

**Limit of random walks  
and  
of random trees**

---

**Elements**

---

**Graz 2009**

Jean-François Marckert  
LaBRI, CNRS, Université de Bordeaux 1  
<http://www.labri.fr/perso/marckert>



## 1 Introduction

In computer science, probability theory, algebra, statistical physics, biology, a lot of objects having a discrete nature appear for various reason. In statistical physics, biology and also in probability theory, discrete objects are used to model some real phenomenons (quantum physics, magnetism, polymer, DNA, genealogical tree, random walk, Markov chain, etc.); in computer science and in algebra the objects are by nature discrete: finite groups, monoid, bases of polynomials,... data structures, algorithms, in computer science, morally, everything is discrete!

A good understanding of the combinatorial properties of these families of objects is then a crucial point, often the first question. What means “to understand the combinatorial properties of a family of objects?” Well, it depends. Sometimes, the objects are so complex, that even the number of objects of size  $n$  is unknown:

– there are  $4^n$  path  $S = (S_k, k = 0, \dots, n)$  in the plane with steps North, East, West, South (with  $S_0 = 0$ ). How many such paths are non self intersecting (starting from 0 and such that  $S_k = S_j \implies k = j$ )? This is not known, neither the answer for the loop cases ( $S_n = S_0$ ), called “non intersecting polygons”

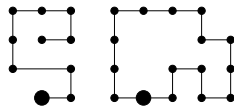


Figure 1: A non self intersecting path with 10 steps; a non intersecting polygon with 16 steps

– How many square matrices  $n \times n$  with entries 0 or 1, such that no two 1 are neighbors? (are neighbors the entries  $a_{k,j}$  and  $a_{k',j'}$  if  $|k' - k| = 1, j = j'$  or  $|j - j'| = 1, k' = k$ ). This is not known.

– Take a honeycomb lattice. The number of directed animals with  $n$  cells is unknown.

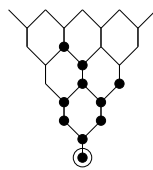


Figure 2: A directed animal on the honeycomb lattice is a subset  $A$  of the lattice such that any point of  $A$  can be reached from the bottom vertex by a raising walk staying inside  $A$ .

In some cases, the objects are more simple, and can be decomposed and counted : we say that we are able to count a family of objects, if the sequence  $(a_n, n \geq 0)$  giving the number of objects of size  $0,1,2,\dots$  is explicitly or implicitly known. Often, only the generating function  $G(z) = \sum_n a_n z^n$  is implicitly known as a solution of a more or less complex system of functional equations.

Examples of counted families of objects include:

– number of compositions of  $n$  (sequence of positive integers summing to  $n$ ); this is  $2^{n-1}$ . Why?

- number of partitions of  $n$  (set of positive integers summing to  $n$ ); the generating function is  $\prod_{k \geq 1} \frac{1}{1-x^k}$ ; why?
- plane partitions with or without imposed symmetries with  $n$  boxes,
- the number of paths  $S = (S_0, \dots, S_n)$  where the increments  $S_k - S_{k-1}$  are taken in a finite set  $I$ , or restricted to satisfy various assumptions, as to end at 0, to be positive all along, etc.,
- number of rooted plane trees with  $n$  nodes (and where the degree of the nodes are restricted to belong to a fixed set, as for example  $\{0, 2, 3\}$ ),
- number of rooted quadrangulations with  $n$  faces (or triangulations), or number of planar graphs with  $n$  points, etc.

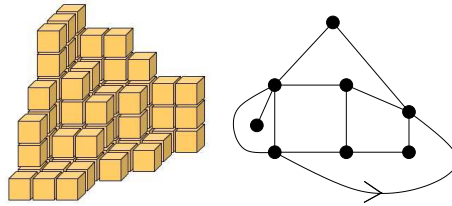


Figure 3: A plane partition and a rooted quadrangulation with 5 (finite) faces.

One can find on the Sloane's web site, more than 160 000 sequences appearing as counting sequences in combinatorics : <http://www.research.att.com/~njas/sequences/> (Can you imagine why this site is made for ?)

Existing techniques to count combinatorial objects are more or less involved, but often tricky : often they rely on the decomposition of the objects; from there some equations involving the searched generating function have to be solved, or some ugly determinants have to be computed; sometimes a bijection exists with a simpler family of objects...

Good overviews are:

- the two books of Richard Stanley, Enumerative Combinatorics, vol. 1 and 2.

<http://www-math.mit.edu/~rstan>

- Analytic combinatorics, by Flajolet and Sedgewick, available for free at

<http://algo.inria.fr/flajolet>

- see also the page of the Austrian research network devoted to Analytic Combinatorics and Probabilistic Number Theory,

<http://dmg.tuwien.ac.at/nfn>

where the different specialties of Austrian research teams on the topics are given.

Of course, studying combinatorial objects doesn't mean only counting. Knowing that there are  $\binom{2n}{n}/(n+1)$  binary trees with  $n$  internal nodes, or that they are  $2^{n-1}$  compositions of  $n$  does

not help a lot to know what is the standard shape of a binary trees with  $10^9$  nodes or what is the standard size of a part in a compositions of  $10^8$ . When one wants to describe a family of combinatorial objects, one of the first questions arising is then to try to give some elements on an additional parameter of the structure. The parameters of interests are various, depends on the applications, but in any case, they generally give the first quantitative information on the behavior of the discrete structures.

For example,

- the standard height of a planar binary tree with  $n$  nodes is about  $\sqrt{n}$ . In other words, if the set of trees with  $n$  nodes is endowed with the uniform distribution, the mean height of a tree is asymptotically  $\sqrt{n}$  (up to a constant factor). [The maximal possible height of a tree with  $n$  internal node is  $n$ : this is trivial and not interesting; the minimal height is around  $\log_2 n$ : this is trivial, and not interesting].
- in a uniform quadrangulation with  $n$  faces, the (mean of the) maximum distance between two vertices is  $n^{1/4}$ ,
- in a binary search tree with  $n$  nodes, the height of the tree is  $4.31 \log n + o(\log n)$  asymptotically (when the data successively inserted  $x_1, \dots, x_n$  are i.i.d. random variables, uniform on  $[0, 1]$  (see Section 2.2)),
- in the cycle-representation of a uniform random permutation of  $\{1, \dots, n\}$ , the cycle containing 1 as a length uniform on  $\{1, \dots, n\}$ .

**Note 1 :** 1) The same set of objects can be endowed with several “natural” distributions. On the above examples, the set of binary trees is first endowed with the uniform distribution, then endowed with the distribution induced by successive insertion of data (the binary search tree construction). These distributions are radically different. Each result concerning the mean or the distribution of a parameter of a family of objects is of course dependent on the chosen distribution on this family.

2) In computer science, data structures are often simple combinatorial objects (trees, arrays). When one wants to analyze the cost of construction or of utilization of a given data structures, often this cost is a simple parameter of the structure: for example, the cost of the construction of a binary search tree relies on the number of comparisons done between the elements; the total number of comparisons done when constructing the tree is equal to the total path length of the tree, the maximum number of comparisons needed to place an element in the tree is given by the height of the tree, etc.

Hence, in the book of Knuth (the art of computer programming), or in those of Flajolet & Sedgewick, analysis of algorithms considerations appear “melted” with combinatorial ones.

During, these last decades appear that some families of combinatorial structures show some remarkable asymptotic behaviors. Some of the parameters of interest present some limiting behavior in distribution; for example, denoting by  $H_n$  the height of a binary tree with  $n$  internal nodes taken uniformly in the set of binary trees with  $n$  nodes, then  $H_n/\sqrt{n}$  converges in distribution. In a binary search tree, the depth of insertion  $D_n$  in a tree of size  $n$ , converges also:  $(D_n - 2/\log n)/\sqrt{\log n}$  converges in distribution to a centered normal variable. Much more than that appears; sometimes the family of objects can be sampled (generated with the right distribution on a computer) and drawn: in some cases, the whole picture of the object seems to present a limiting behavior, see the

examples on Figures 4, 5 and 6.

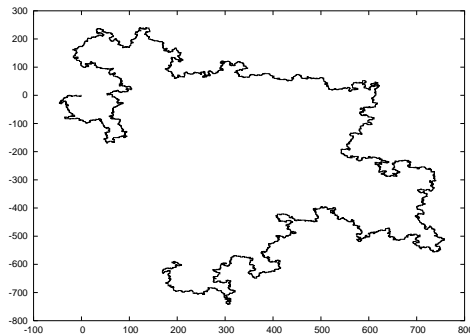


Figure 4: loop erased random walk with 10 000 steps: make a random walk in the plane, and erase the loops as soon as they appear. Stop when the number of (remaining) steps is 10000.

