phi(Q)$ and $\Phi(R)$ by simple computation. Since $\Phi(P) = 2SP$, $\mathbf{E}P(U_n) = \gamma_{n,\tau}(2S) \times P(U_0)$ and the required asymptotics for $\mathbf{E}P(U_n)$ is obtained thanks to Formula (9). *Idem* for $\mathbf{E}Q(U_n)$ and $\mathbf{E}R(U_n)$. The remainder of the exercise is completely left to the reader.]

6.6- (For large urns, the second projected martingale is square-bounded)
Suppose that $\sigma > 1/2$. Expressing u_2^2 as a function of R, u_1 and u_2, show that the martingale $\left(\frac{u_2(U_n)}{\gamma_{n,\tau}(m)}\right)_n$ is bounded in L^2, thus convergent.

[$u_2^2 = R + \frac{bc\sigma^2}{1-2\sigma} u_1 - (b-c)\,\sigma u_2$, so that $\mathbf{E}u_2^2(U_n) = c_1 n^{2\sigma}\left(1 + O(1/n)\right) + c_2 n + c_3 n^{\sigma}\left(1 + O(1/n)\right)$ where c_1, c_2 and c_3 are constants. Since $\sigma > 1/2$, the principal term is the one in $n^{2\sigma}$, proving that the martingale is square bounded (use Formula (9) again to get the asymptotics of $\gamma_{n,\tau}(m)^2$).]

Exercise 7 (triangular urn).
Assume that $b = 0$, so that $R = \begin{pmatrix} S & 0 \\ S-m & m \end{pmatrix}$. Assume also that the initial number of black balls is non zero, *i.e.* that $\beta \neq 0$ (and check that $\beta = 0$ leads to a degenerate process). Let as above u_1 be the linear form $u_1(x, y) = \dfrac{x+y}{S}$ but let here u_2 be the linear form

$$u_2(x, y) = \frac{y}{S}.$$

For any $p \in \mathbb{N}^*$, let also A_p and B_p be the bivariate polynomials

$$A_p = u_1\,(u_1 + 1) \ldots (u_1 + p - 1) = \frac{\Gamma\,(u_1 + p)}{\Gamma\,(u_1)}$$

and

$$B_p = u_2\Big(u_2 + \sigma\Big) \ldots \Big(u_2 + (p-1)\sigma\Big) = \frac{\Gamma\,(u_2 + p\sigma)}{\Gamma\,(u_2)}.$$

Show that $\Phi\,(A_p) = pSA_p$ (as always, even if R is not triangular) and that $\Phi\,(B_p) = pmB_p$ for any $p \geq 1$. Deduce from this that, when n tends to infinity,

$$\mathbf{E}B_p\,(U_n) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S} + p\sigma\right)} \frac{\Gamma\left(\frac{\beta}{S} + p\sigma\right)}{\Gamma\left(\frac{\beta}{S}\right)} n^{p\sigma} \left(1 + O\left(\frac{1}{n}\right)\right).$$

- Assume that $m \geq 1$.

Using the inversion formula

$$u_2^p = \sum_{k=1}^{p} (-\sigma)^{p-k} \left\{ p \atop k \right\} B_k,$$

show that, for any $p \geq 1$,

$$\lim_{n\to\infty} \mathbf{E}\left(\frac{u_2\left(U_n\right)}{n^\sigma}\right)^p = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S}+p\sigma\right)} \frac{\Gamma\left(\frac{\beta}{S}+p\sigma\right)}{\Gamma\left(\frac{\beta}{S}\right)}. \tag{10}$$

so that the number of black balls $U_n^{(2)} = Su_2\left(U_n\right)$ converges in law to a random variable having the right side of Equality (10) as p-th moment (to make a complete proof of that fact, one has to check that a distribution having such a p-th moment is determined by its moments, which can be done by computing the asymptotics of (10) as p tends to infinity with the help of Stirling Formula). This law can be related to stable laws or to Mittag-Leffler ones.

• Assume that $m = 0$. Show that the process is deterministic (degenerate case).

• Assume that $m \leq -1$. Show that the number of black balls tends almost surely to zero (degenerate case again).

References

[1] K.B. Athreya and S. Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist*, 39:1801–1817, 1968.

[2] B. Chauvin, N. Pouyanne, and R. Sahnoun. Limit distributions for large Pólya urns. *Annals Applied Prob.*, 21(1):1–32, 2011.

[3] Philippe Flajolet, Philippe Dumas, and Vincent Puyhaubert. Some exactly solvable models of urn process theory. In Philippe Chassaing, editor, *Fourth Colloquium on Mathematics and Computer Science*, volume AG of *DMTCS Proceedings*, pages 59–118, 2006.

[4] Philippe Flajolet, Joaquim Gabarró, and Helmut Pekari. Analytic urns. *Annals of Probability*, 33:1200–1233, 2005.

[5] S. Janson. Functional limit theorem for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications*, 110:177–245, 2004.

[6] N. Pouyanne. An algebraic approach to Pólya processes. *Annales de l'Institut Henri Poincaré*, 44:293–323, 2008.

Random trees and Probability[1].

Brigitte CHAUVIN
http://chauvin.perso.math.cnrs.fr/

CIMPA Summer School, 2014
Nablus, University An-Najah
August 2014

Contents

1 Abstract/Introduction **90**

2 Binary search trees **90**
2.1 Definition of a binary search tree 90
2.2 Profile of a binary search tree 91
2.2.1 Level polynomial. BST martingale 91
2.2.2 Embedding in continuous time. Yule tree 94
2.2.3 Connection Yule tree - binary search tree 95
2.2.4 Asymptotics of the profile 96
2.3 Path length of a binary search tree 96

3 m-ary search trees **97**
3.1 Definition 97
3.2 Vectorial discrete martingale 99
3.3 Embedding in continuous time. Multitype branching process . . . 101
3.4 Asymptotics 103
3.4.1 Notations 103
3.4.2 Dislocation equations 104

4 Smoothing transformation **106**
4.1 Contraction method 106
4.2 Analysis on Fourier transforms 108
4.3 Cascade type martingales 109

[1]Keywords: branching process, branching property, martingale, analysis of algorithms, Pólya urn, binary search tree, smoothing transformation, fixed point equation, support.

1 Abstract/Introduction

In this school, three examples are developed, involving random trees: binary search trees, Pólya urns and m-ary search trees. For all of them, a same plan runs along the following outline:
(a) A discrete Markovian stochastic process is related to a tree structure. In the three cases, the tree structure is a model coming from computer science and from analysis of algorithms, typically sorting algorithms. The recursive nature of the problem gives rise to *discrete time martingales*.
(b) The process is embedded in continuous time, giving rise to a one type or to a multitype *branching process*. The associated continuous time martingales are connected to the previous discrete time martingales. Thanks to the branching property, the asymptotics of this continuous time branching process is more accessible than in discrete time, where the branching property does not hold.

In all the cases, the limit of the (rescaled) martingale has a non classic distribution. We present some expected properties of these limit distribution (density, support, ...) together with more exciting properties (divergent moment series, fixed point equation, moments, ...).

Sections 2 on binary search trees and Section 3 on m-ary search trees are developped in this course, Pólya urns are developped in Pouyanne's course.

2 Binary search trees

(in short: BST)

2.1 Definition of a binary search tree

A binary search tree is associated with the sorting algorithm "Quicksort" and several definitions can be given with this algorithm in mind (see Mahmoud [15]). Hereunder we give a more probabilistic definition. Let

$$\mathcal{U} = \{\varepsilon\} \cup \bigcup_{n \geq 1} \{0,1\}^n$$

be the set of finite words on the alphabet $\{0,1\}$, where ε denotes the empty word. Words are written by concatenation, the left children of u is $u0$ and the right children of u is $u1$. A *binary complete tree* T is a finite subset of $\mathcal{U}$ such that

$$\begin{cases} \varepsilon \in T \\ \text{if } uv \in T \text{ then } u \in T\,, \\ u1 \in T \Leftrightarrow u0 \in T\,. \end{cases}$$

The root of the tree is ε. The length of a node u is denoted by $|u|$, it is the depth of u in the tree ($|\varepsilon| = 0$). The set of binary complete trees is denoted by $\mathcal{B}$. In a binary complete tree $T \in \mathcal{B}$, a leaf is a node without any children, the set of leaves of T is denoted by ∂T. The other nodes are internal nodes.

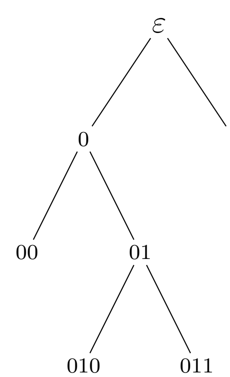


Figure 1: *An example of complete binary tree. At each node is written the word labelling it.*

In the following, we call a random binary search tree the discrete time process $(\mathcal{T}_n)_{n\geq 0}$, with values in $\mathcal{B}$, recursively defined by: $\mathcal{T}_0$ is reduced to a single leaf; for $n \geq 0$, $\mathcal{T}_{n+1}$ is obtained from $\mathcal{T}_n$ by a uniform insertion on one of the $(n+1)$ leaves of $\mathcal{T}_n$. See Figure 2.

2.2 Profile of a binary search tree

2.2.1 Level polynomial. BST martingale

A huge literature exists on binary search trees: see Flajolet and Sedgewick [11] for analytic methods, Devroye [9] for more probabilistic ones and Mahmoud [15] for a book on this topics. In this section, let us focus on the *profile* which expresses the shape of the tree. The profile is given par the sequence

$$U_k(n) := \text{the number of leaves at level } k \text{ in tree } \mathcal{T}_n.$$

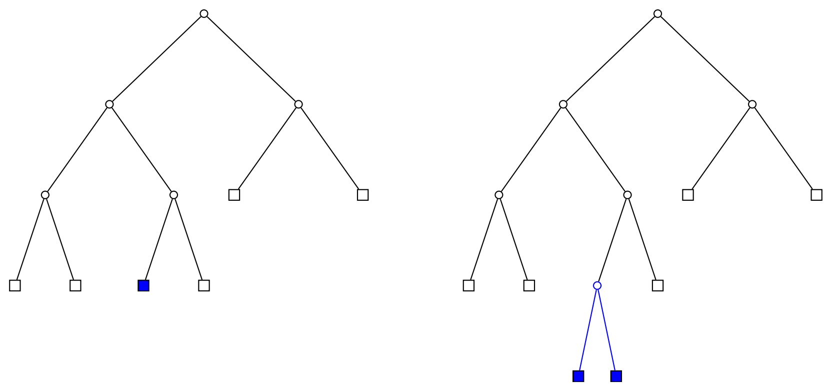


Figure 2: *An example of transition from $\mathcal{T}_5$, a binary search tree of size* 5 *to* $\mathcal{T}_6$, *a binary search tree of size* 6. *The insertion depth equals* 3.

What is the asymptotic behavior of these quantities when $n \to +\infty$? To answer, let's introduce the *level polynomial*, defined for any $z \in \mathbb{C}$ by

$$W_n(z) := \sum_{k=0}^{+\infty} U_k(n)\ z^k = \sum_{u \in \partial \mathcal{T}_n} z^{|u|}. \tag{1}$$

It is indeed a polynomial, since for any level k greater than the height of the tree, $U_k(n) = 0$. It is a random variable, not far from a martingale.

Theorem 2.1 *For any complex number $z \in \mathbb{C}$ such that $z \neq -k, k \in \mathbb{N}$, let*

$$\Gamma_n(z) := \prod_{j=0}^{n-1} \big(1 + \frac{z}{j+1}\big) \quad \textit{and} \quad M_n^{BST}(z) := \frac{W_n(z)}{\mathbb{E}(W_n(z))} = \frac{W_n(z)}{\Gamma_n(2z-1)}.$$

Then, $(M_n^{BST}(z))_n$ is a $\mathcal{F}_n$-martingale with expectation 1, *which can also be written*

$$M_n^{BST}(z) := \frac{1}{\Gamma_n(2z-1)} \sum_{u \in \partial \mathcal{T}_n} z^{|u|}. \tag{2}$$

This martingale is a.s. convergent for any z positive real.
It converges in L^1 to a limit denoted by $M_\infty^{BST}(z)$ for any $z \in]z_-, z_+[$ and it converges a.s. to 0 *for any $z \notin]z_-, z_+[$, where z_- and z_+ are the two solutions of the equation $z \log z - z + 1/2 = 0$. Numerically, $z_- = 0.186\ldots; z_+ = 2.155\ldots$*

Proof. Let d_n be the insertion depth of a new node in the tree $\mathcal{T}_n$ of size n. Remember this insertion is uniform on the $n+1$ leaves of $\mathcal{T}_n$. In other words

$$\mathbb{P}(d_n = k \big| \mathcal{F}_n) = \frac{U_k(n)}{n+1}.$$

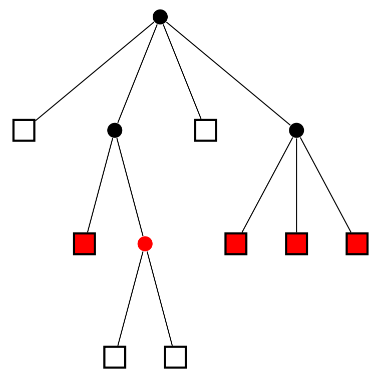


Figure 3: *A non binary tree τ, with height $h(\tau) = 3$, with profile $(0, 2, 4, 2)$. The second generation is in red.*

The number of leaves at level k in the tree $\mathcal{T}_{n+1}$ can be expressed via d_n, see Figure 2:

$$U_k(\mathcal{T}_{n+1}) = U_k(\mathcal{T}_n) - \mathbf{1}_{\{d_n=k\}} + 2\ \mathbf{1}_{\{d_n=k-1\}}.$$

Consequently,

$$\begin{aligned}
\mathbb{E}(W_{n+1}(z) \mid \mathcal{F}_n) &= \mathbb{E}\left(\sum_{k=0}^{+\infty} U_k(\mathcal{T}_{n+1}) z^k \mid \mathcal{F}_n\right) \\
&= \sum_{k=0}^{+\infty} z^k \mathbb{E}\left(U_k(\mathcal{T}_n) - \mathbf{1}_{\{d_n=k\}} + 2\ \mathbf{1}_{\{d_n=k-1\}} \mid \mathcal{F}_n\right) \\
&= \sum_{k=0}^{+\infty} z^k \mathbb{E}\left(U_k(\mathcal{T}_n) - \mathbb{P}(d_n = k \mid \mathcal{F}_n) + 2\mathbb{P}(d_n = k-1 \mid \mathcal{F}_n)\right) \\
&= W_n(z) - \sum_{k=0}^{+\infty} \frac{U_k(\mathcal{T}_n)}{n+1} z^k + 2\sum_{k=1}^{+\infty} \frac{U_{k-1}(\mathcal{T}_n)}{n+1} z^k \\
&= W_n(z) - \frac{1}{n+1} W_n(z) + 2z W_n(z), \\
&= \frac{n+2z}{n+1}\ W_n(z),
\end{aligned} \tag{3}$$

which gives the martingale property, after scaling: indeed, take the expectation in (3) to obtain par recurrence on n

$$\mathbb{E}(W_n(z)) = \prod_{j=0}^{n-1} \frac{j+2z}{j+1} = \Gamma_n(2z-1).$$

and divide by this expectation in (3) to get

$$\mathbb{E}(M_{n+1}^{BST}(z) \mid \mathcal{F}_n) = \mathbb{E}\left(\frac{W_{n+1}(z)}{\Gamma_{n+1}(2z-1)} \Big| \mathcal{F}_n\right) = \left(1 + \frac{2z-1}{n+1}\right) \frac{W_n(z)}{\Gamma_{n+1}(2z-1)} = M_n^{BST}(z).$$

$\square$

2.2.2 Embedding in continuous time. Yule tree

The idea is due to Pittel [16]. Let's consider a continuous time branching process, with an ancestor at time $t = 0$, who lives an exponential time with parameter 1. When he dies, it gives birth to two children who live an exponential time with parameter 1, independently from each other, etc... The tree process thus obtained is called the Yule tree process, it is denoted by $(\mathcal{Y}_t)_t$.