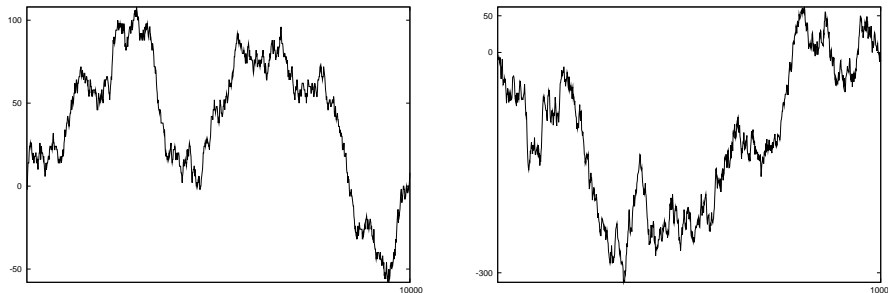


Figure 5: Two random walks with 10000 steps. The first one has increment distribution  $\frac{1}{2}(\delta_1 + \delta_{-1})$ , the second one with increments distribution  $\frac{1}{4}\delta_{-4} + \frac{1}{5}\delta_2 + \frac{1}{5}\delta_3 + \frac{7}{20}\delta_0$ . So different... and so similar

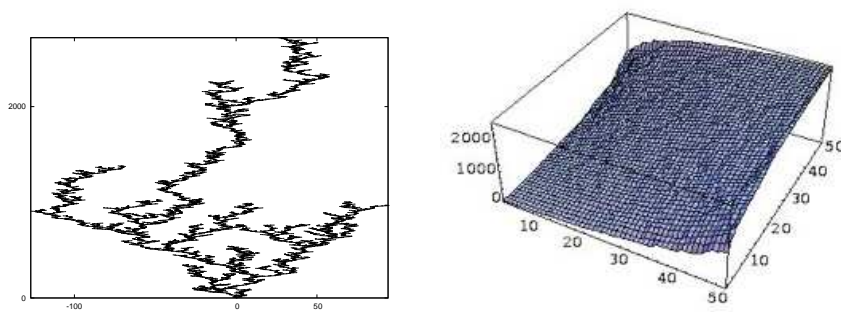


Figure 6: On the first picture, a uniform directed animals on a triangular lattice with 10 000 cells. On the second one, a uniform filling of a square array of size  $n \times n$  by the integers from 1 to  $n^2$ , increasing along rows and columns (picture, taken on Dan Romik's page <http://www.math.ucdavis.edu/~romik>)

## Main questions

Hence, the “limiting picture” can be random or deterministic. The difficulty when one tries to translate the observed phenomenons into theorems or conjectures is immediate: a limiting behavior appears on the picture, yes, but, something has been forgotten in the procedure... the notion of convergence has not been defined! and the limiting object is often not of the same type as the discrete one... it is no more discrete, but present a “continuous” structure. Often, when sampling and drawing “the picture” of the combinatorial object on a picture, a rescaling has been done: morally, what appears at scale  $\sqrt{n}$  have been represented at scale 1 on Figure 5. It has been done more or less consciously by the programmer, but it has been done (since most drawing softwares make an automatic rescaling of the picture to fit with the screen of the computer).

Hence the main questions that appear are the following:

1. what means that a sequence of rescaled combinatorial structures converges?
2. what means that a sequence of rescaled **random** combinatorial structure converges? (in distribution? or in probability?)

Assume that a theorem of convergence has been written...

3. would we be able to deduce something from this global convergence?
4. what can not be deduced from this convergence?

But over all,

5. how can we prove such convergence results ?

## Quick answers

1. Intrinsically, nothing. “To converge” is a notion “topology dependent”. In other words, a topology has to be defined first... it may happen that no pre-existing topology is available in the literature; particularly, this may happen if you want to prove a first result on your preferred structure (all combinatorial structures are different); you do have to invent a topology on a convenient set, containing the discrete rescaled objects and the plausible limit... that you have to create also. Sometimes, several choices of topology are available.

2. Roughly speaking, one can say that if the work in (1) has been done cleanly, the notion of convergence in distribution (or in probability) is well defined at once; there is a notion of weak convergence associated automatically to a given notion of convergence.

3. Again, this is topology dependent. More the topology is thin, more the deductions are numerous and strong. If the limiting object is not trivial, infinitely many things can be deduced in general (for example, the scaling has been chosen in such a way that the limiting object is not infinite or reduced to a single point. These byproducts may concern the whole structure, if the topology is well built. Generically, any parameter which depends continuously of the objects will have a limiting behavior implied by the convergence of the structure.

4. Again, this is topology dependent. In general, infinite many things can not be deduced. For example, if you rescale your structure by  $\sqrt{n}$ , every phenomenon which does not appear at this

scale can not be viewed on the limit, and then can not be studied via this procedure.

5. this is the main question, and of course, the difficult one. Several different problems in nature have to be solved altogether. When a topology has been defined, a first work is to establish some criterion of convergence in distribution with respect to this topology. Then of course, the special case studied must be proved to satisfy the criterion built.<sup>1</sup>

In this course, we will modestly illustrate these questions and constructions for the convergence of a sequence of paths, and for the convergence of a sequence of trees. Some applications to analysis of algorithms will naturally appear.

### Some references

We here give some references that have been used to make the present course; some other references are given for complement, to the interested reader.

Concerning the asymptotic of paths, the main ideas comes from Donsker in the fifties. For this course, the main sources in probability theory are the books of Billingsley (Convergence of Probability Measures, and Probability and Measure), the book of Kallenberg (Foundations of Modern Probability, not recommended to beginners); for beginners, the book of Breiman (Probability) does not contain all the stuff needed for the course, but constitutes a very nice introduction to probability theory. The properties of the Brownian Motion (just evoked at the end of the notes) can be studied in the book of Revuz & Yor (Continuous Martingales and Brownian Motion), or Karatzas & Shreve, (Brownian Motion and Stochastic Calculus).

The convergence of rescaled Galton-Watson tree to a limiting object has been proven/imagined by David Aldous in the nineties

<http://www.stat.berkeley.edu/~aldous/Research/index.html>

in a series of large papers (the continuum random tree 1, 2, and 3).

From this date, the continuum random tree appears to be a central structure, appearing as limit in a lot of applications. One will find on the web page of Jean-François Le Gall a lot of stuff concerning the convergence of rescaled trees (among other its work with T. Dusquesne concerning a generalization of Aldous result to families of Galton-Watson trees, where the offspring distribution is more general than in the present course, letting other limits than the continuum random tree of Aldous, appear); on the page of JFL one will find also some courses on the convergence of random trees

---

<sup>1</sup>Wendelin Werner got the Field medals in 2006, particularly for his works (together with Oded Schramm and Greg Lawler) concerning the limiting behavior for various model of random curbs in the plane : loop erased random walks, contour of percolation clusters, random Peano curbs,... What is remarkable in these impressive works is that limiting behaviors are obtained even if the number of such curbs are not known (even up to an exponential order, for some of them). Hence, this example shows that even if often counting results are the first elements and an important tool to obtain asymptotic results, it exists some cases where they stay unknown when more asymptotic qualitative results are known.



<http://www.dma.ens.fr/~legall/Cornell.pdf>

Properties of branching processes (as Galton-Watson processes) can be studied in the book of Athreya and Ney (Branching processes) and Harris, (The theory of branching processes).

The proof of the convergence of the rescaled random trees we propose here is taken from a paper written in collaboration with M. Mokkadem (The depth first processes of Galton-Watson trees converge to the same Brownian excursion, in 2003).

Before starting with the core of the problem, we recall now some simple facts of probability theory in  $\mathbb{R}$  (for distribution on  $\mathbb{R}$ ). We think that the reader must understand this stuff before being ready to understand the construction of convergence for more complex structures.

## 2 Elements on probability theory in $\mathbb{R}$

### 2.1 Distribution on $\mathbb{R}$ : characterizations and convergences

We write in this part, some elements valid in the real case but also more generally, as it will soon appear.

Let  $\Omega$  be a set. A  $\sigma$ -field  $\mathcal{A}$  on  $\Omega$ , is a part of the power set  $\mathcal{P}(\Omega)$ , stable by countable union, countable intersection, and by complementation. The elements of  $\mathcal{A}$  are the measurable sets, and often called in probability theory *the events*. A probability measure  $\mathbb{P}$  on  $(\Omega, \mathcal{A})$  is a map from  $\mathcal{A}$  into  $[0, 1]$ , which is  $\sigma$ -additive, and such that  $\mathbb{P}(\Omega) = 1$  (by  $\sigma$ -additive, we mean: if  $(A_i, i \geq 1)$  is a sequence of non pairwise intersecting elements of  $\mathcal{A}$ , then  $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$ ).