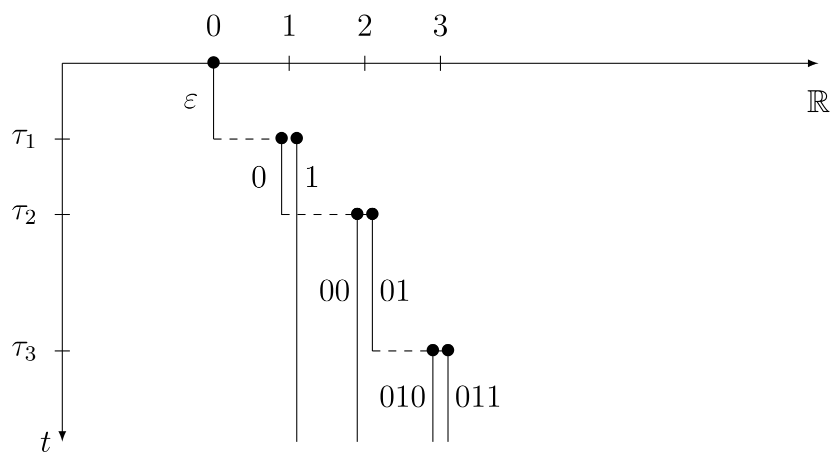


Figure 4: *A representation of a Yule tree. Here $N_t = 4$. The displacements are the generation numbers.*

Let's call N_t the number of leaves in $\mathcal{Y}_t$ (at time t) and denote by

$$0 < \tau_1 < \cdots < \tau_n < \ldots$$

the successive jumping times. For any time t there exists a unique integer n such that $\tau_{n-1} \leq t < \tau_n$ and

$$\{N_t = n\} = \{\tau_{n-1} \leq t < \tau_n\}.$$

Due to the lack of memory of the exponential distribution, $\tau_n - \tau_{n-1}$ is the first time when one of the n living particles splits. Consequently, $\tau_n - \tau_{n-1}$ is the minimum of n independent random variables $\mathcal{E}xp(1)$-distributed[2], so it is $\mathcal{E}xp(n)$-distributed. Moreover, the splitting particle is uniformly chosen among the n living particles. Finally, the continuous-time process stopped at time τ_n and the binary search tree have the same growing dynamics, so that (it is the embedding principle)

$$\left(\mathcal{Y}_{\tau_n}\right)_n \stackrel{\mathcal{L}}{=} \left(\mathcal{T}_n\right)_n . \tag{4}$$

From now, we consider that both processes (the binary search tree and the Yule tree) are built on the same probability space, so that equality in distribution becomes almost sure equality.

2.2.3 Connection Yule tree - binary search tree

On the Yule tree, let us define the "position" of an individual u living at time y by

$$X_u(t) := -|u| \log 2$$

so that the displacements are (up to the constant $\log 2$) like the generation numbers in the tree. See Figure 4. It can be proved (coming from the theory of branching random walks, see Biggins [4] and Bertoin and Rouault [3]) that

Theorem 2.2 *For any $z \in \mathbb{C}$,*

$$M_t^{YULE}(z) := \sum_{u \in \partial \mathcal{Y}_t} z^{|u|} e^{-t(2z-1)}$$

is a $\mathcal{F}_t$-martingale, with expectation 1. This martingale converges a.s. for all z positive real. It converges in L^1 to a limit denoted by $M_\infty^{YULE}(z)$ for all $z \in]z_-, z_+[$ and it converges a.s. to 0 for all $z \notin]z_-, z_+[$, where z_- and z_+ are the solutions of equation $z \log z - z + 1/2 = 0$. Numerically, $z_- = 0.186\ldots ; z_+ = 2.155\ldots$

Moreover, this martingale is connected to the BST martingale $M_n^{BST}(z)$. Indeed, writing $M_n^{BST}(z)$ like in (2), and taking the Yule martingale at time $t = \tau_n$ gives, thanks to the embedding principle (4)

$$\begin{aligned}
M_{\tau_n}^{YULE}(z) &= \sum_{u \in \partial \mathcal{Y}_{\tau_n}} z^{|u|} e^{-\tau_n(2z-1)} \\
&= e^{-\tau_n(2z-1)} \sum_{u \in \mathcal{T}_n} z^{|u|} \\
&= e^{-\tau_n(2z-1)} \Gamma_n(2z-1) M_n^{BST}(z).
\end{aligned}$$

[2]For any positive real λ, $\mathcal{E}xp(\lambda)$ denotes an exponential probability law with parameter λ.

It is not difficult to pass to the limit in the preceding equality, when n tends to infinity, when the parameter z belongs to the L^1-convergence domain of the martingales. In a Yule process, it is known (see for instance Athreya and Ney [1]) that $e^{-t}N_t$ tends to a random limit ξ which is $\mathcal{E}xp(1)$-distributed, when t tends to infinity. Since the stopping times τ_n go to infinity when n goes to infinity, we deduce that $ne^{-\tau_n}$ converges to ξ when n goes to infinity. Finally let us use the Stirling formula to get the estimate

$$\Gamma_n(2z-1) \sim \frac{n^{2z-1}}{\Gamma(2z)},$$

so that we have proved the following proposition.

Proposition 2.3 *For any $z \in]z_-, z_+[$, the following connection holds*

$$M_\infty^{YULE}(z) = \frac{\xi^{2z-1}}{\Gamma(2z)} M_\infty^{BST}(z)$$

where ξ and $M_\infty^{BST}(z)$ are independent and ξ is $\mathcal{E}xp(1)$-distributed.

2.2.4 Asymptotics of the profile

The above connection is one of the main tools leading to the following theorem on the profile of binary search trees. This theorem expresses that, after scaling, the profile tends to the random limit M_∞^{BST}. The asymptotics of the profile is concentrated on the levels k proportional to $\log n$.

Theorem 2.4 *For any compact $K \subset]z_-, z_+[$,*

$$\frac{U_k(n)}{\mathbb{E}(U_k(n))} - M_\infty^{BST}\left(\frac{k}{2\log n}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad a.s.$$

uniformly on $\frac{k}{2\log n} \in K$.

2.3 Path length of a binary search tree

Definition 2.5 (path length of a BST) *The (external) path length L_n of a binary search tree $\mathcal{T}_n$ is the sum of the levels of the leaves of the tree.*

$$L_n := \sum_{u \in \partial \mathcal{T}_n} |u|.$$

This parameter of the tree is interesting in analysis of algorithms, since it represents a cost: $\frac{L_n}{n+1}$ is the mean cost of an insertion in the tree of size n.
Obviously, the path length is related to the level polynomial $W_n(z)$, since

$$L_n = \sum_{k\geq 1} kU_k(n) = W_n'(1).$$

Consequently, elementary computations (taking into account $\mathbb{E}M_n^{BST}(z) = 1$ and $\mathbb{E}M_n'(z) = 0$) lead to

$$\mathbb{E}(L_n) = 2(n+1)(H_{n+1} - 1) \qquad ; \qquad M_n'(1) = \frac{1}{n+1}\left(L_n - \mathbb{E}(L_n)\right),$$

where H_n is the n-th harmonic number, and M_n' is the derivative of M_n^{BST}. Now, the derivative of a martingale is still a martingale, and $z = 1$ is in the L^1-convergence domain of the BST martingale $M_n(z)$, so that it is straightforward to obtain the following theorem.

Theorem 2.6 *After scaling, the path length of a binary search tree, defined by*

$$Y_n := \frac{1}{n+1}\left(L_n - \mathbb{E}(L_n)\right)$$

is a $\mathcal{F}_n$-martingale with mean 0. *It converges almost surely and in L^1 to a random limit denoted by Y.*

The law of Y is sometimes called the "law of Quicksort". It can be viewed as a solution of a distributional equation, in the spirit of Section 4.

3 m-ary search trees

3.1 Definition

For $m \geq 3$, m-ary search trees are a generalization of binary search trees (see for instance Mahmoud [15]). A sequence $(T_n, n \geq 0)$ of m-ary search trees grow by successive insertions of keys in their leaves. Each node of these trees contains at most $(m-1)$ keys. Keys are i.i.d. random variables $x_i, i \geq 1$ with any diffusive distribution on the interval $[0, 1]$. The tree $T_n, n \geq 0$, is recursively defined as follows:
T_0 is reduced to an empty node-root; T_1 is reduced to a node-root which contains x_1, T_2 is reduced to a node-root which contains x_1 and x_2, ... , T_{m-1} has a

node-root containing $x_1, \dots x_{m-1}$. As soon as the $m-1$-st key is inserted in the root, m empty subtrees of the root are created, corresponding from left to right to the m ordered intervals $I_1 =]0, x_{(1)}[, \dots, I_m =]x_{(m-1)}, 1[$ where $0 < x_{(1)} < \dots < x_{(m-1)} < 1$ are the ordered $(m-1)$ first keys. Each following key $x_m, \dots$ is recursively inserted in the subtree corresponding to the unique interval I_j to which it belongs. As soon as a node is saturated, m empty subtrees of this node are created. The process $(T_n)_{n\geq 0}$ is recursively built, where T_n is the m-ary tree of size n, i.e. containing n keys. See Figure 5.

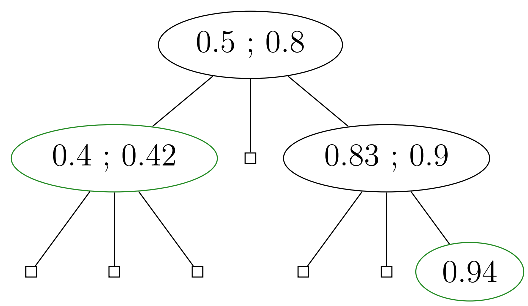


Figure 5: *A m-ary search tree ($m = 3$) of size 7, with 8 gaps, 4 nodes; among them, fringe nodes are in green. The tree has been built with the successive keys:* $0.8; 0.5; 0.9; 0.4; 0.42; 0.83; 0.94$.

To describe such a tree, let us introduce the so-called composition vector of the tree, X_n, which counts the nodes of differents types in the tree. This composition vector of the m-ary search tree provides a model for the space requirement of the sorting algorithm. More precisely, for each $i = \{1, \dots, m\}$ and $n \geq 1$, let

$$X_n^{(i)} := \text{ number of nodes in } T_n \text{ which contain } (i-1) \text{ keys (and } i \text{ gaps).}$$

Such nodes are named nodes of type i. Counting the number of keys in T_n with the $X_n^{(i)}$, we get the relation:

$$n = \sum_{i=1}^{m} (i-1) X_n^{(i)},$$

which allows to only study $m-1$ variables $X_n^{(i)}$ instead of m. We choose to forget the saturated nodes, which are internal nodes and to only count the non saturated nodes, which are at the fringe of the tree.

When the data are i.i.d. random variables, one gets a *random* m-ary search tree. With this dynamics, the insertion of a new key is *uniform* on the gaps. We want to describe the asymptotic behavior of the vector X_n as n tends to infinity.
Remark here the urn model, when considering the *gaps*. Call the gap process $(G_n)_n$. Write the replacement matrix. Notice that $G_n^{(i)} = iX_n^{(i)}$.

3.2 Vectorial discrete martingale

The dynamics of the nodes is illustrated by Figure 6 and it gives the expression

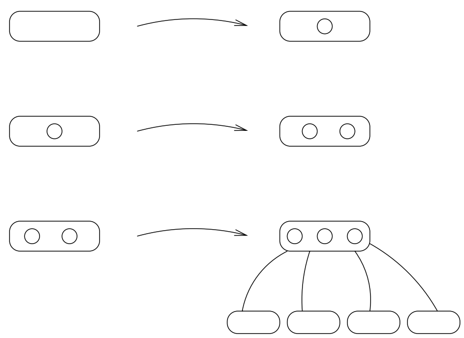


Figure 6: Dynamics of insertion of data, in the case $m = 4$.

of X_{n+1} as a function of X_n. The $(n+1)$-st data is inserted in a node of type i, $i = 1, \dots, m-1$ with probability $\dfrac{iX_n^{(i)}}{n+1}$ and in this case, the node becomes a node of type $i+1$ for $i = 1, 2, \dots, m-2$, and gives m nodes of type 1, if $i = m-1$.
In other words, for $i = 1, \dots, m-1$, let

$$\begin{cases} \Delta_1 = (-1, 1, 0, 0, \dots) \\ \Delta_2 = (0, -1, 1, 0, \dots) \\ \quad \vdots \\ \Delta_{m-2} = (0, \dots, 0, -1, 1) \\ \Delta_{m-1} = (m, 0, \dots, 0, -1). \end{cases},$$

Then

$$\mathbb{P}(X_{n+1} = X_n + \Delta_i | X_n) = \frac{iX_n^{(i)}}{n+1}.$$

The remarkable fact is that the transition from X_n to X_{n+1} is *linear* in X_n. Notice also that $\sum_{i=1}^{m-1} \frac{iX_n^{(i)}}{n+1} = 1$. When we note

$$A = \begin{pmatrix} -1 & & & & & m(m-1) \\ 1 & -2 & & & & \\ & 2 & -3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & -(m-2) & \\ & & & & m-2 & -(m-1) \end{pmatrix}$$

then

$$\mathbb{E}\Big(X_{n+1}|X_n\Big) = \sum_{i=1}^{m-1} (X_n + \Delta_i) \frac{iX_n^{(i)}}{n+1} = \Big(I + \frac{A}{n+1}\Big) X_n.$$

We call Γ_n the polynomial

$$\Gamma_n(z) := \prod_{j=0}^{n-1} \big(1 + \frac{z}{j+1}\big),$$

and we deduce first, by taking the expectation, and then by induction that: $\mathbb{E}(X_n) = \Gamma_n(A)X_0$. Dividing by $\Gamma_n(A)$, we get:

Proposition 3.7 *Let $(X_n)_n$ be the composition vector of a m-ary search tree. Then, $(\Gamma_n(A)^{-1}X_n)_n$ is a $\mathcal{F}_n$ vectorial martingale.*

The spectrum of matrix A gives the asymptotic behavior of X_n. The eigenvalues are the roots of the characteristic polynomial

$$\chi_A(\lambda) = \prod_{k=1}^{m-1}(\lambda + k) - m! = \frac{\Gamma(\lambda+m)}{\Gamma(\lambda+1)} - m! \tag{5}$$

where Γ denotes Euler's Gamma function. In other words, each eigenvalue λ is a solution of the so-called characteristic equation

$$\boxed{\prod_{k=1}^{m-1}(\lambda + k) = m!} \tag{6}$$

All eigenvalues are simple, 1 being the one having the largest real part. Let λ_2 be the eigenvalue with a positive imaginary part τ_2 and with the greatest real part σ_2 among all the eigenvalues different from 1. The asymptotic behaviour of X_n is different depending on $\sigma_2 \leq \frac{1}{2}$ or $\sigma_2 > \frac{1}{2}$. The proofs of the following theorem can be found in [15, 12, 7, 17].

Theorem 3.8

• *When* $\sigma_2 < \frac{1}{2}$, $m \leq 26$ *then*

$$\frac{X_n - nv_1}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma^2)$$

where v_1 *is an eigenvector for the eigenvalue* 1*, and where* Σ^2 *can be calculated.*

• *When* $1 > \sigma_2 > \frac{1}{2}$, $m \geq 27$ *then*

$$X_n = nv_1 + \Re\big(n^{\lambda_2} W^{DT} v_2\big) + o(n^{\sigma_2})$$

where v_1, v_2 *are deterministic, nonreal eigenvectors;* W^{DT} *is a* $\mathbb{C}$*-valued random variable with a martingale limit; the notation* DT *stands for discrete time;* $o(\ )$ *means a convergence a.s. and in all the* L^p, $p \geq 1$*; the moments of* W^{DT} *can be recursively computed.*

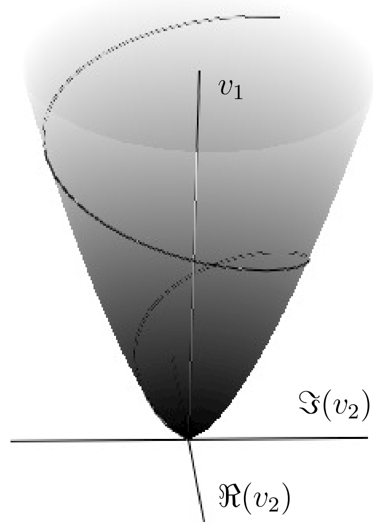
Geometrically speaking: let us denote by φ any argument of the complex number W^{DT}. The trajectory of the random vector X_n, projected in the 3-dimensional real vector space spanned by the vectors $(\Re(v_2), \Im(v_2), v_1)$ is almost surely asymptotic to the (random) spiral

$$\begin{cases} x_n = |W| n^{\sigma_2} \cos(\tau_2 \log n + \varphi), \\ y_n = -|W| n^{\sigma_2} \sin(\tau_2 \log n + \varphi), \\ z_n = n, \end{cases}$$

drawn on the (random) revolution surface

$$|W|^2 z^{2\sigma_2} = x^2 + y^2,$$

when n tends to infinity.



3.3 Embedding in continuous time. Multitype branching process

For $m \geq 3$, define a continuous time multitype branching process, with $m-1$ types

$$X^{CT}(t) = \begin{pmatrix} X^{CT}(t)^{(1)} \\ \vdots \\ X^{CT}(t)^{(m-1)} \end{pmatrix}$$

with $X^{CT}(t)^{(j)} = \#$ particles of type j alive at time t.
Each particle of type j is equipped with a clock $\mathcal{E}xp(j)$-distributed. When this clock rings, the particle of type j dies and gives birth to
$\rightarrow$ a particle of type $j+1$ when $j \leq m-2$
$\rightarrow$ m particles of type 1 when $j = m-1$.

Call $0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots$ the successive jumping times. The arguments on the exponential distribution are the same ones as for binary search trees embedding. Considering the process of *gaps* instead of nodes, it is easy to see that $\tau_n - \tau_{n-1}$ is $\mathcal{E}xp(u+n-1)$-distributed, where $u = \sum_{k=1}^{m-1} kX^{CT}(0)^{(k)}$ is the numbers of gaps at time 0.

The embedding principle can be expressed

$$\big(X^{CT}(\tau_n)\big)_n \stackrel{\mathcal{L}}{=} (X_n)_n \,,$$

and as for BST, we consider that both processes are built on the same probability space, so that this equality holds almost surely. For this multitype branching process, it is classical to see that

Proposition 3.9

$$\big(e^{-tA}X^{CT}(t)\big)_{t\geq 0}$$

is a $\mathcal{F}_t$ vectorial martingale.

By projection on the eigenlines (v_1, v_2 are eigenvectors and u_1, u_2 are eigen linear forms), we get

Theorem 3.10 *([5], Janson [12])*

$$X^{CT}(t) = e^t \xi v_1\big(1+o(1)\big) + \Re\left(e^{\lambda_2 t}W^{CT}v_2\right)\big(1+o(1)\big) + o(e^{\sigma_2 t})$$

where ξ is a real-valued random variable Gamma(u)-distributed;

$$W^{CT} := \lim_{t\to\infty} e^{-\lambda_2 t}u_2(X^{CT}(t))$$

is a complex valued random variable, which admits moments of any order $p \geq 1$; $o(\ )$ means a convergence a.s. and in all the L^p, $p \geq 1$. Moreover, the following martingale connection holds

$$W^{CT} = \xi^{\lambda_2}\ W^{DT} \qquad a.s.$$

with ξ and W^{DT} independent.

The geometric interpretation with a random curve on a spiral can be done like in discrete time. Nonetheless, notice the random first term in the expansion of $X^{CT}(t)$.

3.4 Asymptotics

3.4.1 Notations

In the following, we denote

$$T = \tau_{(1)} + \cdots + \tau_{(m-1)}. \tag{7}$$

where the $\tau_{(j)}$ are independent of each other and each $\tau_{(j)}$ is $\mathcal{E}xp(j)$ distributed. Let us make precise some elementary properties of T. By induction on m, let us prove that T has

$$f_T(u) = (m-1)e^{-u}(1-e^{-u})^{m-2}\mathbf{1}_{\mathbb{R}_+}(u), \quad u \in \mathbb{R}, \tag{8}$$

as a density. Indeed, this is true for $m = 2$; when X and Y have f_X and f_Y as densities respectively, then the convolution formula gives that $Z = X + Y$ has f_Z as a density, where

$$f_Z(z) = \int_0^z f_X(z-y)f_Y(y)dy.$$

Consequently, taking $X = T$, with f_T given by (8), and $Y = \tau_{(m)}$ having $f_Y(y) = me^{-my}$ as a density, we get

$$\begin{aligned} f_Z(z) &= \int_0^z (m-1)e^{-(z-y)}\left(1-e^{-(z-y)}\right)^{m-2} me^{-my}dy & (9)\\ &= m(m-1)e^{-z}\int_0^z e^{-y}\left(e^{-y}-e^{-z}\right)^{m-2}dy & (10)\\ &= m(m-1)e^{-z}\left[-\frac{\left(e^{-y}-e^{-z}\right)^{m-1}}{m-1}\right]_0^z & (11)\\ &= me^{-z}(1-e^{-z})^{m-1}. & (12) \end{aligned}$$

We deduce from (8) that e^{-T} has a Beta distribution with parameters 1 and $m-1$. A straightforward change of variable ($x = e^{-u}$) under the integral shows that for any complex number λ such that $\Re(\lambda) > -1$,

$$\begin{aligned} \mathbb{E}e^{-\lambda T} &= \int_0^{+\infty} e^{-\lambda u} f_T(u)du = (m-1)B(1+\lambda, m-1) & (13)\\ &= \frac{(m-1)!}{\prod_{k=1}^{m-1}(\lambda+k)}, & (14) \end{aligned}$$

where B denotes Euler's Beta function:

$$B(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \quad \Re x > 0, \Re y > 0. \tag{15}$$

In particular,

$$m\mathbb{E}|e^{-\lambda T}| = m\mathbb{E}e^{-\Re(\lambda)T} = \frac{(1+m-1)\dots(1+1)}{(\Re(\lambda)+m-1)\dots(\Re(\lambda)+1)} \begin{cases} <1 \text{ if } \Re(\lambda)>1, \\ =1 \text{ if } \Re(\lambda)=1, \\ >1 \text{ if } \Re(\lambda)<1. \end{cases} \tag{16}$$

3.4.2 Dislocation equations

We would like a complete description of the $\mathbb{C}$-valued random variable W^{CT}. It is a limit of a branching process after projection and scaling, remember that

$$W^{CT} := \lim_{t\to\infty} e^{-\lambda_2 t} u_2(X^{CT}(t)).$$

Let us see now how the branching property applied at the first splitting time provides fixed point equations on the limit distributions.

Let us write dislocation equations for the continuous time branching process at finite time t. We write $X_j(t)$ for $X^{CT}(t)$ when the process starts from $X^{CT}(0) = e_j$, where e_j denotes the j-th vector of the canonical basis of $\mathbb{R}^{m-1}$ (whose j-th component is 1 and all the others are 0). This means that the process starts from an ancestor of type j.

Notice that the distribution of the first splitting time τ_1 depends on the ancestor's type; denote by $\tau_{(j)}, j = 1, \dots, m-1$, the first splitting time when the process starts from $X(0) = e_j$. Thus $\tau_{(j)}$ is $\mathcal{E}xp(j)$ distributed.

The branching property applied at the first splitting time gives:

$$\forall t > \tau_1, \quad \begin{cases} X_1(t) \overset{\mathcal{L}}{=} X_2(t-\tau_{(1)}), \\ X_2(t) \overset{\mathcal{L}}{=} X_3(t-\tau_{(2)}), \\ \dots \\ X_{m-2}(t) \overset{\mathcal{L}}{=} X_{m-1}(t-\tau_{(m-2)}), \\ X_{m-1}(t) \overset{\mathcal{L}}{=} [m]X_1(t-\tau_{(m-1)}), \end{cases} \tag{17}$$

where the notation $[m]X$ denotes the sum of m independent copies of the random variable X.

After projections of the variables $X_j(t)$ with the form u_2, scaling with $e^{-\lambda_2 t}$ and taking the limit when t goes to infinity, we get the variables

$$W_j := \lim_{t\to+\infty} e^{-\lambda_2 t} u_2(X_j(t)),$$

so that the system (17) on $X_j(t)$ leads to the following system of distributional equations on W_j:

$$\begin{cases} W_1 \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(1)}} W_2, \\ W_2 \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(2)}} W_3, \\ \dots \\ W_{m-2} \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(m-2)}} W_{m-1}, \\ W_{m-1} \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(m-1)}} [m] W_1. \end{cases} \tag{18}$$

Since W_1 is the distribution of W^{CT} starting from a particle of type 1 (which is indeed the case for the m-ary search tree), this shows that W_1 is a solution of the following fixed point equation:

$$Z \overset{\mathcal{L}}{=} e^{-\lambda_2 T}(Z^{(1)} + \cdots + Z^{(m)}), \tag{19}$$

where T is defined in (7) and where $Z^{(i)}$ are independent copies of Z, which are also independent of T. Several results can be deduced from this equation, namely the existence and the unicity of solutions, properties of the support. Some are described in the following section.
In terms of the Fourier transform

$$\varphi(t) := \mathbb{E}\exp\{i\langle t, Z\rangle\} = \mathbb{E}\exp\{i\Re(\bar{t}Z)\}, \quad t \in \mathbb{C},$$

where $\langle x, y\rangle = \Re(\overline{x}y) = \Re(x)\Re(y) + \Im(x)\Im(y)$, Equation (19) reads

$$\varphi(t) = \int_0^{+\infty} \varphi^m(te^{-\overline{\lambda_2}u}) f_T(u)du, \quad t \in \mathbb{C}, \tag{20}$$

where f_T is defined by (8). Notice that this functional equation can also be written in a convolution form: if $\Phi(t) := \varphi(e^{\overline{\lambda_2}t})$ for any $t \in \mathbb{C}$, then Φ satisfies the following functional equation:

$$\Phi(t) = \int_0^{+\infty} \Phi^m(t-u) f_T(u)du, \quad t \in \mathbb{C}. \tag{21}$$

4 Smoothing transformation

In this section, inspired from the case of m-ary search trees (see [5]), the following fixed point equation coming from the previous multitype branching process is studied, thanks to several methods. These methods are general ones, they are used for other distributional equations. Let us just mention analogous results for:
• binary search trees, where the quicksort distribution is studied in Rösler [18];
• Pólya urns where the limit distribution occurring for large urns is studied in [8, 6].
The following smoothing equation comes from m-ary search trees, studied in Section 3.

$$W \overset{\mathcal{L}}{=} e^{-\lambda T}(W^{(1)} + \cdots + W^{(m)}), \tag{22}$$

where $\lambda \in \mathbb{C}$, T is defined in (7), $W^{(i)}$ are $\mathbb{C}$-valued independent copies of W, which are also independent of T. We successively see:
• the contraction method, in order to prove existence and unicity of a solution, in a suitable space of probability measure, in Section 4.1;
• some analysis on the Fourier transforms in order to prove that W has a density, in Section 4.2;
• a cascade type martingale which is a key tool to obtain the existence of exponential moments for W, in Section 4.3.

4.1 Contraction method

This method has been developed in Rösler [18] and Rösler and Rüschendorf [19] for many examples in analysis of algorithms. The idea is to get existence and unicity of a solution of Equation (22) thanks to the Banach fixed point Theorem. Notice that we already have the existence, thanks to Section 3. The key point is to chose a suitable metric space of probability measures on $\mathbb{C}$ where the hereunder transformation $K : \mu \mapsto K\mu$ is a contraction.

$$K\mu := \mathcal{L}\big(e^{-\lambda T}(X^{(1)} + \cdots + X^{(m)})\big), \quad (\mathcal{L} : \text{law}) \tag{23}$$

where T is given by (7), $X^{(i)}$ are independent random variables of law μ, which are also independent of T.

First step: the metric space.

For any complex number C, let $\mathcal{M}_2(C)$ be the space of probability distributions on $\mathbb{C}$ admitting a second absolute moment and having C as expectation. The first point is to be sure that K maps $\mathcal{M}_2(C)$ into itself, this is given by the following lemma.

Lemma 4.11 *If λ is a root of the characteristic equation (6) such that $\Re(\lambda) > -\frac{1}{2}$ and if C is any complex number, then K maps $\mathcal{M}_2(C)$ into itself.*

PROOF. Since $\Re(\lambda) > -1$, the random variable $e^{-\lambda T}$ has an expectation. See (13). Furthermore, by (13) again, $m\mathbb{E}e^{-\lambda T} = 1$ as λ is a root of (5). This ensures the conservation of the expectation by K. Since $\Re(\lambda) > -\frac{1}{2}$, then $\mathbb{E}|e^{-\lambda T}|^2 < \infty$ and $K\mu$ admits a second absolute moment whenever μ does. Therefore $K\mu \in \mathcal{M}_2(C)$ whenever $\mu \in \mathcal{M}_2(C)$. $\square$

Now, define d_2 as the Wasserstein distance on $\mathcal{M}_2(C)$ (see for instance Dudley [10]): for $\mu, \nu \in \mathcal{M}_2(C)$,

$$d_2(\mu, \nu) = \left(\min_{(X,Y)} \mathbb{E}\left(|X - Y|^2\right) \right)^{\frac{1}{2}}, \tag{24}$$

where the minimum is taken over couples of random variables (X, Y) having respective marginal distributions μ and ν; the minimum is attained by the Kantorovich-Rubinstein Theorem – see for instance Dudley [10], p. 421. With this distance d_2, $\mathcal{M}_2(C)$ is a complete metric space.

Second step: K is a contraction on $(\mathcal{M}_2(C), d_2)$.