The three tuple  $(\Omega, \mathcal{A}, \mathbb{P})$  is called a probability space.

**Note 2 :** When  $\Omega$  is a finite or countable set, usually  $\mathcal{A}$  is taken equals to the whole set  $\mathcal{P}(\Omega)$ ; there is a good reason for that... The measure  $\mathbb{P}$  gives a measure to the events; then in general we want the singletons  $\{\omega\}$  to have a probability, that is to belong to the  $\sigma$ -field  $\mathcal{A}$ . In a countable set, if the singletons are in the  $\sigma$ -field, then every subsets are also in the  $\sigma$ -field (since this latter is close under countable union). Hence, in the discrete case, considerations about  $\sigma$ -fields are often omitted in the first courses of probability. In the case of distribution on  $\mathbb{R}$  this can not longer be omitted, since no satisfying construction of measures can avoid this point on  $\mathbb{R}$ .

**Example 1 :** • Famous distributions on discrete sets:

– on finite sets:

. the uniform distribution

. Bernoulli parameter  $p$ . This is the distribution  $p\delta_1 + (1-p)\delta_0$  meaning,  $\mathbb{P}(\{1\}) = p, \mathbb{P}(\{0\}) = 1-p$

. Binomial distribution with parameter  $(n, p)$ ; this is the distribution which assigns the weight  $\binom{n}{k} p^k (1-p)^{n-k}$  to  $k \in \llbracket 0, n \rrbracket$ .

– on  $\mathbb{N}$  :

. the geometrical distribution:  $\mathbb{P}(\{k\}) = p(1-p)^k$  for  $k \geq 1$ ,

. Poisson distribution with parameter  $\lambda$ ,  $\mathbb{P}(\{k\}) = \lambda^k \exp(-\lambda)/k!$ , for  $k \geq 0$

• on  $\mathbb{R}$  (famous distributions having a density with respect to the Lebesgue measure):

. Uniform distribution,  $f(x) = \mathbb{1}_{[0,1]}(x)$ ,

. Exponential distribution with parameter  $a$  :  $f(x) = ae^{-ax}, x \geq 0$ ,

. Normal distribution  $\mathcal{N}(m, \sigma^2)$ , with mean  $m$  and variance  $\sigma^2$ ,

$$f(x) = \frac{\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}.$$

When a density exists, the set of measurable sets can not be the entire power set  $\mathcal{P}(\mathbb{R})$  for the reason that (a measure having the same properties as) the Lebesgue measure can not be defined on all  $\mathcal{P}(\mathbb{R})$ . Hence, even if  $\mathbb{P}$  is the uniform distribution on  $[0, 1]$ ,  $\mathbb{P}(B)$  is not (can not be) defined for any subset  $B$  of  $[0, 1]$ ; even if this fact seems at the first glance not important, it implies in some sense all the “complications” coming from the  $\sigma$ -field considerations.

The  $\sigma$ -field used on  $\mathbb{R}$  is the  $\sigma$ -field generated by the open sets (the minimal  $\sigma$  field containing the open sets); this  $\sigma$ -field is called the Borelian  $\sigma$ -field, and denoted by  $\mathcal{B}(\mathbb{R})$ ; its elements are called the Borelians, or Borelian sets. If  $\mathbb{P}$  has a density  $f$ , then for any Borelian  $A \in \mathcal{B}(\mathbb{R})$ ,  $\mathbb{P}(A) = \int_A f(x)dx$ , and in any case, we write  $\mathbb{P}(A) = \int_A d\mathbb{P}(x)$  meaning that we integrate with respect to the measure  $\mathbb{P}$ .

We recall that when  $\mathbb{P}_X$  is discrete then for any measurable function  $g$

$$\mathbb{E}(g(X)) = \sum_{x \in A} \mathbb{P}(X = x)g(x)$$

when  $\mathbb{P}_X$  has a density  $f$  then

$$\mathbb{E}(g(X)) = \int f(x)g(x)dx$$

in any case, we write

$$\mathbb{E}(g(X)) = \int g(x)d\mathbb{P}_X(x).$$

To end, notice that  $\mathcal{B}(\mathbb{R})$  contains all the singletons on  $\mathbb{R}$ , and then, it contains any  $\sigma$ -field build on any countable set (for example  $\mathbb{Z}$ ,  $\mathbb{N}$ ,  $\mathbb{Q}$ ); for this reason, in general when a variable is considered as a real random variable, the considered  $\sigma$ -field will always be  $\mathcal{B}(\mathbb{R})$ .

Let  $(E, \mathcal{E})$  be a set  $E$  equipped with a  $\sigma$ -field  $\mathcal{E}$ . A **random variable**  $X$  on  $(\Omega, \mathcal{A})$  taking its values in  $(E, \mathcal{E})$  is a measurable map from  $(\Omega, \mathcal{A})$  into  $(E, \mathcal{E})$  (measurable means that  $X^{-1}(C) \in \mathcal{A}$  for any  $C \in \mathcal{E}$ ).

If  $(\Omega, \mathcal{A}, \mathbb{P})$  is a probability space, the distribution of  $X$  is a probability measure  $\mathbb{P}_X$  on  $(E, \mathcal{B}(E))$  defined by

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)).$$

Hence, the distribution of a random variable  $X$  is a measure  $\mathbb{P}_X$ . This distribution is called the image distribution of  $X$  (and also, the pushed-forward distribution of  $X$  on  $(E, \mathcal{B}(E))$ ).

---

**Example 2 :** a) Assume that  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$ ,  $\mathbb{P}$  is the uniform distribution. Then  $X : \Omega \mapsto \mathbb{R}$  defined by  $X(i) = (i - 2)^2$  is a random variable in  $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$  with distribution  $\mathbb{P}_X$  as follows:  $\mathbb{P}_X(\{i\}) = 1/6$  for  $i \in \{0, 4, 9, 16\}$  and  $\mathbb{P}_X(\{1\}) = 2/6$ .

b) If  $\Omega =$  set of trees with 12 nodes,  $\mathcal{A} = \mathcal{P}(\Omega)$ ,  $\mathbb{P} =$ uniform distribution. Then  $H : (\Omega, \mathcal{P}(\Omega)) \mapsto (\mathbb{N}, \mathcal{P}(\mathbb{N}))$  defined by  $H(t) =$  height of the tree  $t$ , is a random variable.  $\mathbb{P}_H(\{5\}) = \mathbb{P}(H^{-1}(5)) = \mathbb{P}(\{t : H(t) = 5\}) = \#\{t \in \Omega, H(t) = 5\} / \#\Omega$

c) If  $\Omega = \mathbb{R}$ ,  $\mathcal{A} := \mathcal{B}(\mathbb{R})$  the set of Borelian subsets of  $\mathbb{R}$  (defined below),  $\mathbb{P}$  being the normal distribution  $\mathcal{N}(0, 1)$ . Then  $X : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , defined by  $X(x) = \cos(x)$  is a random variable with values in  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . We have  $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\cos(X) \in A) = \int f(x)\mathbb{1}_A(\cos(x))dx$  where  $N$  is normal  $N(0, 1)$ , and  $f$  its density.

---

**Note 3 :** This note can be skipped; its aim is to discuss the fact that often, in probability theory, the probability space is not given, but only distribution and random variables.

Assume that  $X$  is the identity function on  $(\Omega, \mathcal{A})$  (it is measurable as a map from  $(\Omega, \mathcal{A})$  onto itself, and then a random variable). Trivially in this case  $\mathbb{P}_X = \mathbb{P}$  : hence  $\mathbb{P}$  appears to be the initial distribution of a probability space and to be the distribution of a random variable. Hence any distribution on a set is the distribution of a random variable! This artificial case shows that there are no difference in nature between probability distribution of a space, as introduced at the beginning of this section, or the distribution of a random variable. This identity is used everywhere in probability theory, without indication. This is the reason why, often in probability theory, the probability set  $\Omega$  is not given explicitly (specially when we are working on  $\mathbb{R}$ ): often, the story begins with a random variable taking its values in a given set  $E$  having a certain distribution  $\mathbb{P}_X$ ; this does not mean that there didn't be at "the beginning of the time a set  $\Omega$ ", but this set is not useful for the computations, and if needed, it can be taken equal to  $E$ ; in which case taking  $\mathbb{P} = \mathbb{P}_X$  as the "initial probability measure on  $\Omega$ ", the identity has distribution  $\mathbb{P}_X$ .

### 2.1.1 Characterizations of measures on $\mathbb{R}$

Here are some examples of characterization of measures on  $\mathbb{R}$ ; they are written in term of random variables, but does not concern the equality of random variables, but of their distributions.

- 1) if  $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$  for all open ball  $A$ , or open sets  $A$ , or closed sets  $A$ , then  $\mathbb{P}_X = \mathbb{P}_Y$ ,
- 2) if  $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$  for all continuous bounded function  $f$  then  $\mathbb{P}_X = \mathbb{P}_Y$  (same criterion with continuous + bounded support instead),
- 3) the cumulative function  $x \mapsto F_X(x) := \mathbb{P}(X \leq x)$  characterizes the distribution: if  $F_Y(x) = F_X(x)$  for every  $x \in \mathbb{R}$  then the two distributions  $\mathbb{P}_X$  and  $\mathbb{P}_Y$  are equal,
- 4) Fourier transforms:  $t \mapsto \Phi_X(t) := \mathbb{E}(e^{itX})$ , for  $t \in \mathbb{R}$ . If  $\Phi_X(t) = \Phi_Y(t)$  on  $\mathbb{R}$ , then  $\mathbb{P}_X = \mathbb{P}_Y$ ,
- 5) Laplace transforms:  $t \mapsto \Psi_X(t) := \mathbb{E}(e^{tX})$ , for  $t \in \mathbb{R}$  (or a subset of  $\mathbb{R}$ ). If  $\Psi_X(t) = \Psi_Y(t)$  on  $\mathbb{R}$  (or on an interval of  $\mathbb{R}$ ), then  $\mathbb{P}_X = \mathbb{P}_Y$  (Laplace transforms do not exist in all generality),
- 6) Probability generating function  $t \mapsto G_X(t) := \mathbb{E}(t^X)$ , for  $t \in \mathbb{R}$  (or a subset of  $\mathbb{R}$ ); If  $G_X(t) = G_Y(t)$  on  $\mathbb{R}$  (or on an interval of  $\mathbb{R}$ ), then  $\mathbb{P}_X = \mathbb{P}_Y$  (this is used for distribution having their support on  $\mathbb{N}$ ),
- 7) moments: if for any  $k \in \mathbb{N} := \{0, 1, 2, \dots\}$ ,  $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$  then it may happen that  $\mathbb{P}_X \neq \mathbb{P}_Y$ . Under the so-called Carleman's condition, namely if  $\sum (\mathbb{E}(X^{2k}))^{-1/(2k)} = +\infty$ , then there exists only one measure with these moments, and then  $\mathbb{P}_X = \mathbb{P}_Y$ .

**Exercise 1 :** a) Let  $X$  be uniform on  $[0, 1]$ ; using the characterization (3) find the law of  $X^2$ ; do the same work with characterization (2). Compare these two methods for a general r.v.  $Y = f(X)$ .

b) Recall that if  $X$  and  $Y$  are independent then  $f(X)$  is independent of  $g(Y)$  for any measurable function  $f$  and  $g$  (in some measurable spaces). Using characterization (1) find the distribution of  $X + Y$  if  $X$  and  $Y$  are independent, and have the Poisson distribution with respective parameters  $\lambda$  et  $\mu$  (this is equivalent to compute  $\mathbb{P}(X + Y = k)$  for any integer  $k$ ... yes?). Do the same using (4) (discuss what happens if instead of (4), (5) or (6) is used). Discuss the difference in the general case when one wants to compute the law of  $X_1 + \dots + X_n$  using (1) and using (4).

### 2.1.2 Characterization of the convergence of measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

First, we recall some definitions. **Definition :** Let  $X$  be a r.v. and  $(X_n, n \geq 0)$  a sequence of random variables defined on the same probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  taking their values in  $\mathbb{R}$ .

– We say that  $(X_n, n \geq 0)$  converges in probability to  $X$  if :

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(\{\omega \mid |X_n(\omega) - X(\omega)| \geq \varepsilon\}) \xrightarrow[n]{} 0.$$

– We say that  $(X_n, n \geq 0)$  converges almost surely to  $X$  if :

$$\mathbb{P}(\lim_n X_n = X) = \mathbb{P}(\{\omega \mid \lim X_n(\omega)\}) = 1.$$

– We say that  $(X_n, n \geq 0)$  converges to  $X$  in  $L^p$  (for  $p > 0$ ) if :

$$X_n \in L^p, \quad \text{for each } n, \text{ and if } \mathbb{E}(|X_n - X|^p) \xrightarrow[n]{} 0.$$

We put together these three notions, because we see them to be from the same type: they are convergence of functions; all the random variables  $X, X_1, X_2$  are defined on the same space, and they are compared point by point. These convergences are convergence of the "standard type"; more  $n$  is large, and more  $X_n$  is close to the limit  $X$ , which can be random (see Exercise 2.1.2).

Archetype of a.s. convergence is the strong law of large numbers (see also Figure 7) :

**Theorem 1 (Strong law of large number)** *Let  $(Y_n, n \geq 1)$  be a sequence of i.i.d. random variables with mean  $m$ , then*

$$\frac{\sum_{i=1}^n Y_i}{n} \xrightarrow[n]{a.s.} m.$$

The weak law of large number is the same statement with the convergence in probability instead. To prove the weak version is easy if one assumes  $\text{var}(Y_1) < +\infty$ . Try! (Use Bienaymé-Tchebichev inequality). A simple proof of the strong law of large numbers exists if a moment of order  $2 + \varepsilon$  exists: inequality of Markov + Borel Cantelli lemma gives the result.

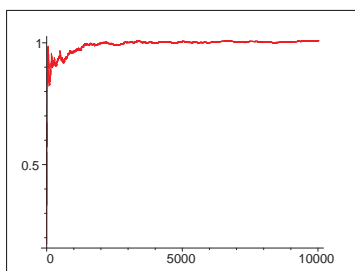


Figure 7: Drawing of a random sequence  $(\frac{\sum_{i=1}^n Y_i}{n}, n \geq 1)$ , with i.i.d. r.v.  $Y_i$  having mean 1.

**Exercise 2 :** Let  $(X_n, n \geq 0)$  be a sequence i.i.d. r.v. Bernoulli  $B(1/2)$ .

a) Show that that the sequence  $(Y_n, n \geq 1)$  defined by  $Y_n = \sum_{i=1}^n X_i 2^{-i}$  converges a.s.

- b) For each  $n \geq 1$ , let  $Z_n$  be defined by  $Z_n = \sum_{i=1}^n X_i 2^{-(n-i+1)}$ ; prove that  $Z_n$  has the same distribution as  $Y_n$  (that is  $\mathbb{P}_{Y_n} = \mathbb{P}_{Z_n}$ ), for each  $n$ .
- c) Prove that the sequence  $(Z_n)$  does not converge a.s. (neither in probability).

The convergence in distribution is not a convergence of r.v., but a convergence of distributions! Let  $\mu, \mu_1, \mu_2, \dots$  be a sequence of probability measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . The sequence of measures  $\mu_n$  is said to converge weakly to  $\mu$  (we write  $\mu_n \xrightarrow[n]{weak} \mu$ ) if for any **bounded continuous function**  $f$

$$\lim \int_{\mathbb{R}} f(x) d\mu_n(x) = \int_{\mathbb{R}} f(x) d\mu(x).$$

**Definition :** We say that a sequence of random variables  $X_n$  (defined on possibly different probability spaces  $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ ) converges in distribution, if the sequence of measures  $(\mathbb{P}_{X_n}, n \geq 1)$  converges weakly to a limiting probability measure  $\mu$  (which is then the probability distribution of a random variable  $X$ ). In other words, a sequence of random variables  $(X_n)$  converges in distribution to  $X$  (we note  $X_n \xrightarrow[n]{(d)} X$ ) if

$$\mathbb{E}(f(X_n)) \xrightarrow[n]{} \mathbb{E}(f(X))$$

for any  $f : \mathbb{R} \mapsto \mathbb{R}$  bounded, continuous.

(We could have written  $\mathbb{E}_n$  instead of  $\mathbb{E}$  meaning that the expectation is taken with respect to  $\mathbb{P}_n$ ).

See below some more usual characterization of the convergence. What is important to see here, is that the variables need not to be defined on the same  $\Omega$ , and also that  $X_n$  and  $X$  are not compared as random variables:  $X_n$  is not close to  $X$ , but  $\mathbb{E}(f(X_n))$  is close to  $\mathbb{E}(f(X))$ .