It is a small calculation, taking some care when choosing the random variables: let (X, Y) be a couple of complex-valued random variables such that $\mathcal{L}(X) = \mu$, $\mathcal{L}(Y) = \nu$ and $d_2(\mu, \nu) = \sqrt{\mathbb{E}|X - Y|^2}$. Let $(X_i, Y_i), i = 1, \ldots, m$ be m independent copies of the d_2-optimal couple (X, Y), and T be a real random variable with density f_T defined by (8), independent from any (X_i, Y_i). Then,

$$\mathcal{L}(e^{-\lambda T} \sum_{i=1}^{m} X_i) = K\mu \quad \text{and} \quad \mathcal{L}(e^{-\lambda T} \sum_{i=1}^{m} Y_i) = K\nu,$$

so that (remember that for all i, $\mathbb{E}(X_i) = \mathbb{E}(Y_i) = C$)

$$\begin{aligned}
d_2(K\mu, K\nu)^2 &\leq \mathbb{E}\left|\left(e^{-\lambda T}\sum_{i=1}^{m} X_i\right) - \left(e^{-\lambda T}\sum_{i=1}^{m} Y_i\right)\right|^2 \\
&= \mathbb{E}\left|e^{-\lambda T}\sum_{i=1}^{m}(X_i - Y_i)\right|^2 = \mathbb{E}\left|e^{-\lambda T}\right|^2 \mathbb{E}\left|\sum_{i=1}^{m}(X_i - Y_i)\right|^2 \\
&= \mathbb{E}\left|e^{-\lambda T}\right|^2\left(\sum_{i=1}^{m}\mathbb{E}\left|X_i - Y_i\right|^2 + \sum_{i\neq j}\mathbb{E}\left(X_i - Y_i\right)\left(\overline{X_j} - \overline{Y_j}\right)\right) \\
&= m\mathbb{E}\left|e^{-2\lambda T}\right| \ d_2\left(\mu,\nu\right)^2.
\end{aligned}$$

With Equation (16), we know that $m\mathbb{E}\left|e^{-2\lambda T}\right| < 1 \Longleftrightarrow \Re(\lambda) > \dfrac{1}{2}$, which happens for a large urn. Therefore K is a contraction on $\mathcal{M}_2(C)$. We have proved the following theorem.

Theorem 4.12 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic equation (6) such that $\Re(\lambda) > \frac{1}{2}$, and let $C \in \mathbb{C}$. Then K is a contraction on the complete metric space $(\mathcal{M}_2(C), d_2)$, and the fixed point equation (22) has a unique solution W in $\mathcal{M}_2(C)$.*

4.2 Analysis on Fourier transforms

The aim is to prove that W solution of Equation (22) has the whole complex plane $\mathbb{C}$ as its support and that W has a density with respect to the Lebesgue measure on $\mathbb{C}$. The method relies on Liu [13, 14] adapted in [5] for $\mathbb{C}$-valued variables. It runs along the following lines.
Let φ be the Fourier transform of any solution W of (22). It is a solution of the functional equation

$$\varphi(t) = \int_0^{+\infty} \varphi^m(te^{-\overline{\lambda}u}) f_T(u)du, \quad t \in \mathbb{C}, \tag{25}$$

where f_T is defined by (8).
We first prove that φ is dominated by $|t|^{-a}$ for some $a > 1$ so that the inverse Fourier transform provides a density for W. It will prove that φ is in $L^2(\mathbb{C})$ (for a distributional equation in $\mathbb{R}$, it is proved that φ is in $L^1(\mathbb{R})$).

To prove that $\varphi(t) = O(|t|^{-a})$ when $|t| \to \infty$, for some $a > 1$, we use a Gronwall-type technical Lemma which holds as soon as $A := e^{-\lambda T}$ has good moments and once we prove that $\lim_{|t|\to+\infty} \varphi(t) = 0$. It is the same to prove that $\lim_{r\to+\infty} \psi(r) = 0$ where

$$\psi(r) := \max_{|t|=r} |\varphi(t)|.$$

This comes from iterating the distributional equation (25) so that

$$\psi(r) \leq \mathbb{E}(\psi^m(r|A|)).$$

By Fatou lemma, we deduce that $\limsup_r \psi(r)$ equals 0 or 1. And it cannot be 1 because of technical considerations and because the only point where $\psi(r) = 1$ is $r = 0$. This key fact comes from a property of the support of W strongly related to the distributional equation with a non lattice type assumption: as soon as a point z is in the support of W, then the whole disc $D(0, |z|)$ is contained in the support of W. Finally, the result is

Theorem 4.13 *Let W be a complex-valued random variable solution of the distributional equation*

$$W \stackrel{\mathcal{L}}{=} e^{-\lambda T}(W^{(1)} + \cdots + W^{(m)}),$$

where λ is a complex number, $W^{(i)}$ are independent copies of W, which are also independent of T. Assume that $\lambda \neq 1$, $\Re(\lambda) > 0$, $\mathbb{E}W < \infty$ and $\mathbb{E}W \neq 0$. Then

(i) The support of W is the whole complex plane $\mathbb{C}$;

(ii) the distribution of W has a density with respect to the Lebesgue measure on $\mathbb{C}$.

4.3 Cascade type martingales

The distributional equation (22) suggests to use Mandelbrot's cascades in the complex setting (see Barral [2] for independent interest about complex Mandelbrot's cascades).

As in Section 3, take $\lambda \in \mathbb{C}$ be a root of the characteristic equation (6) with $\Re(\lambda) > 1/2$. Still denote $A = e^{-\lambda T}$. Then $m\mathbb{E}A = 1$ because λ is a root of the characteristic equation (6) and $m\mathbb{E}|A|^2 < 1$ because $\Re(\lambda) > 1/2$ (see (16)). Let $A_u, u \in U$ be independent copies of A, indexed by all finite sequences of integers

$$u = u_1...u_n \in U := \bigcup_{k\geq 1} \{1, 2, \ldots, m\}^k$$

and set $Y_0 = 1$, $Y_1 = mA$ and for $n \geq 2$,

$$Y_n = \sum_{u_1 \dots u_{n-1} \in \{1,\dots,m\}^{n-1}} mAA_{u_1}A_{u_1u_2}\dots A_{u_1\dots u_{n-1}}. \tag{26}$$

As $m\mathbb{E}A = 1$, $(Y_n)_n$ is a martingale, with expectation 1.
This martingale has been studied by many authors in the real random variable case, especially in the context of Mandelbrot's cascades, see for example [14] and the references therein. It can be easily seen that

$$Y_{n+1} = A\sum_{i=1}^{m} Y_{n,i} \tag{27}$$

where the $Y_{n,i}$ for $1 \leq i \leq m$ are independent of each other and independent of A and each of them has the same distribution as Y_n.
Therefore for $n \geq 1$, Y_n is square-integrable and

$$\operatorname{Var} Y_{n+1} = (\mathbb{E}|A|^2m^2 - 1) + m\mathbb{E}|A|^2 \operatorname{Var} Y_n,$$

where $\operatorname{Var} X = \mathbb{E}\left(|X - \mathbb{E}X|^2\right)$ denotes the variance of X. Since $m\mathbb{E}|A|^2 < 1$, the martingale $(Y_n)_n$ is bounded in L^2, so that (see Theorem 2.14 in Mailler's course) the following result holds.

$$Y_n \to Y_\infty \text{ a.s. and in } L^2$$

where Y_∞ is a (complex-valued) random variable with

$$\operatorname{Var}(Y_\infty) = \frac{\mathbb{E}|A|^2m^2 - 1}{1 - m\mathbb{E}|A|^2}.$$

Notice that, passing to the limit in (27) gives a new proof of the existence of a solution W of Equation (22) such that $\mathbb{E}W = 1$ and W has a finite second moment whenever $\Re(\lambda) > 1/2$.
The previous convergence allows to consider Y_∞ instead of W and a technical lemma then leads to the following theorem, showing that the exponential moments of W exist in a neighborhood of 0, so that the characteristic function of W is analytic at 0.

Theorem 4.14 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic equation (6) with $\Re(\lambda) > 1/2$ and let W be a solution of Equation (22). There exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $t \in \mathbb{C}$ with $|t| \leq \varepsilon$,*

$$\mathbb{E}e^{\langle t,W\rangle} \leq e^{\Re(t)+C|t|^2} \quad \textit{and} \quad \mathbb{E}e^{|tW|} \leq 4e^{|t|+2C|t|^2}.$$

References

[1] K.B. Athreya and P. Ney. *Branching Processes*. Springer, 1972.

[2] J. Barral, X. Jin, and B. Mandelbrot. Convergence of complex multiplicative cascades. *Ann. Appl. Probab.*, 20(4):1219–1252, 2010.

[3] J Bertoin and A. Rouault. Discretization methods for homogeneous fragmentations. *J. London Math. Soc.*, 72(1):91–109, 2005.

[4] J.D. Biggins. Martingale convergence in the branching random walk. *Adv in Appl. Prob.*, 10:62–84, 1978.

[5] B. Chauvin, Q. Liu, and N. Pouyanne. Limit distributions for multitype branching processes of m-ary search trees. *Ann. Inst. Henri Poincaré*, to appear, 2013.

[6] B. Chauvin, C. Mailler, and N. Pouyanne. Smoothing equations for large Pólya urns. *J. of Theor. Probab.*, pages 1–37, 2013.

[7] B. Chauvin and N. Pouyanne. m-ary search trees when m > 26: a strong asymptotics for the space requirements. *Random Structures and Algorithms*, 24(2):133–154, 2004.

[8] B. Chauvin, N. Pouyanne, and R. Sahnoun. Limit distributions for large Pólya urns. *Annals Applied Prob.*, 21(1):1–32, 2011.

[9] L. Devroye. *Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms*. Probabilistic Methods for Algorithmic Discrete Mathematics. Springer, M. Habib et al. edition, 1998.

[10] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002 edition.

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

[12] S. Janson. Functional limit theorem for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications*, 110:177–245, 2004.

[13] Q. Liu. Asymptotic properties of supercritical age-dependent branching processes and homogeneous branching random walks. *Stochastic Processes and their Applications*, 82(1):61–87, 1999.

[14] Q. Liu. Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stochastic Processes and their Applications*, 95:83–107, 2001.

[15] H.M. Mahmoud. *Evolution of Random Search Trees*. John Wiley & Sons Inc., New York, 1992.

[16] B. Pittel. On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.

[17] N. Pouyanne. An algebraic approach to Pólya processes. *Ann. Inst. Henri Poincaré*, 44(2):293–323, 2008.

[18] U. Rösler. A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 42:195–214, 1992.

[19] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001.

Automata and Motif Statistics

Pierre Nicodème

1 Motivation

Automata are used

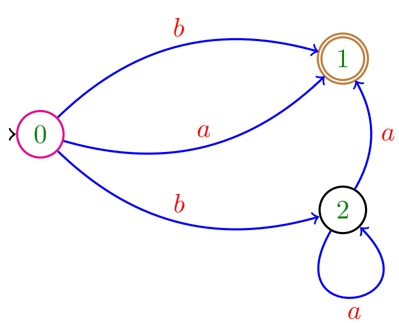
- in **hardware technology** (circuits)
- in **compilers** and **lexical analyzers**
- for **pattern matching**
- to build **groups** with **specific cogrowth**
- to compute **statistics of motifs** when a Motif is an infinite language or a very large language described by a regular expression (linguistics, bioinformatics, Web analysis)

2 Overview of the course

- Basics of Automata theory
- Pattern Matching
- Counting with automata in random texts
- Applications

3 Finite automata

3.1 What is an automaton?



An Automaton is

- **A directed graph,**
- **where vertices are called states,**
- **edges are called transitions,**
- **and labelled by letters of a finite alphabet;**
- **there is a specific state called start,**
- **and there are accepting states.**
- The function mapping the **states to their successors** is called **"transition function"**

AUTO $= (\mathcal{A}, Q, \text{start}, \delta, F)$

The automaton above

1. Alphabet - $\mathcal{A} = \{a, b\}$
2. Set of States - $Q = \{1, 2, 3\}$
3. start = $\{0\}$
4. Transition function δ: $\begin{cases} \delta(0,a) = \{1\} & \delta(0,b) = \{1,2\} \\ \delta(1,a) = \{\} & \delta(1,b) = \{\} \\ \delta(2,a) = \{2,1\} & \delta(2,b) = \{\} \end{cases}$
5. Accepting states: $F = \{1\}$

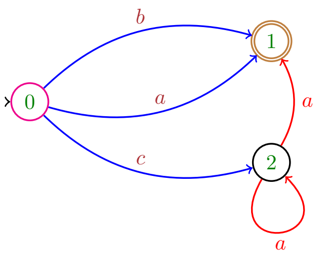
- A **run** of length n is a sequence $(q_0, q_1, \ldots, q_n)$ such that
  1. q_0 = start
  2. there exists $a_1 a_2 \ldots a_n \in \mathcal{A}^n$ and $q_{i+1} \in \delta(q_i, a_{i+1})$
- A word $w = a_1 a_2 \ldots a_n$ is **accepted** if there is **at least a run** of length n **spelling its letters** and **ending** in an **accepting state**.
- The **set of words accepted** by the automaton is the **language recognized** by the automaton.
  (A **language** is a **possibly infinite set of words**)

Examples

- **Some not accepted words:**
  $c, a^m, ab, b^n \qquad (m \geq 2, n \geq 2)$
- **Some accepted words:**
  $a, b, ca^n \qquad (n \geq 1)$
- **The recognized language** (an infinite set of words in the present case)
  $a + b + ca^+ \qquad (a^+ = \sum_{n \geq 1} a^n)$

3.2 Different classes of Automata

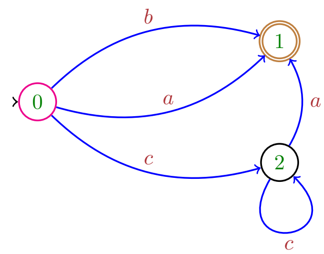
3.2.1 Deterministic and Non-Deterministic automata



A **NFA**
(**Non-deterministic Finite Automaton**)

$|\delta(2, a)| = \big|\{2, 1\}\big| > 1$

Several successors
with the **same letter**



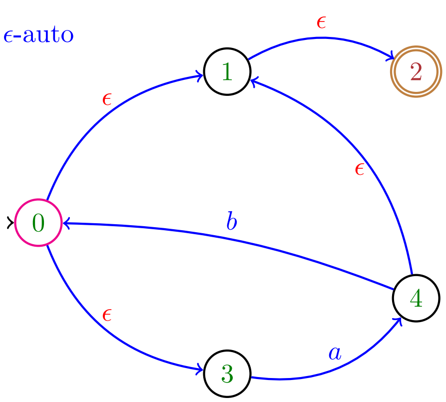
A **DFA**
(**Deterministic Finite Automaton**)

$\forall q \in Q, \forall \ell \in \mathcal{A}, |\delta(q, \ell)| = 1$

Only one successor
with **one letter** at **each state**

3.2.2 Finite Automata with $\epsilon-$transitions

ϵ-auto



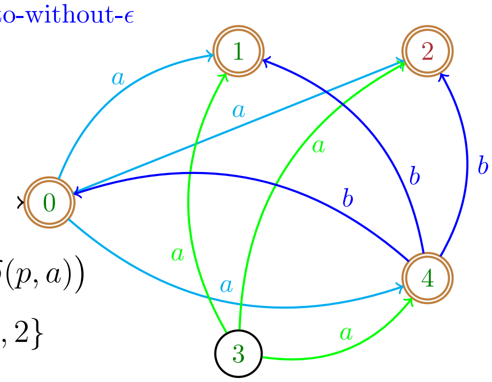
$$\epsilon\text{-}cl(4) = \{4, 1, 2\}$$

$$F' = \{0, 1, 4, 2\}$$

$$\epsilon\text{-}cl(0) = \{0, 1, 2, 3\}$$

$$\Delta(0, a) = \epsilon\text{-}cl\left(\bigcup_{p \in \{0,1,2,3\}} \delta(p, a)\right)$$
$$= \epsilon\text{-}cl\left(\{4\}\right) = \{4, 1, 2\}$$

auto-without-ϵ



- ϵ-auto $= \big(\mathcal{A} = \{a, b, \epsilon\}, Q = \{0, 1, 2, 3, 4\}, s = 0, \delta, F = \{2\}\big)$
- An ϵ-transition consumes **no input** (no letter of the alphabet different of ϵ)
- **ϵ-closure**: $\forall q \in Q,\ \epsilon\text{-}cl(q) := \{p \mid p \text{ is accessible from } q \text{ without consuming input}\}$

Build an **automaton without ϵ-transition** that recognizes the **same language**

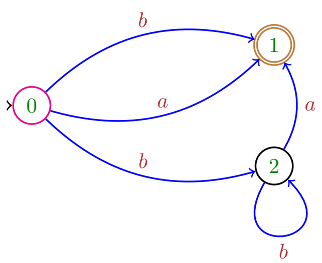
- auto-without-ϵ $= \big(\mathcal{A} = \{a, b\}, Q, s, \Delta, F'\big)$
- $F' = F \bigcup \{q \mid \epsilon\text{-}cl(q) \bigcap F \neq \emptyset\} = \{0, 4, 1, 2\}$
- $\Delta(q, \ell) = \epsilon\text{-}cl(\bigcup_{p \in \epsilon\text{-}cl(q)} \delta(p, \ell))$

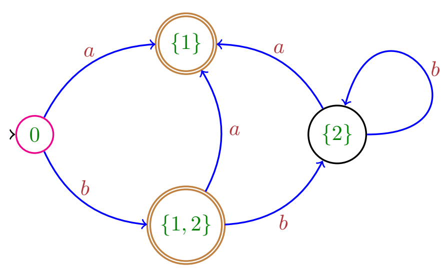
$$
\begin{array}{lll}
\epsilon\text{-}cl(0) = \{0,1,2,3\} & \Delta(0,a) = \{4,1,2\} & \Delta(3,a) = \{4,1,2\} \\
\epsilon\text{-}cl(1) = \{1,2\} & \Delta(0,b) = \{\} & \Delta(3,b) = \{\} \\
\epsilon\text{-}cl(2) = \{2\} & \Delta(1,a) = \Delta(1,b) = \{\} & \Delta(4,a) = \{\} \\
\epsilon\text{-}cl(3) = \{3\} & \Delta(2,a) = \Delta(2,b) = \{\} & \Delta(4,b) = \{0,1,2\} \\
\epsilon\text{-}cl(4) = \{4,1,2\} & &
\end{array}
$$

Remark. Usually, the resulting automaton is a NFA.

3.2.3 Determinization of an automaton

$M_{\text{NFA}} = (\mathcal{A}, Q, 0, \delta, F)$ $\qquad$ $M'_{\text{DFA}} = (\mathcal{A}, Q', 0, \Delta, F')$





$\Delta(0,a) = \{1\}$ $\qquad$ $\Delta(\{1,2\},b) = \{2\}$

$\Delta(0,b) = \{1,2\}$ $\qquad$ $\Delta(\{2\},b) = \{2\}$

$\Delta(\{1,2\},a) = \{1\}$ $\qquad$ $\Delta(\{2\},a) = \{1\}$

- $Q' \subset 2^Q$ (the **subsets** of Q)
- $s' = s$
- $F' = \{f \in Q'; \quad f \cap F \neq \emptyset\}$ $\left\{ \begin{array}{l} \text{the subsets that contain} \\ \text{at least one accepting state of } M \end{array} \right.$
- $\forall S \in Q', \forall \ell \in \mathcal{A}, \quad \Delta(S, \ell) = \bigcup_{q \in S} \delta(q, \ell)$

Definition 3.1. ***Two automata*** *$M = (Q, \mathcal{A}, s, \delta, F)$ and $M' = (Q', \mathcal{A}', s', \delta', F')$ are* ***equivalent*** *if they recognize the* ***same language*** *($\mathcal{L}(M) = \mathcal{L}(M')$)*

The automata M_{NFA} and M'_{DFA} are equivalent

- each **accepted run** of M_{NFA} **translates** to **an accepted run** of M_{NFA}
- each **non accepted run** of M'_{DFA} is the **translation** of a **non accepted run** of M_{DFA}

Theorem 3.1 (Rabin-Scott 1959)**.** *Let $M = (Q, \mathcal{A}, s, \Delta, F)$ be a NFA. Then there exists a DFA $M' = (Q', \mathcal{A}', s', \delta', F')$ that is equivalent to M.*

Proof by induction

Remark 3.1. *Each* **DFA** *is a* **NFA**.

Corollary 3.1. *(i) The NFA's are no more powerful than the DFAs in terms of the languages they accept.*
(ii) The NFA's and DFA's recognize the same set of languages.

4 Regular Expressions and Regular Languages

Surprisingly, there is another fully different characterization of languages recognized by Finite Automata, the Regular Languages.

4.1 What is a Regular Language?

Definition 4.1. *Let $\mathcal{A}$ be a* **finite alphabet**.
The collection of **regular languages** *over $\mathcal{A}$ is defined* **recursively** *by*

1. *$\emptyset$ is a regular language*
2. *$\{\epsilon\}$ is a regular language*
3. *$\{\ell\}$ is a regular language for each $\ell \in \mathcal{A}$*
4. *if A and B are regular languages, so are*
   - $A \bigcup B$ (Ex: $\{ab\} \bigcup \{c\} = \{ab, c\}$)
   - $A \bullet B$ (Ex: $\{ab, c\} \bullet \{d, e\} = \{abd, cd, abe, ce\}$)
   - $A^\star$ (Ex: $\{ab\}^\star = \{\epsilon, ab, abab, \dots, (ab)^n, \dots\}$)
5. *No other languages over $\mathcal{A}$ are regular*

Regular expressions are **shorthands** for **regular languages**

$a+b$	denotes	$\{a, b\} = \{a\} \bigcup \{b\}$
ab	denotes	$\{ab\} = \{a \bullet b\}$
$a^\star$	denotes	$\{a\}^\star$
a^+	denotes	$a.a^\star = a \bullet a^\star$

4.2 Formal definition of Regular Expressions

Regular expressions are defined recursively by

1. $\emptyset$ and ϵ are regular expressions
2. ℓ is a regular expressions for each $\ell \in \mathcal{A}$
3. if r and s are regular expressions, so are

- $r + s$
- $r.s$
- $r^\star$

4. No other sequence of symbols is a regular expression.

Lemma 4.1. *Every regular language can be accepted by a finite automaton*

Lemma 4.2. *Every language accepted by a finite automaton is regular*

Theorem 4.1 (Kleene 1956)**.** *A language is regular if and only if it is accepted by a Finite Automaton*

Proof of Lemma 4.1.

1. **Atomic Languages**

$$\begin{array}{ccc} \emptyset & \text{is accepted by} & (\mathcal{A}, \{0\}, 0, \delta = \emptyset, \emptyset) \\ \epsilon & \text{is accepted by} & (\mathcal{A}, \{0\}, 0, \delta = \emptyset, \{0\}) \\ \ell \in \mathcal{A} & \text{is accepted by} & (\mathcal{A}, \{0,1\}, 0, \delta(0, \ell) = \{1\}, \{1\}) \end{array}$$

2. let $\mathcal{L}_1$ and $\mathcal{L}_2$ **regular languages** respectively **accepted** by automata A_1 and A_2.

$$\begin{array}{ccc} \mathcal{L}_1.\mathcal{L}_2 & \text{is accepted by} & A_1.A_2 \\ \mathcal{L}_1 + \mathcal{L}_2 & \text{is accepted by} & A_1 \cup A_2 \\ \mathcal{L}_1{}^\star & \text{is accepted by} & A_1{}^\star \end{array}$$

Starting from the atomic languages, one **builds recursively** a **ϵ-NFA** **recognizing a given regular expression**

Proof of Lemma 4.2 - From Finite Automata to Regular Expressions.

$A = (\mathcal{A} + \epsilon, \{q_1, q_2, \ldots, q_m\}, S \subseteq Q, \delta, F \subseteq Q)$ a finite automaton

1. let $L(i,j,k) = \left\{ w \;\middle|\; \begin{array}{l} w \text{ is the label of a path from } q_i \text{ to } q_j \\ \text{where intermediate nodes have labels } \leq k \end{array} \right\}$
2. $L(i,j,0)$ has no intermediate labels $\Longrightarrow L(i,j,0) \subseteq \mathcal{A} \cup \epsilon$ is regular
3. Assume $L(i,j,k)$ regular and consider $L(i,j,k+1)$
   Let p be a path form q_i to q_j where **intermediate nodes** have **labels** $\leq k+1$.
   - (a) $p \in L(i,j,k)$ (the path p does not reach q_{k+1})
   - (b) p begins at q_i,reaches q_{k+1} a **first time**, possibly **other times**, until a **last time**, and ends at q_j

   Cases (a) and (b) give
   $L(i,j,k+1) = L(i,j,k) \cup L(i,k+1,k) L(k+1,k+1,k)^\star L(k+1,j,k)$
   Therefore $L(i,j,k+1)$ is **regular**
4. In particular $L(i,j,m)$ is regular

Conclusion: $L(A) = \bigcup \left\{ L(i,j,m) | q_i \in S, q_j \in F \right\}$ is **regular**, since it is a **finite union of regular languages**

5 Counting

5.1 Generating Functions of Languages

$\mathcal{L}$ a language (a possibly infinite set of words)

– **Enumeration**

$$L(z) = \sum_{w \in \mathcal{L}} z^{|w|} = \sum_{n \geq 0} l_n z^n$$

where l_n is the **number of words** of **length** n of $\mathcal{L}$

– **Weighted generating Function**

$$W(z) = \sum_{w \in \mathcal{L}} \mathbf{P}(w) z^{|w|} = \sum_{n \geq 0} p_n z^n$$

where p_n is the **probability** that a **random word of length** n belongs to $\mathcal{L}$

– Enumeration

$L(a,b) = \sum_{w \in \mathcal{L}} a^{|w|_a} b^{|w|_b} = \sum_{i,j} l_{i,j} a^i b^j$

$l_{i,j}$ = **number of words in the language with** $\begin{cases} i \text{ letters } a \\ j \text{ letters } b \end{cases}$

$F(z) = L(z,z) = \sum_n f_n z^n$, f_n = **number of words of length n in the language**

– Weighted counting $F(z) = L(\mathbf{P}(a)z, \mathbf{P}(b)z) = \sum_n p_n z^n$

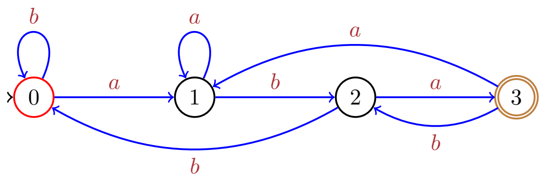
p_n = **probability that a word of length** n **is in the language**

5.2 Generating Function of a Regular Expression

The following algorithm is usually attributed to Chomsky-Schützenberger (1963), but may be older.
We provide here a trivial example, but the algorithm used is fully general.

$P = \mathcal{A}^\star aba = (a+b)^\star aba$

Build the automaton that **accepts** the language defined by P; it recognizes the set of words **terminating** with aba



Define $\mathcal{L}_i$ as the **language of runs** $\begin{cases} \text{that } \textbf{start at state } i \\ \text{and } \textbf{terminate in an accepting state} \end{cases}$

$$\mathcal{L}_1 = ba + a^\star ba + ba(ba)^\star + \dots$$

$$\begin{array}{ll}
\mathcal{L}_0 = a.\mathcal{L}_1 + b.\mathcal{L}_0 & L_0(a,b) = a \times L_1(a,b) + b \times L_0(a,b) \\
\mathcal{L}_1 = a.\mathcal{L}_1 + b.\mathcal{L}_2 & L_1(a,b) = a \times L_1(a,b) + b \times L_2(a,b) \\
\mathcal{L}_2 = a.\mathcal{L}_3 + b.\mathcal{L}_0 & L_2(a,b) = a \times L_3(a,b) + b \times L_0(a,b) \\
\mathcal{L}_3 = a.\mathcal{L}_1 + b.\mathcal{L}_2 + \epsilon & L_3(a,b) = a \times L_1(a,b) + b \times L_2(a,b) + 1
\end{array}$$

solve: $\quad L_0(a,b) = \dfrac{1}{1-(a+b)} \times aba \qquad F(z) = \sum p_n z^n = L_0(\mathbf{P}(a)z, \mathbf{P}(b)z)$

The resulting generating function is always the solution of a linear system of equations, and therefore a rational function.

5.3 Asymptotics of a rational function

– if $F(z) = \dfrac{P(z)}{Q(z)}$ with $P(\rho \neq 0),\ Q(\rho = 0)$

– and ρ **real**, **positive**, **dominant singularity** of **order** k

Then,

$$f_n = [z^n]F(z) = \frac{P(\rho)}{Q(\rho)} \times \rho^{-n} \times (n-k+1) \times (1+A^n) \qquad (A<1)$$

Expand the **polynomial** $P(z)$ at ρ

$$P(z) = P(\rho) + (z-\rho)P'(\rho) + \frac{1}{2!}(z-\rho)^2 P''(\rho) + \dots$$

to get a **full expansion**

Generating Functions of Regular Languages

1. Any regular expression is recognized by a Finite Automaton
2. The Chomsky-Schützenberger algorithm **applies** to **any regular expression**.

Theorem 5.1 (Chomsky-Schützenberger 1963)**.** *The generating function of a regular language is rational.*

Corollary 5.1. *Let $\mathcal{R}$ a regular language and $\mathcal{R}_n = \mathcal{R} \cap \mathcal{A}^n$. $\exists n_0,\ \forall n > n_0,\quad |\mathcal{R}_n| = p_1(n)\lambda_1^n + \dots + p_k(n)\lambda_k^n$, with $p_i(n)$ complex polynomials and $\lambda_i \in \mathbb{C}$*

5.4 An asymptotic test of non-regularity

For any regular language $\mathcal{R}$, there exists a real positive number λ and a polynomial $p(n)$ such that

$$\lim_{n\to\infty} r_n = \lambda^n \times p(n), \qquad r_n = \left|\mathcal{R} \bigcap \mathcal{A}^n\right|$$

- The number of words of length $2n$ in **Dyck Languages** ((()(()))) is the Catalan number $\binom{2n}{n}/(2n+1)$ asymptotic to $\dfrac{4^n}{n^{3/2}\sqrt{\pi}}$.

  Dyck languages are **not regular** and **cannot be recognized by a DFA**; however they can be recognized by a push-down automaton, and they have an algebraic generating function.
- Let $\pi(x)$ be the **number of prime numbers less** than $x \in \mathbb{R}^+$.