Consider the following example: if you take a sequence of i.i.d. random variables  $(X_n)$  where all the  $X_n$  have the same distribution Bernoulli 1/2: they converge in distribution, since for any function  $f$  (continuous or not)  $\mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ ; but  $X_n$  does not converge in probability, since a.s.  $(X_n, n \geq 1)$  has two accumulation points 0 and 1.

### Characterization of the convergence in distribution

If one of the following facts holds, then  $X_n \xrightarrow[n]{(d)} X$ ,

- 1) Convergence of the cumulative distribution function  $F_n(x) = \mathbb{P}(X_n \leq x) \rightarrow F(x) = \mathbb{P}(X \leq x)$ , for all  $x$  where  $F$  is continuous
- 2) Convergence of the Fourier transform  $\Phi_n(t) = \mathbb{E}(e^{itX_n}) \rightarrow \Phi(t) = \mathbb{E}(e^{itX})$ , for all  $t$
- 3) Convergence of the Laplace transform  $t \mapsto \Psi_{X_n}(t) := \mathbb{E}(e^{tX_n}) \rightarrow \Psi_X(t) := \mathbb{E}(e^{tX})$ , for  $t$  in an interval of  $\mathbb{R}$ .
- 4) Convergence of the moments : if for any integer  $k \geq 1$ ,  $\mathbb{E}(X_n^k) \rightarrow \mathbb{E}(X^k)$  and  $X$  satisfying the Carleman's condition.

Compare these criterions with the criterions of characterization of measures in Section 2.1.1. One may state a meta theorem as follows:

“Any criterion characterizing the probability measures on  $\mathbb{R}$  characterizes also the convergence of measures, under the additional condition that the limit is a probability measure”.

**Exercise 3 :** a) What we mean by the meta theorem is: the point-wise convergence of  $F_n$  to some function  $F$  is not sufficient ( $F$  must be a cumulative function), the convergence of the Fourier transform is not sufficient, the limit must be the Fourier transform of a probability measure, etc. Find an example of sequence of r.v.  $(X_n)$ , such that  $(F_n)$  converges point-wise and such that the sequence  $(X_n)$  does not converge in distribution.  
 b) Find an example of sequence of r.v.  $(X_n)$  such that  $X_n \xrightarrow{d} X$  but  $\mathbb{E}(X_n) \not\rightarrow \mathbb{E}(X)$ .

See also below, the classical Portmanteau theorem :

**Theorem 2 (Portmanteau)** Let  $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$  be a sequence of measures on the same Polish space  $(S, \mathcal{S})$ . The five following assertions are equivalent :

- (i)  $\mathbb{P}_n \xrightarrow[n]{weak} \mathbb{P}$ .
- (ii)  $\int f d\mathbb{P}_n \rightarrow \int f d\mathbb{P}$  for any  $f$  bounded, uniformly continuous.
- (iii)  $\limsup_n \mathbb{P}_n(F) \leq \mathbb{P}(F)$  for any closed set  $F$ .
- (iv)  $\liminf_n \mathbb{P}_n(\mathcal{O}) \geq \mathbb{P}(\mathcal{O})$  for any open set  $\mathcal{O}$ .
- (v) For all  $A \in \mathcal{S}$  such that  $\mathbb{P}(\partial A) = 0$ ,  $\mathbb{P}_n(A) \rightarrow \mathbb{P}(A)$ .

An archetype of the convergence in distribution is the following theorem (see also Figure 8)

**Theorem 3 (Central limit theorem)** Let  $(Y_i)$  be a sequence of i.i.d. random variables, having mean  $m$  and variance  $\sigma^2 \in (0, +\infty)$ . Let

$$X_n := \frac{\sum_{i=1}^n (Y_i - m)}{\sqrt{\sigma^2 n}}$$

be the rescaled sum (with mean 0 and variance 1). We have  $X_n \xrightarrow[n]{(d)} \mathcal{N}(0, 1)$ .

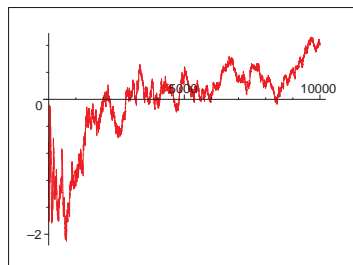


Figure 8: We have simulated a sequence  $(\frac{\sum_{i=1}^n (Y_i - m)}{\sqrt{\sigma^2 n}}, n \geq 1)$  satisfying the requirement of Theorem 3. It does not converge a.s., but in distribution.

**Exercise 4 :** a) Find in the literature the hierarchy of the convergences.

b) Using characterization (3) show the central limit theorem for a sum of i.i.d. random variables having distribution the Bernoulli distribution  $B(p)$ ,  $Poisson(\lambda)$  (subtract the mean, divide by the square root of the variance).

## 2.2 Exercise: analysis of the mean profile of binary search trees

We will use here some stuff introduced in the next Section, concerning simple notion on trees; the reader unaware of these notions, can skip this exercise in a first reading.

We present under an exercise the analysis of the mean profile of a binary search tree. Lot is known on binary search trees, among other thanks to the efforts of Luc Devroye <http://cg.scs.carleton.ca/~luc/> (there is also an entire book by Mahmoud, on the topic). Here we present a simple result, which can be proved by several methods, and which can be attributed to Guy Louchard.

A binary search tree is a structure associated with quick-sort. Quick-sort is an algorithm used to sort a list of data  $l = (x_1, \dots, x_n)$  from a set  $E$ , equipped with a total order  $<$ ; it works recursively as follows (the list is assumed to contain distinct elements). Take the first element  $x_1$  of  $l$ , and by successive comparison with  $x_1$ , build two lists  $l_{<}^{x_1}$  and  $l_{>}^{x_1}$  of elements of  $(x_2, \dots, x_n)$  smaller and larger than  $x_1$ . Apply recursively the same algorithm to these lists, until each list contains 0 or 1 element.

The binary search tree is practically used to store/retrieve data in a computer. Formally, it is a binary labeled tree, that is iteratively constructed as follows;

At the beginning the tree is empty. When a new data  $x_i$  is inserted in a binary search tree,

- . if the tree is empty it is placed at the root,
- . if the tree is not empty,  $x_i$  is compared to the data  $y$  placed at the root; if  $x_i$  is larger than  $y$ ,  $x_i$  is inserted in the right subtree of  $y$ , if is smaller,  $x_i$  is inserted in the left subtree rooted  $y$ .

Hence, when  $x_1, \dots, x_n$  have been successively inserted, the data in the left subtree of  $x_1$  is  $l_{<}^{x_1}$ , and those in the right one is  $l_{>}^{x_1}$ . Let  $BST_n$  be the binary tree having as internal nodes the labeled nodes hence constructed (if  $n$  data are inserted, the tree has size  $2n + 1$ ). Denote also by  $t_n$  the underlying tree of  $BST_n$ , that is the tree without the labels.

1) Assume  $E = \mathbb{R}$ , endowed with the usual order  $<$ . Draw  $BST_5$  corresponding to the data  $(x_1, x_2, x_3, x_4, x_5) = (0.3, 0.5, 0.7, 0.4, 0.45)$ .

2) Let  $x_6$  be a new data in  $[0, 1]$  to be inserted. Discuss the difference between  $BST_6$  and  $BST_5$  according to the value of  $x_6$ .

3) Discuss in terms of  $BST_n$  the following quantities (cost=number of comparisons): cost to build  $BST_n$ , maximal cost for a successful query (that is, to retrieve a data which is in the tree), maximal cost of an unsuccessful query (that is, to be sure that a data is not in the tree).

In order to analyze the stochastic behavior of the binary search tree, we now consider what is called the permutation model. We assume from now on that the sequence  $(x_i)$  is a sequence of i.i.d. random variables, uniform on  $[0, 1]$ .

The underlying tree  $t_n$  associated with the list  $l$  does not depend on the value of the list, but only



of the relative order of the data, that is, it depends on the permutation  $\sigma^n$  in the symmetric group  $S_n$ , such that

$$x_{\sigma_1^n} < \dots < x_{\sigma_n^n}. \quad (1)$$

4) Under the permutation model show that the permutation  $\sigma$  in (1) is uniform in  $S_n$ ; is it still the case if the  $x_i$  are not uniform, but say follows another distribution having a density? What is law of the rank of (a new data)  $x_{n+1}$  in the list  $x_1, \dots, x_{n+1}$ ? Give the evolution rule for  $t_n$  (be careful, the question is on  $t_n$ , not on  $BST_n$ ).

5) Under the permutation model, conditionally on  $x_1$ , how many nodes in the left subtree of  $BST_n$  (rooted on the first child of the root) when  $x_2, \dots, x_n$  have been inserted?. When  $n \rightarrow +\infty$ , what can be said? Discuss the same question for another subtree.

6) Using (3), what is the law of the size of the left subtree of  $t_n$ ? Compute this again using (4), by integration of the value of  $x_1$  (we can use the following formula  $\int_0^1 x^a(1-x)^b dx = \frac{a!b!}{(a+b+1)!}$  valid for  $a > 0, b > 0$ ).

7) We enter now in the study of the mean profile. Let  $Z(n) := (Z_k(n), k \geq 1)$  be the (leaf) profile of  $t_n$ , that is the successive number of leaves at level  $k \geq 1$  in  $t_n$  (recall that  $n$  counts the number of internal nodes).

a) Using (4), show that for  $n \geq 0$ ,

$$\mathbb{E}(Z_k(n+1) \mid (Z_k(n), k \geq 0)) = Z_k(n) \left(1 - \frac{1}{n+1}\right) + 2\frac{Z_{k-1}(n)}{n+1}, \quad (2)$$

this being initialized by  $Z_0(1) = 1, Z_i(0) = 0$  for  $i \geq 1$ .

From (2) we deduce, by taking the expectation

$$\mathbb{E}(Z_k(n+1)) = \frac{n\mathbb{E}(Z_k(n))}{n+1} + 2\frac{\mathbb{E}(Z_{k-1}(n))}{n+1}, \quad (3)$$

and thus, setting  $z_k(n) = \frac{\mathbb{E}(Z_k(n))}{n+1}$  this gives

$$z_k(n+1) = \frac{nz_k(n) + 2z_{k-1}(n)}{n+2} \quad \text{and } z_0(1) = 1. \quad (4)$$

b) For  $n \geq 0$ , consider a random walk  $W_n = 1 + X_1 + \dots + X_n$  where the  $X_i$  are independent but not i.i.d.:  $\mathbb{P}(X_k = 1) = \frac{2}{k+1}, \mathbb{P}(X_k = 0) = 1 - \frac{2}{k+1}$ . Express  $p_k(n+1) := \mathbb{P}(W_{n+1} = k)$  in terms of the vector  $(p_j(n), j \geq 0)$ . Compare to (4).

c) Check the Lindeberg's condition<sup>2</sup> for  $W_n$ , and state a limit theorem for  $W_n$ ; what can be deduced

---

<sup>2</sup>Lindeberg's condition for central limit theorem: Let  $(X_k)$  be a sequence of real r.v., independent, defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Assume the expected values  $\mathbb{E}(X_k) = \mu_k$  and variances  $\text{var}(X_k) = \sigma_k^2$  exist and are finite, and let  $s_n^2 := \sum_{k=1}^n \sigma_k^2$ . If this sequence of independent r.v.  $X_k$  satisfies the Lindeberg's condition:

$$\text{for all } \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}((X_k - \mu_k)^2 1_{|X_k - \mu_k| > \varepsilon s_n}) = 0,$$

then the central limit theorem holds, i.e. the random variable

$$Z_n := \frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n}$$

converges in distribution to a standard normal random variable as  $n \rightarrow \infty$ .

for the profile of BST?

d) Make some simulations: draw the profile of a binary search tree with 100 000 nodes associated with a sequence of i.i.d. random variables  $x_i$ , uniform on  $[0, 1]$ . What can be conjectured?

### 3 Paths and trees : discrete considerations

#### 3.1 Planar trees

We will use the formalism introduced in probability theory by Neveu, which is particularly well adapted to probabilist considerations, and which is moreover the usual formalism in computer science. The trees considered in this course are rooted, and planar, meaning that the children of each node are ordered.

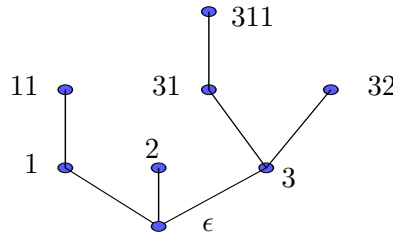


Figure 9: Representation of the tree  $E = \{\epsilon, 1, 2, 3, 11, 31, 32, 311\}$

Each node is a word; a tree is a set of words. We proceed as follows. Let  $U = \epsilon \cup \bigcup_{n \geq 1} \mathbb{N}^{*n}$  be the set of finite words that can be written on the alphabet  $\mathbb{N}^* := \{1, 2, 3, \dots\}$ ;  $\epsilon$  designates the empty word. Denote by  $uv$  the concatenation of the words  $u$  and  $v$ ; by convention  $\epsilon u = u\epsilon = u$ .

**Definition :** A part  $T$  of  $U$  is a tree if it has the following properties :

- $\epsilon \in E$
- if  $uv \in E$  then  $u \in E$
- if  $uk \in E$  for  $k > 1$  and  $u \in U$  then  $u(k-1) \in E$ .

Some vocabulary : let  $T$  be a tree. The elements of  $T$  are called nodes. The word  $\epsilon$  is called the root of  $T$ ; the length of a word  $u$  in  $T$ , denoted by  $|u|$ , is called the height of  $u$  in  $T$  or the level of  $u$  in  $T$  (we have  $|\epsilon| = 0$ ). For  $u$  in  $T$ ,

$$c_u(T) = \max\{k \mid uk \in T\}$$

is the number of children of  $u$  in  $T$  (and, the  $u_i$ 's, for  $i \in \mathbb{N}$  are the children of  $u$ ). Hence, if it exists,  $u_1$  is called the first child of  $u$ ,  $u_2$  the second, etc.. The cardinality of  $T$  is called the size of the tree, and denoted by  $|T|$ .

The properties imposed in the definition of the tree are clear : a tree must have a root, all the prefixes=ancestors of a node must be in the tree, and the children of a node  $u$  in  $T$  are ordered from  $u_1$  to the last one  $u_{c_u(T)}$  (these properties ensure a unique tree for “each tree shape”).

The lexicographical order on  $U$  denoted by  $<_{lex}$  can be projected on any tree  $T$ , making of it a totally ordered set. Trees are often represented as done on Figure 9, omitting the words; it is then understood, that the root is at the bottom, that the children of each node are sorted from left to right.

The set of trees will be denoted by  $\mathcal{T}$ , and the set of trees with  $n$  nodes by  $\mathcal{T}_n$ .

One finds in the literature several other definitions for this same families of rooted planar tree : one of them, classical in computer science is the following one: a tree is either a single node, or a ordered (finite) sequence of trees (this is a recursive definition).

### 3.2 Encoding of finite trees

A very rich and efficient method to represent a tree (from a combinatorial point of view) consists in encoding them by some paths. For this, are used some traversal algorithms.

#### Breadth first search

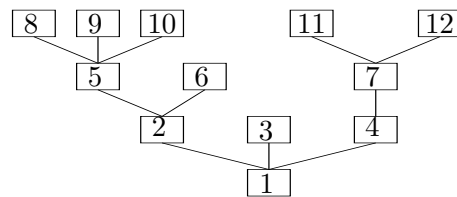


Figure 10: Order induced by the breadth first search

In the breath first search, the levels of the tree are visited successively, from the root to the top of the tree, and on a given level, the nodes are visited from left to right. On the above tree is indicated the visit order of the different nodes. [Formally, the breadth first order is defined as follows:  $u <_{br} v$  if  $(|u| < |v|$  or if  $(|u| = |v|$  and  $u <_{lex} v)$ ).

#### Depth first search

The nodes are visited according to the lexicographical order. First the root, then 1, then 11, etc.

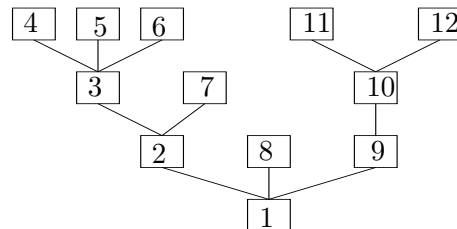


Figure 11: Depth first search order.

### 3.3 Encoding of trees by paths

#### 3.3.1 Queue representation

(We here use the word "queue", where usually in computer science is used the word "stack").

Let us make a depth first traversal of a tree  $T$  of  $\mathcal{T}$  (exactly the same work could be made with the breadth first search instead). Sort the nodes  $(u_1(T), \dots, u_{|T|}(T))$  of  $T$  according to the lexicographical order (hence  $u_1(T) = \epsilon$ ). Denote by  $y_i(T) = c_{u_i(T)}(T)$  the number of children of  $u_i(T)$  in  $T$ . We thus have associated with any tree  $T$  of  $\mathcal{T}$  a sequence

$$D(T) = (y_1(T), \dots, y_{|T|}(T))$$

as done on Figure 12, that we call description of  $T$ .