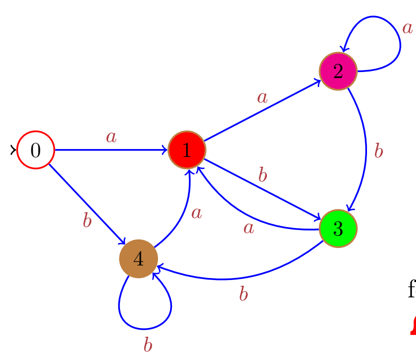
$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1$$

  There is **no known generating function enumerating the primes**. Would one find one it would **not be regular**. It is **not possible to enumerate** the **primes** by an **automaton**.

6 Some classical pattern matching algorithms

6.1 Aho-Corasick (1975) - Finite Motif - Multiple Counting

$P = \{a, aa, ab, b\}$



$\delta(2, a) = 2 \quad w_2.a = a.aa$

$\delta(2, b) = 3 \quad w_2.b = a.ab$

$\delta(3, a) = 1 \quad w_3.a = ab.a$

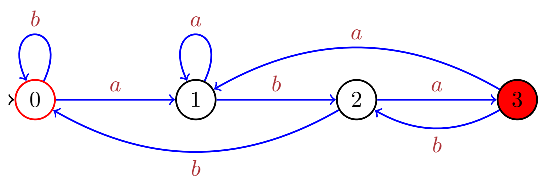
$\delta(3, b) = 4 \quad w_3.b = ab.b$

for **each specific match** ring a **bell**

1. Build a **trie** (a tree that is here equivalent to an automaton) recognizing all the words of P
   - let Q be the set of nodes of the trie: $Q = \{0, 1, 2, 3, 4\}$
   - $\forall q \in Q$, let w_q the word spelling the run from 0 to q – (as instance $w_3 = ab$)
2. for each **node** q with a **missing transition** ℓ
   - **add a transition** $\delta(q, \ell)$ to state q'
   - such that $w_{q'}$ is **the longuest possible suffix** of $w_q.\ell$

6.2 Knuth-Morris-Pratt automaton (1977) - Only one word

$P = aba$



– **same construction** as Aho-Corasick

– for **each match** ring the **bell**

– $aaaaaba$🔔$bbaba$🔔ba🔔bb

7 Statistics of Motifs

We learned how to compute the **number of matches** of a **finite pattern in a random text**.

What about **counting** the **occurrences** of a **Regular Expression** in such texts?

7.1 Tools and Aim - Generating Functions

For a **given pattern** P, we want to compute

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

$$\text{where} \quad f_{n,k} = \mathbf{P}\left(\begin{array}{l} P \text{ occurs } k \text{ times} \\ \text{in a } \textbf{random text of length } n \end{array}\right)$$

If X_n is the random variable

– counting the number of occurrences of P

– in a random text of size n

$$F(z, u) = \sum_{n\geq 0, k\geq 0} f_{n,k} u^k z^n = \sum_{n\geq 0} z^n \sum_{k\geq 0} \mathbf{P}(X_n = k) u^k$$

The variables z and u are formal variables

– z is related to the length of the texts

– u is related to the number of occurrences of P

7.2 Counting with Regular Expressions - The right language

1. **Input**:

   – a finite alphabet $\mathcal{A}$

   – a regular expression $\mathcal{R}$

2. **Output**: $F(z, u) = \sum_{n\geq 0, k\geq 0} f_{n,k} u^k z^n$,

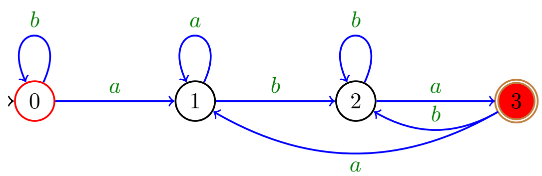
   where $f_{n,k}$ is the number of occurrences of the pattern $\mathcal{R}$ in a random sequence of length N.

Method

1. Build the DFA recognizing $\mathcal{A}^\star.\mathcal{R}$
2. Use a variant of Chomsky-Schützenberger to ring the bell and produce the variable u

Counting the number of occurrences of ab^+a

$$P = \mathcal{A}^\star ab^+a = (a+b)^\star ab^+a$$



We define $\mathcal{L}_i$ as the language of words that are recognized by the automaton, with the condition that **state i is chosen as initial state**.
This leads to the linear set of equations on languages $\{\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3\}$.
There is a particular case for **state** 3, where you must **ring the bell**; this **translates** to the **formal parameter u**.

$$\begin{aligned}
\mathcal{L}_0 &= a.\mathcal{L}_1 + b.\mathcal{L}_0 & L_0(a,b,u) &= a\times L_1(a,b,u) + b\times L_0(a,b,u)\\
\mathcal{L}_1 &= a.\mathcal{L}_1 + b.\mathcal{L}_2 & L_1(a,b,u) &= a\times L_1(a,b,u) + b\times L_2(a,b,u)\\
\mathcal{L}_2 &= a.\mathcal{L}_3 + b.\mathcal{L}_2 & L_2(a,b,u) &= a\times u\times L_3(a,b,u) + b\times L_2(a,b,u)\\
\mathcal{L}_3 &= a.\mathcal{L}_1 + b.\mathcal{L}_2 + \epsilon & L_3(a,b,u) &= a\times L_1(a,b,u) + b\times L_2(a,b,u) + 1
\end{aligned}$$

Solve:

$$L_0(a,b,u) = \frac{1-b+ab-uab}{1-a-2b+2ab+b^2-ab^2-u(ab-ab^2)}, \quad F(z,u) = \sum f_{n,k}u^k z^n = L_0(\mathbf{P}(a)z, \mathbf{P}(b)z, u)$$

$$\mathbf{P}(a) = \mathbf{P}(b) = \frac{1}{2} \quad \rightsquigarrow \quad F(z,u) = \frac{8-4z+2z^2-2uz^2}{8-12z+6z^2-z^3-u(2z^2-z^3)}$$

Once again, **this method is fully general**.

7.3 Exploiting the generating Function

$$R = ab^+a, \quad F(z,u) = \frac{8-4z+2z^2-2uz^2}{8-12z+6z^2-z^3-u(2z^2-z^3)}$$

– Expand in series with respect to z in the neighborhood of 0

$$F(z,u) = 1+z+z^2+\left(\frac{1}{8}u+\frac{7}{8}\right)z^3+\left(\frac{5}{16}u+\frac{11}{16}\right)z^4+\left(\frac{1}{2}+\frac{15}{32}u+\frac{1}{32}u^2\right)z^5+\mathcal{O}(z^6)$$

– Compute the generating function of the expectations of the number of occurrences of the pattern

$$E(z) = \sum_n \mathbf{E}(X_n)z^n = \left.\frac{\partial F(z,u)}{\partial u}\right|_{u=1} = -\frac{1}{2}\frac{z^2}{1-z}+\frac{1}{4}\frac{z^2}{1-\frac{1}{2}z}+\frac{1}{4}\frac{z^2}{(1-z)^2}$$

– Get $\mathbf{E}(X_n)$

$$\mathbf{E}(X_n) = -\frac{1}{2}+2^{-n}+\frac{1}{4}(n-1) = \frac{1}{4}(n-3)+2^{-n}$$

$$R = ab^+a, \quad F(z,u) = \frac{8-4z+2z^2-2uz^2}{8-12z+6z^2-z^3-u(2z^2-z^3)}$$

- **Generating function of the Second Moment** $M_2(z) = \sum_{n\geq 0} \mathbf{E}(X_n^2) z^n = \frac{\partial}{\partial u} u \frac{\partial F(z,u)}{\partial u}\Big|_{u=1}$

$$M_2(z) = \frac{1}{4}\frac{z^2(z^2-2)}{1-z} - \frac{1}{4}\frac{z^2(z^2-1)}{(1-z)^2} - \frac{1}{8}\frac{z^2(z^2-2)}{1-\frac{z}{2}} + \frac{1}{8}\frac{z^4}{(1-z)^3}$$

- **Extract** the nth. **Taylor coefficient**

$$\mathbf{E}(X_n^2) = [z^n]M_2(z) = \frac{1}{16}n^2 - \frac{5}{16}n + \frac{5}{8} - 2^{-n}$$

- **Standard Deviation** σ_n

$$\sigma_n = \sqrt{\mathbf{E}(X_n^2) - \mathbf{E}^2(X_n)} = \frac{1}{4}\sqrt{n + 1 - 2^{-n+3}n + 2^{-n+3} - 4^{-n+2}}$$

7.4 Limit law

- **Laplace transform L** of a **random variable** X of density function $f(x)$

$$\mathbf{L}(X,t) = \mathbf{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

- **Laplace transform** of a **standard Gaussian variable** $\mathcal{N}$

$$\mathbf{L}(\mathcal{N},t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}dx = e^{t^2/2}$$

Theorem 7.1 (Paul Lévy Continuity Theorem - 1925)**.**

If, for $t \in [-\alpha, +\alpha]$, $\lim_{n\to\infty} \mathbf{E}(e^{tX_n}) = \mathbf{L}(\mathcal{N}) = e^{t^2/2}$,

then $X_n \xrightarrow{\mathcal{D}} \mathcal{N}$ *(convergence in distribution or law)* : $\lim_{n\to\infty} \mathbf{P}(X_n < x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-w^2/2}dw$

7.5 Limit law of the number of occurrences of ab^+a.

We assume that $\mathbf{P}(a) = \mathbf{P}(b) = 1/2$

$$F(z,u) = \frac{8-4z+2z^2-2uz^2}{8-12z+6z^2-z^3-u(2z^2-z^3)} = -\frac{1-u}{u\left(1-\frac{z}{2}\right)} + \frac{1+\sqrt{u}}{2u\left(1-z\frac{1+\sqrt{u}}{2}\right)} + \frac{1-\sqrt{u}}{2u\left(1-z\frac{1-\sqrt{u}}{2}\right)} \quad (1)$$

$$\Psi_n(u) = [z^n]F(z,u) = \frac{1}{u}\left(\frac{1+\sqrt{u}}{2}\right)^{n+1} + O\left(\frac{1}{2^n}\right) \qquad \text{for } u \text{ close of } 1$$

We consider $\Psi_n(e^t) = \mathbf{E}(e^{tX_n})$ and the **normalised law** $\frac{X_n - \mu_n}{\sigma_n}$

$$\Phi_n(t) = \Psi_n\left(t\frac{X_n-\mu_n}{\sigma_n}\right) = \mathbf{E}\left[\exp\left(\frac{t(X_n-\mu_n)}{\sigma_n}\right)\right] = \exp\left(-\frac{\mu_n t}{\sigma_n}\right)\mathbf{E}\left[\exp\left(\frac{tX_n}{\sigma_n}\right)\right]$$

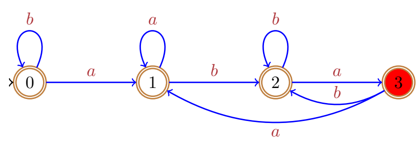
We substitute: $\mu_n = \dfrac{n-3}{4} + \mathcal{O}(2^{-n}), \quad \sigma_n = \dfrac{\sqrt{n+1}}{4} + \mathcal{O}(2^{-n})$

In a **neighborhood** of $t=0$, we **expand** $\log(\Phi_n(t))$

$$\log(\Phi_n(t)) = \frac{t^2}{2} - \frac{t^4}{12(n+1)} + \mathcal{O}\left(\frac{t^6}{n^2}\right), \qquad \lim_{n\to\infty} \log(\Phi_n(t)) = \frac{t^2}{2}$$

7.6 The Gaussian law is general

$R = ab^+a \qquad P = \mathcal{A}^\star ab^+a$



$$\begin{aligned} L_0(z,u) = L_0 &= zp_aL_1 + zp_bL_0 + 1, \\ L_1 &= zp_bL_2 + zp_aL_1 + 1, \\ L_2 &= zp_a u L_3 + zp_bL_2 + 1 \\ L_3 &= zp_aL_1 + zp_bL_2 + 1 \end{aligned}$$

General case: $\mathbf{L} = \begin{pmatrix} L_0 \\ \vdots \\ L_n \end{pmatrix} = z\mathbb{T}(u)\mathbf{L} + \mathbf{1}$, and $\mathbb{T}(u)$ **positive** $n \times n$ matrix for $u \geq 0$

Theorem 7.2 (Perron-Frobenius, 1907-1912)**.** *If $\mathbb{T}(u)$ is positive, irreducible and aperiodic, the dominant eigenvalue is unique, real and positive.*

$$L_0(z,u) = \frac{P(z,u)}{Q(z,u)} = \frac{P(z,u)}{(1-z\lambda_1(u))\cdots(1-z\lambda_n(u))} \qquad \lambda_i(u) \text{ eigenvalue of } \mathbb{T}(u)$$

$$\lambda_1(u) \text{ dominant } \implies \frac{1}{|\lambda_1(u)|} < \frac{1}{|\lambda_2(u)|} \leq \dots$$

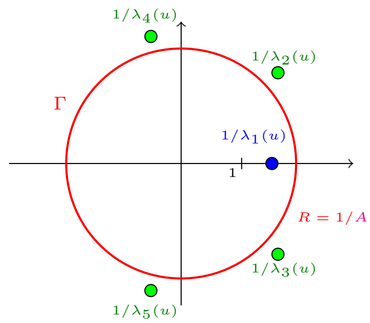
Perron-Frobenius conditions in the context of automata.

- **irreducibility**: from any state, any other state can be reached (The above automaton is not irreducible)
- **primitivity**: there exists a large enough e such that any state can be reached by any other state in exactly e steps

Remark 7.1.

- *The above automaton with initial state* 1 *and states* $1,2,3$*, is irreducible and primitive*
- *The automaton with states* $0,1,2,3$ *is such that* $L_0 = \dfrac{L_1}{1-zp_b} + \dfrac{1}{1-zp_b}$
- *For* $u=1$*, we have* $L_0 = L_1 = L_2 = L_3 = 1/(1-z)$
- *by continuity,* $\lambda_1(u)$ *is close of* 1 *for* $u \in [1-\epsilon, 1+\epsilon]$
- *for* L_0*, we have* $\dfrac{1}{\lambda_1(u)} < \dfrac{1}{p_b}$

Uniform Separation Property with respect to n



$$\begin{aligned} p_n(u) &= [z^n]F(z,u) = \frac{1}{2i\pi}\oint_\Gamma \frac{dz}{z^{n+1}}F(z,u) = \frac{1}{2i\pi}\oint_\Gamma \frac{c(u)}{z^{n+1}(1-\lambda_1(u)z)} + \frac{1}{z^{n+1}}g(z,u)\,dz, \quad (2)\\ &= c(u)\lambda_1(u)^n(1+O(A^n)) \qquad (A<1), \end{aligned}$$

where $g(z,u)$ has no singularity inside the disk $|z| \leq$ radius of Γ.