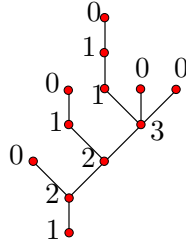


Figure 12: the sequence  $(1,2,0,2,1,0,3,1,1,0,0,0)$  is the sequence of length 12 corresponding to this tree of size 12

**Lemma 1** *The map  $D$  is an injection from  $\mathcal{T}$  onto  $\cup_{n \geq 1} \mathbb{N}^n$ .*

**Proof .** Take two distinct trees  $T_1$  and  $T_2$  in  $\mathcal{T}$  . If  $c_\epsilon(T_1) \neq c_\epsilon(T_2)$  then  $y_1(T_1) \neq y_1(T_2)$ , and then  $D(T_1) \neq D(T_2)$ , it's done. Assume that  $c_\epsilon(T_1) = c_\epsilon(T_2)$ , and consider  $u$  the minimal element of  $U$  which is not in both  $T_1$  and  $T_2$ ; assume that  $u$  is in  $T_1$ . By definition of the depth first traversal, if  $u$  is the  $l$ th node in  $T_1$ , then  $(u_1(T_1), \dots, u_{l-1}(T_1)) = (u_1(T_2), \dots, u_{l-1}(T_2))$ , that is both trees contains the same  $l - 1$  first nodes. Let us see why  $(y_1(T_1), \dots, y_{l-1}(T_1)) \neq (y_1(T_2), \dots, y_{l-1}(T_2))$ .

The father  $v$  of  $u$  belongs to both  $T_1$  and  $T_2$ , and has the same rank in both trees, in the lexicographical order (say it is  $v = u_j(T_1) = u_j(T_2)$  for  $j \leq l - 1$ , since it is visited before its child  $u$ ). But it is clear that  $c_v(T_1) > c_v(T_2)$ . Indeed, let  $k$  be the rank of  $u$  among the children of  $v$ , that is let  $k$  such that  $u = vk$ . Since  $u \in T_1$  then  $c_v(T_1) \geq k$ ; in the other hand, since  $u \notin T_2$  then  $c_v(T_2) < k$ . Hence,  $(y_1(T_1), \dots, y_{l-1}(T_1)) \neq (y_1(T_2), \dots, y_{l-1}(T_2))$ , and thus  $D(T_1) \neq D(T_2)$ .  $\square$

The map  $D$  is obviously not a surjection on  $\cup_{n \geq 1} \mathbb{N}^n$ . For example the sequence  $(2, 4)$  has no pre-image. For any  $n$ , consider  $\mathbf{D}_n$  the set of sequences of non negative integers  $y = (y_1, \dots, y_n)$  satisfying the following set of constraints :

$$(C) = \begin{cases} y_1 & \geq 1, \\ y_1 + y_2 & \geq 2, \\ & \vdots \\ y_1 + \dots + y_{n-1} & \geq n-1, \\ y_1 + \dots + y_n & = n-1. \end{cases}$$

(for  $n = 1$ , the only constraint is  $y_1 = 0$ ).

**Proposition 1** For any  $n \geq 1$ , the restriction of  $D$  on  $\mathcal{T}_n$  is a bijection on  $\mathbf{D}_n$ .

Before proving this result, for any  $n \geq 1$ , we introduce the sets  $\mathbf{L}_n$ , of paths  $s := (s_0, s_1, \dots, s_n)$  with increments in  $\{-1, 0, 1, 2, \dots\}$ , satisfying the constraints

$$(C') = \begin{cases} s_0 & = 0, \\ s_1 & \geq 0, \\ & \vdots \\ s_{n-1} & \geq 0, \\ s_n & = -1. \end{cases}$$

In the literature, the elements of  $\cup_n \mathbf{L}_n$  are called Lukaciewish paths.

For any  $n \geq 1$ , there is an obvious bijection  $\Psi_n$  between  $\mathbf{D}_n$  and  $\mathbf{L}_n$ . The image of  $(y_1, \dots, y_n) \in \mathbf{D}_n$  being simply  $s := \Psi_n((y_1, \dots, y_n)) = (s_0, s_1, \dots, s_n)$  the path with increments the  $y_i - 1$  :

$$s_0 = 0, \quad \text{and for } 1 \leq j \leq n, \quad s_j = (y_1 - 1) + \dots + (y_j - 1). \quad (5)$$

**Proof of Proposition 1 .** In the proof,  $n \geq 1$  is fixed.

1) First, take  $T \in \mathcal{T}_n$  and consider  $D(T) = (y_1(T), \dots, y_n(T))$ ; we have to check that  $D(T)$  belongs to  $\mathbf{D}_n$ , that is, satisfies (C): a tree with  $n$  nodes contains  $n - 1$  children, then  $y_1(T) + \dots + y_n(T) = n - 1$ . In the lexicographical order, a node is visited before its children. Hence,  $y_1(T) + \dots + y_k(T)$  counts the number of children of  $u_1(T), \dots, u_k(T)$ , which include  $u_2(T), \dots, u_k(T)$  and also  $u_{k+1}(T)$  if  $k+1 \leq n$ , since the  $k+1$ th node if it exists, is a child of one of the nodes  $u_1(T), \dots, u_k(T)$ ; therefore  $y_1(T) + \dots + y_k(T) \geq k$  if  $k \leq n - 1$ . Hence  $D(T)$  is indeed  $\mathbf{D}_n$ . Moreover by Lemma 1, since  $D$  is an injection, its restriction to  $\mathcal{T}_n$  is then an injection from  $\mathcal{T}_n$  on  $\mathbf{D}_n$ .

2) To end the proof, it suffices to show that for any element  $d = (y_1, \dots, y_n) \in \mathbf{D}_n$ , there exists  $T$  in  $\mathcal{T}_n$  such that  $d = D(T)$ . For this we use a counting argument and show that  $\#\mathbf{D}_n = \#\mathcal{T}_n$ ; first, one has

$$\#\mathcal{T}_n = \sum_{k=1}^{n-1} \sum \#\mathcal{T}_{x_1} \times \dots \times \#\mathcal{T}_{x_k};$$

where the second sum ranges on the  $k$  tuples  $(x_1, \dots, x_k)$  such that  $x_1 \geq 1, \dots, x_k \geq 1$  and  $x_1 + \dots + x_k = n - 1$  (this is a simple decomposition of the trees of  $\mathcal{T}_n$  according to the degree  $k$  of the root, and the number of nodes  $x_1, \dots, x_k$  in the  $k$  subtrees rooted at the children of the root).

Since we already know that  $\#\mathbf{D}_n = \#\mathbf{L}_n$ , let us establish that  $\#\mathbf{L}_n = \#\mathcal{T}_n$ . Any path  $s = (s_0, s_1, \dots, s_n) \in \mathbf{L}_n$  can be decomposed as follows: consider  $s_1$  (the first step). Since  $s$  has its increment in  $\{-1, 0, 1, 2, \dots\}$ ,  $s_1 = j$  belongs to  $\{0, 1, \dots, n-2\}$ . Now, the path  $(s_1 - s_1, s_2 - s_1, \dots, s_n - s_1)$  starts from 0, end at position  $-1 - j$ , and has length  $n-1$ ; it can be decomposed into  $j+1$  parts as on Figure 13, corresponding to the sections of the paths between the first hitting

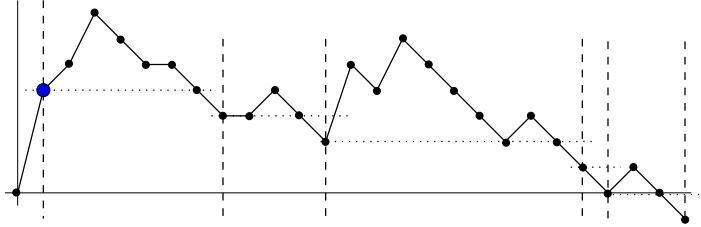


Figure 13: Decomposition of a Lukaciewish path after its first steps, into Lukaciewish paths.

time of  $-1, -2, \dots, -j-1$ . Each of these parts is an element of  $\cup_{m \geq 1} \mathbf{L}_m$ . This decomposition immediately gives

$$\#\mathbf{L}_n = \sum_{k=0}^{n-2} \sum \#\mathbf{L}_{x_1} \times \dots \times \#\mathbf{L}_{x_k},$$

and the second sum ranges on the same  $x_i$ 's as above. Since  $\#\mathbf{L}_1 = |\mathcal{T}_1| = 1$ , this suffices by induction to end the proof.  $\square$

We here sum up some information that can be deduced from this construction.