Hwang's **quasi-power** theorem $\rightarrow$ limiting **Gaussian distribution**

Variability condition: $\lambda''(1) + \lambda'(1) - \lambda'(1)^2 \neq 0 \qquad (\lambda(u) = \lambda_1(u))$

7.7 Statistics of one regular motif

Let X_n count the number of occurrences of a regular motif R in a random text of length n. With $g(z,u)$ defined as in Equation (2), we have

$$F(z,u) = \sum_{n,k} \mathbf{P}(X_n = k)u^k z^n = \frac{c(u)}{1-\lambda(u)z} + g(z,u)$$

Theorem 7.3 (N, Salvy, Flajolet - 1999)**.** *Both in the* **Bernoulli** *and* **Markov** *model, with* $\mathbb{T}(u)$ *the fundamental matrix, and* $\lambda(u)$ *its dominant eigenvalue,*

1. $F(z,u)$ *is* **rational** *and* **can be computed** *explicitly*
2. **Moments** $\begin{cases} \mathrm{E}(X_n) & = \lambda'(1)n + c_1 + O(A^n), \quad (c_1 = c'(1)) \\ \mathrm{Var}(X_n) & = (\lambda''(1) + \lambda'(1) - \lambda'(1)^2)n + c_2 + O(A^n) \\ & \qquad\qquad (c_2 = c''(1) + c'(1) - c'(1)^2) \end{cases}$
3. **Limit Gaussian law**: $\qquad \Pr\left(\frac{X_n - \mu n}{\sigma\sqrt{n}}\right) \rightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt$

[Bourdon, Vallée - 2006] Extension to **dynamical sources**

Counts of $R = ab^+a$ Assuming again $\mathbf{P}(a) = \mathbf{P}(b) = \frac{1}{2}$,

with X_n number of occurrences of R in a random text of size n

we have

$$\sigma_n = \sqrt{\mathbf{Var}(X_n)} = \frac{\sqrt{n+1}}{4} + \mathcal{O}(2^{-n})$$

The **Variability condition:** is **verified**

$$\mathbf{Var}(X_n) = (\lambda''(1) + \lambda'(1) - \lambda'(1)^2)n + c_2 + O(A^n) = \Theta(n)$$

We have $\mathbf{Var}(X_n) = \Theta(n) \implies$ **normal limit law**

Counts of $R = ab^\star$ $\mathbf{P}(a) = \mathbf{P}(b) = \frac{1}{2}$

$$F(z,u) = \sum_{n\geq 0}\sum_{k\geq 0} \mathbf{P}(X_n = k)u^k z^n = \frac{uz/2 - 1}{1 - z/2 - uz + uz^2}$$

$$\begin{cases} \mathbf{E}(X_n) = n - 1 + 2^{-n} \\ \mathbf{E}(X_n^2) = n^2 - 2n + 3 - 3\times 2^{-n} \\ \mathbf{Var}(X_n) = 2 - (2n+1)2^{-n} - 4^{-n} \\ \lim_{n\to\infty} \mathbf{Var}(X_n) = 2 \end{cases}$$

- The **variation condition** is **not verified**
- The **limiting law** is **not normal**

Hwang's Quasi-Power theorem - Gaussian form

$$\textbf{Notation: } m(f) = \frac{f'(1)}{f(1)}, \qquad v(f) = \frac{f''(1)}{f(1)} + \frac{f'(1)}{f(1)} - \left(\frac{f'(1)}{f(1)}\right)^2$$

Theorem 7.4 (Hwang 1994)**.** *Let the X_n be non-negative discrete random variables (supported by $\mathbb{Z}_{\geq 0}$) with probability generating function $p_n(u)$. Assume that, uniformly in a complex neighborhood of $u = 1$, for sequences $\beta_n, \kappa_n \to \infty$, there holds*

$$p_n(u) = A(u).B(u)^{\beta_n}\left(1 + \mathcal{O}\left(\frac{1}{\kappa_n}\right)\right),$$

where $A(u), B(u)$ are analytic at $u = 1$ and $A(1) = B(1) = 1$. Assume finally that $B(u)$ satisfies the so-called "variability condition",

$$v(B(u)) \equiv B''(1) + B'(1) - B'(1)^2 \neq 0.$$

Under these conditions, the mean and variance of X_n satisfy

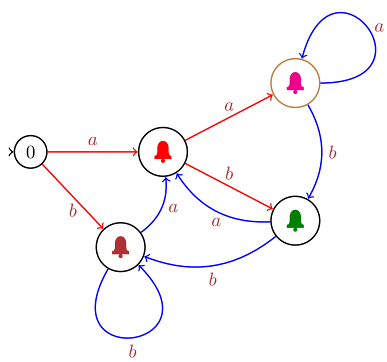
$$\begin{aligned}\mu_n \equiv \quad & \mathbf{E}(X_n) = & \beta_n m(B(1)) + m(A(1)) + \mathcal{O}\left(\kappa_n^{-1}\right)\\ \sigma_n^2 \equiv \quad & \mathbf{Var}(X_n) = & \beta_n v(B(1)) + v(A(1)) + \mathcal{O}\left(\kappa_n^{-1}\right).\end{aligned}$$

The distribution of X_n is, after standardization, asymptotically Gaussian,

$$Pr\left\{\frac{X_n - \mathbf{E}(X_n)}{\sqrt{\mathbf{Var}(X_n)}} \leq x\right\} = \mathcal{N}(x) + \mathcal{O}\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\beta_n}}\right),$$

7.8 What about counting with several motifs simultaneously?

$P = \{a, aa, ab, b\}$ **Several Finite Motifs**



Where are the **bells**?
Easy: upon some **nodes** of the **trie**

It is **not so easy** for several **general regular motifs**

7.9 Product of Marked Automata

The product of automata is classical in automata theory. For two automata

- $\text{Auto}_1 = (\mathcal{A}, Q_1, s_1, \delta_1, F_1)$,
- $\text{Auto}_2 = (\mathcal{A}, Q_2, s_2, \delta_2, F_2)$,

The product automaton $\mathbf{P} = \text{Prod}(\text{Auto}_1, \text{Auto}_2)$ is defined as: $\mathbf{P} = (\mathcal{A}, \mathbf{Q} \subseteq Q_1 \times Q_2, (s_1, s_2), \Delta, \mathbf{F})$ where

$$\forall q_1 \in Q_1, q_2 \in Q_2, \forall \ell \in \mathcal{A}, \quad \Delta\big((q_1, q_2), \ell\big) = \big(\delta_1(q_1, \ell), \delta_2(q_2, \ell)\big)$$

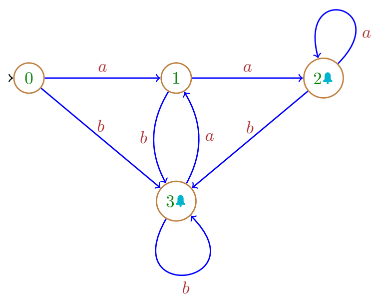
$$\mathbf{F} = \{(q_i, q_j) \text{with } q_i \in F_1 \text{ or } q_j \in F_2$$

Remarks. Like for the determinization of an automaton, the algorithm generating the product automaton starts from the initial state (s_1, s_2), and only the accessed states encountered during the algorithm are generated to build $\mathbf{Q}$
The construction is quadratic in the worst case with respect of the size of the two initial automata.
The product of more than two automata follows the same rules.

We need however to **distinguish the type of terminal states** with respect to the **corresponding match** within the **multiple pattern** by **assigning different marks** to them.

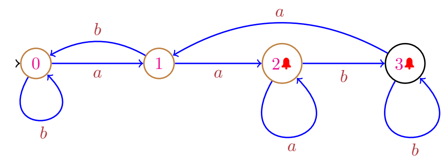
$U = aa + b$

$\mathrm{AutoU} = (\mathcal{A}, 0, Q, \delta, F = Q, \mathrm{Mark} = \{2, 3\})$

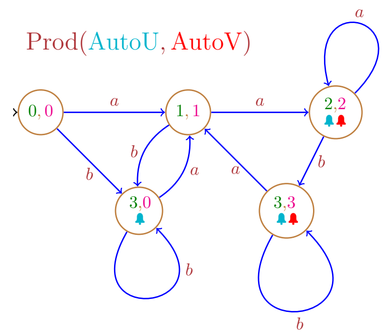


$V = b^\star aab^\star$;

$\mathrm{AutoV} = (\mathcal{A}, 0, Q, \delta, F = Q, \mathrm{Mark} = \{2, 3\})$



Prod(AutoU, AutoV)

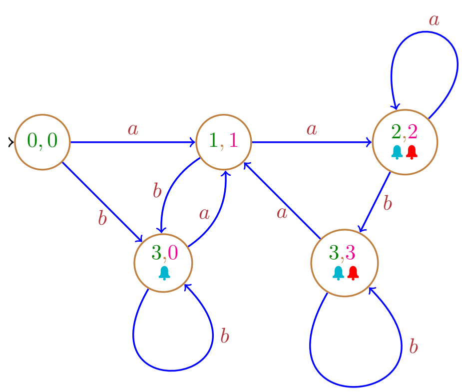


$$\mathrm{Prod}(\mathrm{AutoU}, \mathrm{AutoV}) = \Big(\mathcal{A}, (0,0), \mathbf{Q} \subseteq Q \times Q, \Delta, \mathbf{F} = \mathbf{Q},$$
$$\mathrm{Mark}_1 = \{(2,2), (3,0), (3,3)\},$$
$$\mathrm{Mark}_2 = \{(2,2), (3,3)\} \Big)$$

$$\Delta\big((q_i, q_j), (\ell_1, \ell_2)\big) = \big(\delta(q_i, \ell_1), \delta(q_j, \ell_2)\big)$$

$$\mathrm{Mark}_1 = \mathbf{Q} \bigcap \Big(\bigcup_{q \in \mathrm{Mark}} q \times Q\Big) \qquad \mathrm{Mark}_2 = \mathbf{Q} \bigcap \Big(\bigcup_{q \in \mathrm{Mark}} Q \times q\Big)$$

Getting the Multivariate generating Function



$U = aa + b, \qquad V = b^\star aab^\star$

Chomsky-Schützenberger again

$$\begin{aligned}
L_{00} &= \pi_a z L_{11} + \pi_b z u L_{30} + 1\\
L_{11} &= \pi_a z uv L_{22} + \pi_b z u L_{30} + 1\\
L_{30} &= \pi_a z L_{11} + \pi_b z u L_{30} + 1\\
L_{22} &= \pi_a z uv L_{22} + \pi_b z uv L_{33} + 1\\
L_{33} &= \pi_a z L_{11} + \pi_b z uv L_{33} + 1
\end{aligned}$$

$\mathbf{P}(a) = \pi_a \quad \mathbf{P}(b) = \pi_b$

Assume $\pi_a = \pi_b = \frac{1}{2}$ $\begin{cases} U_n \text{ number of occurrences of } U \text{ in texts of length } n\\ V_n \text{ number of occurrences of } V \text{ in texts of length } n \end{cases}$

$$F(z, u, v) = \sum_{n \geq 0} z^n \sum_{\substack{u \geq 0\\ v \geq 0}} \mathbf{P}(U_n = r, V_n = s) u^r v^s$$
$$= \frac{8 + 4z - 8uvz - 2uv(1 - uv)z^2}{8 - 4uz - 8uvz - 2u(1 - 2uv - uv^2)z^2 - u^2v^2(1 + u)z^3}$$

Covariance of U_n and V_n

$$\sum_{n\geq 0} \mathbf{E}(U_n)z^n = \left.\frac{\partial F(z,u,1)}{\partial u}\right|_{u=1}, \quad \sum_{n\geq 0} \mathbf{E}(U_n^2)z^n = \left.\frac{\partial}{\partial u} u \frac{\partial F(z,u,1)}{\partial u}\right|_{u=1}$$
$$\sum_{n\geq 0} \mathbf{E}(V_n)z^n = \left.\frac{\partial F(z,1,v)}{\partial v}\right|_{v=1}, \quad \sum_{n\geq 0} \mathbf{E}(V_n^2)z^n = \left.\frac{\partial}{\partial v} v \frac{\partial F(z,1,v)}{\partial v}\right|_{v=1}$$

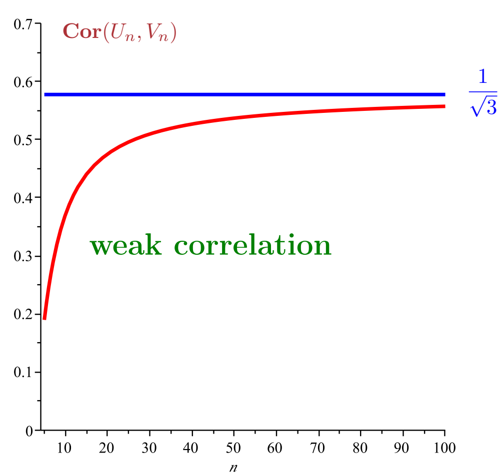
$$\sum_{n\geq 0} \mathbf{E}(U_n V_n)z^n = \left.\frac{\partial}{\partial u}\frac{\partial}{\partial v} F(z,u,v)\right|_{\substack{u=1\\v=1}} = \frac{z^2}{8} \times \frac{8+8z-14z^2+5z^3-z^4}{(1-z)^3(2-z)^2}$$

$$\mathbf{E}(U_n V_n) = \frac{3}{8}n^2 - \frac{3n+1}{4} + 2^{-n}n \qquad \begin{cases} \mathbf{E}(U_n) = \dfrac{3n-1}{4} \\ \mathbf{E}(V_n) = \dfrac{n-2}{2} + 2^{-n} \end{cases}$$

$$\mathbf{Cov}(U_n, V_n) = \mathbf{E}(U_n V_n) - \mathbf{E}(U_n)\mathbf{E}(V_n) = \frac{n-4}{8} + 2^{-n}\frac{n+1}{4}$$

Correlation of $U_n = aa + b$ and $V_n = b^\star aab^\star$

$$\mathbf{Cor}(U_n, V_n) = \frac{\mathbf{Cov}(U_n, V_n)}{\sigma_{U_n}\sigma_{V_n}} = \frac{\mathbf{E}(U_n V_n) - \mathbf{E}(U_n)\mathbf{E}(V_n)}{\sigma_{U_n}\sigma_{V_n}}$$
$$= \frac{n-4+2^{-(n-1)}(n+1)}{\sqrt{(n+1)(3n-6-2^{-n}(4n-12)-4^{-(n-1)})}}$$



Remark:

For $n = 100$, we would get by **exhaustive enumeration** $2^{100} \approx 1.27 \times 10^{30}$ texts

7.10 More on Marked-Automata

1. The **Marked-States** have the **same properties** as the **Accepting-States**, with respect to

   – **determinization** of NFAs

   – **minimization** of DFAs

2. It is possible to make the **product of any finite number of automata**; this is not limited to the product of two automata. The automata need only be complete.

7.11 Reg-Exp to NFA by Glushkov (1961) or Berry-Sethi (1986) algorithm

$R = (a + b)^* aba$

1. **Index** the occurrences of letters $R' = (a_1 + b_1)^* a_2 b_2 a_3$

2. Use the **constructors** $\text{first}, \text{last}, \text{follow}$, while considering that you are "looking" from left to right to the regular expression for first and follow and backwards for last

   – first: the set of indexed letters that you can access by reading "only" one indexed letter from the left; you can bypass "stared" expressions $H^\star$ for any sub-regular-expression within the indexed original regular expression.
   ▶ $\text{first}(R') = \{a_1, b_1, a_2\}$

   – last: symmetric of first while reading backward.
   ▶ $\text{last}(R') = \{a_3\}$

   – $\text{follow}(R', \ell)$: you put yourself at the position ℓ, where ℓ is a marked letter of R', and you compute the set of indexed letters you can get by a single "read"; the conditions are identical to those of first.
   ▶ $\text{follow}(R', b_1) = \{a_1, b_1, a_2\}$

3. **Build the Automaton**

   – **indexed letters** $\rightarrow$ states

   – **suppression of the indices** $\rightarrow$ transitions
   ▶ $\delta(b_1, a) = \{a_1, a_2\}, \qquad \delta(b_1, b) = \{b_1\}, \quad etc.$

Glushkov and Berry-Sethy algorithm.

Recursive definition of $\text{first}, \text{last}, \text{follow}$ and nullable

$\text{nullable}(R) = \texttt{true} \quad \texttt{if } \epsilon \in \text{language of } R$

$$\text{first}(R_1 R_2) = \begin{cases} \text{first}(R_1) \cup \text{first}(R_2) & \texttt{if} \quad \text{nullable}(R_1), \\ \text{first}(R_1) & \text{otherwise} \end{cases}$$

$\text{follow}(R_1 R_2, x) =$

$$\begin{cases} \text{follow}(R_2, x) & \texttt{if} \quad x \in R_2, \\ \text{follow}(R_1, x) \cup \text{first}(R_2) & \texttt{if} \quad x \in \text{last}(R_1) \\ \text{follow}(R_1, x) & \texttt{otherwise} \end{cases}$$

$\text{follow}(R^*, x) =$

$$\begin{cases} \text{follow}(R, x) \cup \text{first}(R) & \texttt{if} \quad x \in \text{last}(R), \\ \text{follow}(R, x) & \texttt{otherwise} \end{cases}$$

Technical Condition $\Rightarrow$ quadratic complexity

8 Fast exact extraction of Taylor coefficients

$$F(z,u) = \frac{P(z,u)}{Q(z,u)} \Longrightarrow \begin{cases} E(z) = \sum_{n\geq 0} \mathbf{E}(X_n) z^n = \left.\dfrac{\partial F(z,u)}{\partial u}\right|_{u=1} = \dfrac{U(z)}{V(z)}, \\ M_2(z) = \sum_{n\geq 0} \mathbf{E}(X_n^2) z^n = \left.\dfrac{\partial}{\partial u} u \dfrac{\partial F(z,u)}{\partial u}\right|_{u=1} = \dfrac{H(z)}{K(z)} \end{cases},$$

where $U(z), V(z), H(z)$ and $K(z)$ are polynomials.
We are looking for $\mathbf{E}(X_n)$ and $\mathbf{E}(X_n^2)$ that are Taylor coefficients of order n of a rational function.
$\mathbf{E}(X_n) = [z^n]E(z), \quad \mathbf{E}(X_n^2) = [z^n]M_2(z)$

Aim: we want to perform a **fast extraction** of the **nth Taylor coefficient** of a **rational function**

Method: (a) find a recurrence for the coefficients.

$$E(z) = \frac{\sum_{0\leq i\leq j} u_i z^i}{\sum_{0\leq i\leq k} v_i z^i} = \sum_{n\geq 0} e_n z^n \Longrightarrow \sum_{0\leq i\leq k} v_i z^i \sum_{n\geq 0} e_n z^n = \sum_{0\leq i\leq j} u_i z^i$$
$$\Longrightarrow e_m v_0 + e_{m-1} v_1 + \cdots + e_{m-k} v_k = 0 \qquad (m > j)$$

(b) Build a **matrix recurrence** of **order** 1.

$$E_m^t = \mathbb{A}^{m-k} E_k^t \begin{cases} E_m = (e_m, e_{m-1}, \ldots, e_{m-k}) \\ E_{m+1}^t = \mathbb{A} \times E_m^t \end{cases} \quad \text{with } \mathbb{A} = \begin{pmatrix} -v_1/v_0 & -v_2/v_0 & \ldots & -v_k/v_0 \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & & & \end{pmatrix} \text{ square matrix}$$

(c) Use an algorithm known as **binary exponentiation** to compute $\mathbb{A}^{m-k}$: $\mathbb{A}^4 = (\mathbb{A}^2)^2$, $\mathbb{A}^8 = (\mathbb{A}^4)^2, \ldots$

Example - $R = aba$, $\mathbf{P}(a) = \mathbf{P}(b) = 0.5$ **-** $\mathbf{E}(400000)$**?**

$$\sum_{n\geq 0} \mathbf{E}(X_n) z^n = \frac{z^3/2}{4 - 8z + 5z^2 - 2z^3 + z^4}$$

$$e_n = 2e_{n-1} - \frac{5}{4}e_{n-2} + \frac{1}{2}e_{n-3} - \frac{1}{4}e_{n-4}$$

$$E^t_{400000} = \begin{pmatrix} 2 & -5/4 & 1/2 & -1/4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}^{399997} \begin{pmatrix} 1/8 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \begin{array}{l} 399997 \\ = 1100001101001111101 \\ \quad \text{(base 2)} \quad \text{(19 bits)} \end{array}$$

19 matrix products, 11 matrix by vector products (number of bits equal to 1)

$$E(X_{400000}) = \frac{399998}{8} \text{ (0.001sec)}, \quad E(X_{4000000}) = \frac{3999998}{8} \text{ (0.002sec)}$$

Complexity $O(\log n)$ **number of operations** for the computation of the **nth coefficient**

$$\log(4000000)/\log(400000) \approx 1.179 \qquad \text{beware of } \textbf{bit complexity}$$

Automatic computations - Lib. `regexpcount` **(N.-Salvy)**

```
> with(regexpcount):
> GRAM:={a=Atom,b=Atom,R=Prod(a,Sequence(b),a)};
```

$$GRAM := \{R = Prod(a, Sequence(b), a), a = Atom, b = Atom\}$$

```
> autoR:=regexptomatchesgram(GRAM,S,[[R,u,'overlap']]);
```

$$autoR := \{S = Union(\mathrm{E}, Prod(a, w3), Prod(b, S)), a = Atom, b = Atom, u = \mathrm{E}, w2 = Union(\mathrm{E}, Prod(a, u, w2), Prod(b, w3)), w3 = Union(\mathrm{E}, Prod(a, u, w2), Prod(b, w3))\}$$

```
> EQS:={seq(eval(subs(Prod=`*`,Union=`+`,Epsilon=1,Atom=var,i)),i=
  autoR)};
```

$$EQS := \{S = 1 + a\,w3 + b\,S, a = var, b = var, u = 1, w2 = 1 + a\,u\,w2 + b\,w3, w3 = 1 + a\,u\,w2 + b\,w3\}$$

```
> for i in {u,p} do EQS:=EQS minus {i=1} end do:for i in {a,b} do
  EQS:=EQS minus {i=var} end do:EQS;
```

$$\{S = 1 + a\,w3 + b\,S, w2 = 1 + a\,u\,w2 + b\,w3, w3 = 1 + a\,u\,w2 + b\,w3\}$$

```
> VAR:={seq(op(1,i),i=EQS)};
```

$$VAR := \{S, w2, w3\}$$

```
> SOLabu:=subs(solve(EQS,VAR),S);
```

$$SOLabu := -\frac{-a - 1 + b + a\,u}{a\,u\,b + 1 - 2\,b + b^2 - a\,u}$$

```
> SOLzu:=subs(a=z/2,b=z/2,SOLabu);
```

$$SOLzu := -\frac{-1 + \frac{1}{2}\,z\,u}{\frac{1}{4}\,z^2\,u + 1 - z + \frac{1}{4}\,z^2 - \frac{1}{2}\,z\,u}$$

```
> E(z):=subs(u=1,diff(SOLzu,u));
```

$$E(z) := -\frac{1}{2}\,\frac{z}{\frac{1}{2}\,z^2 + 1 - \frac{3}{2}\,z} + \frac{\left(-1 + \frac{1}{2}\,z\right)\left(\frac{1}{4}\,z^2 - \frac{1}{2}\,z\right)}{\left(\frac{1}{2}\,z^2 + 1 - \frac{3}{2}\,z\right)^2}$$

Automatic computations - Library `gfun` (Salvy-Zimmerman)

```
> E(z):=subs(u=1,diff(SOLzu,u));
```

$$E(z) := -\frac{1}{2}\,\frac{z}{\frac{1}{2}z^2+1-\frac{3}{2}z} + \frac{\left(-1+\frac{1}{2}z\right)\left(\frac{1}{4}z^2-\frac{1}{2}z\right)}{\left(\frac{1}{2}z^2+1-\frac{3}{2}z\right)^2}$$

```
> with(gfun):
> rec:=diffeqtorec(E(z)-y(z),y(z),u(n));
```

$$rec := \left\{-2\,u(n)+4\,u(1+n)-n,\ u(0)=0,\ u(1)=0,\ u(2)=\frac{1}{4}\right\}$$

```
> PROC:=rectoproc(rec,u(n)):
> time();
                              28.099

> t:=time():evalf(PROC(4000));time()-t;
                           1999.000000
                              0.017
> t:=time():evalf(PROC(40000));time()-t;
                           19999.00000
                              0.050
> evalf(log(40000)/log(4000));
                           1.277618919
> BITS_NUMER_4000:=evalf(log(numer(PROC(4000)))/log(10));
                    BITS_NUMER_4000 := 1207.420796
> BITS_NUMER_40000:=evalf(log(numer(PROC(40000)))/log(10));
                    BITS_NUMER_40000 := 12045.50083
```

Automatic computations - Lib. `gfun` (Salvy-Zimmerman)

```
> rec:=diffeqtorec(E(z)-y(z),y(z),u(n));
> PROC:=rectoproc(rec,u(n)):
> PP[4000]:=PROC(4000);
PP_4000 :=
    2635089982768455257107675698678946135004854303025452265416514097033272099297556075760842449422107194424874057220495292925385518857078692198218039384257185
    9305080421432939102370161231490121586614313651929557477946317553459544224891886439856056029528327596611892940174861761639624766804651481286292192266934515
    7616564579653326533152116751987169088120903711326493141941024724576377193712598746393325621172177608837740203359481599889055366167377732651823103641252738
    3414595947305921112047909415647520431952859914117438164391026906958299508209849625986824965595364911397136016960799660468743294709168796848751485808242686
    7241375223153740399360830333474596148120365212952448142610805840455064559134667030298312790704687079353135482454175849720688564093174157630595504000635446
    6721761910115113310158034393947696250371314763046987742511356760633590078090299313664684231393397949869929645078593616544758290908108701063787860954672055
    0588031354659109364559093701888447103900767515491188517270411454003385284483931369862245370562401767813732876381923948403361697372455955466938975834412295
    69969818738803692554948517286086011692293139371521278025991623604578896409740965332055287304082286646838114987802625 /
    1318204093430943100103889794236591363184019161093272769092803450241756928112834455107975212317212203314094075648071682303844681769424058128173106245251218
    3854467444438688895632897064277199393003658655292424951448883218338941583237562000928492260894611103857875407791326544091858312558605043164728460363649082
    5000782681167246890021068910448808948534719215270882011976500612594485839776187466930127874523350479658699451405443521705380373270324028340081592616934836
    9947271609457689400724316866256888660306583248683060612501764335646973240725287456721773369482423667532334175568183922195469382045607202025388437122682684
    58636194212875139566587445390068014747975813971748114770439248826688866712923795412855584187446066572963049265860017933827257911002088122876736120060347897
    2016889399757435372765399896922309279825570166606797269890623692162876477283791552608646438916157053461695670374484050297527909408758729896842351653162609
    9838935144902005685122107904896671887894330923207197857563987720862123704094012691276761065814107937875804340361142545474418057715085520493716346090251273
    51260539639221457005977247266676344018155647509515396711351487546062479444592779055555421362722504575706910949376
>
> evalf(log10(numer(PP[4000])));
                              1207.420796
> evalf(log10(denom(PP[4000])));
                              1204.119983
```

9 An application to biology - Protein Motifs Statistics

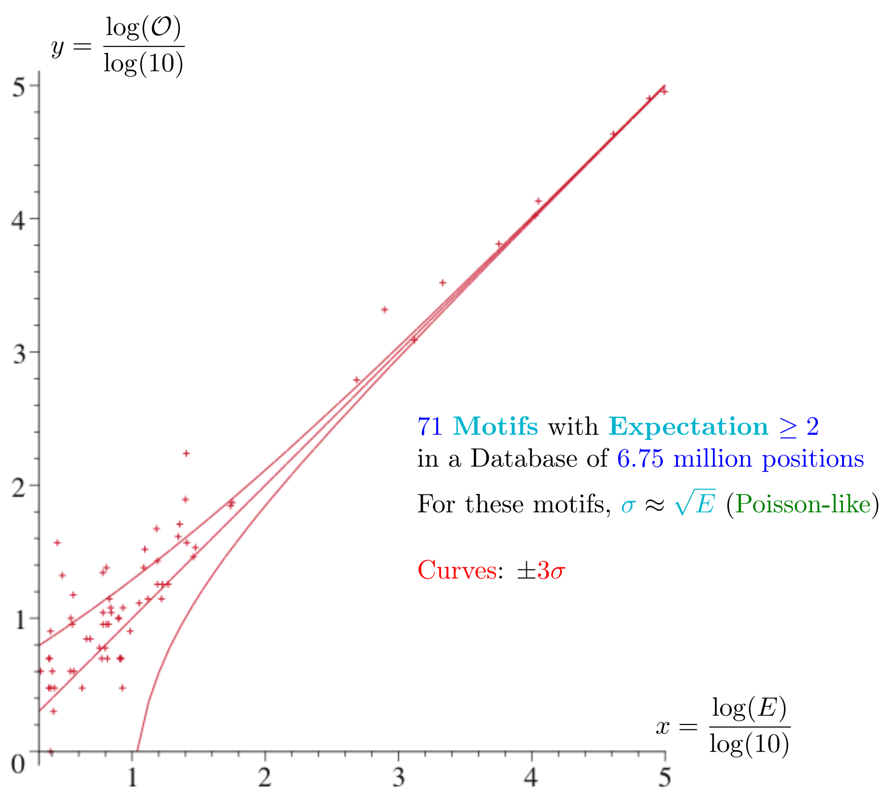
Motif PS00844 (1998): DALA_DALA_LIGASE_2

[LIV]-x(3)-[GA]-x-[GSAIV]-R-[LIVCA]-D-[LIVMF](2)-x(7,9)-[LI]-x-E-[LIVA]-N-[STP]-x-P-[GA]

- $\mathcal{A}$: alphabet of the proteins (20 letters)
- [LIV] $= L + I + V$
- [LIVMF](2) $= (L + I + V + M + F)^2$
- x $= \mathcal{A}$
- x(3) $=$ x^3
- x(7,9) $=$ x^7 $+$ x^8 $+$ x^9

The **automaton** recognizing $\mathcal{A}^\star$.PS00844 and **counting the matches of the motif** in a **random non-uniform Bernoulli** text has **946 states** while the **number of words** of the **finite language** generated by the **motif** is about 2×10^{26}

Comparison of Observed and Predicted Counts



From [Nicodème, Salvy, Flajolet] - Motif Statistics, TCS2002

10 Short Bibliography

– KELLEY, D. *Automata and Formal Languages, an Introduction*. Prentice Hall, 1995. A very clear introduction to the subject.

– KOZEN, D. C. *Automata and Computability*, Springer Verlag, 1997. Probably more complete than Kelley's book, but more difficult to read.

– NICODÈME,P. , SALVY, B., FLAJOLET, F. *Motif Statistics*, TCS 2002

– NICODÈME, P. *Regexpcount, a symbolic package for counting problems on regular expressions and words*, Fundamentae Informaticae, 2003.

– NICAUD, C., PIVOTEAU, C., RAZET, B. *Average Analysis of Glushkov Automata under a BST-Like Model*, FSTTCS'10, 2010

– NUEL, G., DUMAS, J.-G. *Sparse approaches for the exact distribution of patterns in long state sequences generated by a Markov source*, TCS 2012