**Proposition 2** *The depth first search (resp. breadth first search) induced a bijection  $\Gamma_{lex}$  (resp.  $\Gamma_{br}$ ) from  $\mathcal{T}$  onto  $\cup_{n \geq 1} \mathbf{L}_n$ , sending bijectively  $\mathcal{T}_n$  onto  $\mathbf{L}_n$  : for  $t \in \mathcal{T}_n$ , if  $\Gamma_{lex}(t) = s$  (resp.  $\Gamma_{br}(t) = s$ ) then the  $i$ th increment of  $s$  is  $c_{u_i}(t) - 1$  the number of children minus 1 of the  $i$ th node of  $t$  according to the lexicographical order (resp. breadth first order), for  $i$  from 1 to  $n$ .*

### Breadth first queue (BFQ)

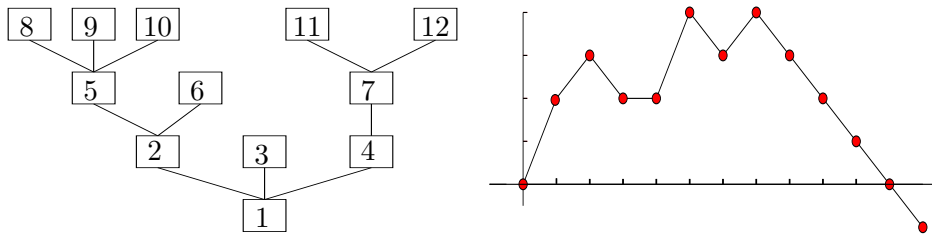


Figure 14: A tree and its BFQ

Let  $T \in \mathcal{T}$  be a tree. The Lukaciewish path  $S_T = \Gamma_{br}(T)$  is called the breadth first queue

(BFQ) of  $T$ . As said above,

$$S_T(0) = 0 \quad \text{and for } j \in \{1, \dots, |T|\}, \quad S_T(j) = \sum_{i=1}^j (y_i(T) - 1),$$

where  $y_i(T)$  is the number of children of the  $i$ th node of  $T$  according to the breadth first order.

### Depth first search (DFS)

The DFS allows to associate with each tree  $T \in \mathcal{T}$ , three different paths.

– First, the Lukaciewish path  $S_T = \Gamma_{lex}(T)$  is called the depth first queue (DFQ) of  $T$ . As said above,

$$S_T(0) = 0 \quad \text{and for } j \in \{1, \dots, |T|\}, \quad S_T(j) = \sum_{i=1}^j (y_i(T) - 1),$$

where  $y_i(T)$  is the number of children of the  $i$ th node of  $T$  according to the breadth first order.

– The second path we associate to a given tree, is the *contour process*. This construction is classical in combinatorics: a tree  $T$  having  $n$  nodes has  $n - 1$  edges. We will a function  $F_T$  (called the contour function) from  $\llbracket 0, 2(n - 1) \rrbracket$  taking its value in the set of the nodes of  $T$ . Set  $F_T(0) = \epsilon$ , the root of  $T$ .  $F_T(k)$  is built as follows : if the node  $F_T(k - 1)$  has some non visited children (meaning, not in the list  $\{F_T(0), \dots, F_T(k - 2)\}$ ) then  $F_T(k)$  is the leftmost non visited child of  $F_T(k - 1)$ . If all children of  $F_T(k - 1)$  have been visited, then  $F_T(k)$  is the father of  $F_T(k - 1)$ . The contour function is the function  $v_T$  from  $\llbracket 0, 2(|T| - 1) \rrbracket$  taking its values in  $\mathbb{R}^+$  defined by

$$V_T(k) = |F_T(k)|,$$

the distance to the root of  $F_T(k)$ . The function  $V_T$  gives the height of a fly walking clockwise around the tree, one edge by unit of time, according to the time. It's important to notice the huge similarity between a tree and its contour process; this will appear to be crucial in the construction of a notion of limiting tree. By construction  $V$  is a path starting at 0, ending at 0, having increments  $\pm 1$ , and non negative. Such a path is called "Dyck path".

– The third path associated with  $T$  is its "height process" :

$$H_T(i) = |u_{i+1}(T)| \quad \text{for } 0 \leq i \leq |T| - 1.$$

**Proposition 3** *Each of the path, BFQ, DFQ, height process and contour process, characterizes the tree it is coming from.*

(We mean by this assertion that the map which associates with a tree its height process (for example) is an injection.)

**Exercise 5 :** Let  $T$  be a tree with  $n$  nodes. Denote by  $Z_0(T), Z_1(T), Z_2(T), \dots$  the number of nodes at distance 0, 1, 2, ... from the root. For example, on the tree Figure 15,  $Z_0 = 1, Z_1 = 2, Z_2 = 3, Z_3 =$

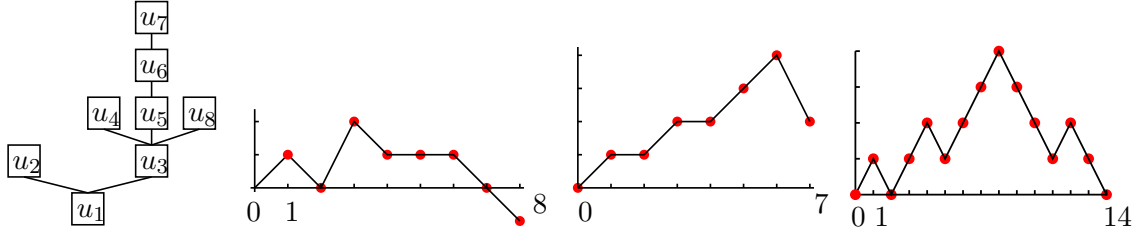


Figure 15: : a tree, its DFQ, its height process and its contour process.

$1, Z_4 = 1, Z_5 = 0 \dots$ . The width of the tree  $T$  is defined by  $W(T) = \max Z_k(T)$ , and the height is defined by  $H(T) = \max\{k, Z_k(T) \neq 0\}$  (as usual).

- 1) Express  $W(T)$  and  $H(T)$  in term of the contour process  $V_T$
- 2) Express  $W(T)$  and  $H(T)$  in term of the BFQ  $S_T := \Gamma_{br}(T)$  of  $T$ .
- 3) Let  $N_k(T)$  be the number of nodes in  $T$  having  $k$  children. Express  $N_k(T)$  in terms of the BFQ  $S_T$ , and in terms of the DFQ  $S'_T := \Gamma_{lex}(T)$  of  $T$ .

### 3.4 Relation between $H_T$ , $V_T$ and $S_T$

Further, a probability distribution on trees will be introduced; for this distribution, it will appears that the DFQ  $S_T$  is a simple random process. The contour process  $V_T$  is the rich process from a topological point of view, since it deeply resembles to the tree. Studying the relation between these processes will be an important tool to prove the convergence of the contour process, and then later, of the tree itself. The height process appears to be an intermediate on this way.

#### 3.4.1 Link between $H_T$ and $V_T$

Here, we just transform into formulas a simple fact: the height process as well as the contour process gives the successive height of the nodes in the tree traversal. The contour process encodes several times the height of the nodes, and the height process only once.

Let  $T$  be a tree of  $\mathcal{T}_n$ , and let  $(u_1(T), \dots, u_n(T))$  be its nodes, sorted according to the lexicographical order. For any  $l \in \{1, \dots, n\}$  set

$$m_T(l) = \inf\{k \mid F_T(k) = u_{l+1}(T)\};$$

$m_T(l)$  gives the first visit time of  $u_{l+1}(T)$  by the contour function  $F_T$ . Hence  $V_T(m_T(l)) = |F_T(k)| = |u_{l+1}| = H_T(l)$  for any  $l \in \{1, \dots, n\}$ .

**Lemma 2** For any  $l \in \{1, \dots, n\}$ ,

$$m_T(l) + H_T(l) = 2l.$$



**Proof .** for  $l = 0$ , it's true. Then we proceed inductively, observing that  $m_T(l)$  gives the successive first visit time of nodes sorted according to the lexicographical order. Two cases appear :

- if  $u_{l+2}(T)$  is the first son of  $u_{l+1}(T)$ , then  $m_T(l+1) = m_T(l) + 1$  and  $H_T(l+1) = H_T(l) + 1$ , for which we see that  $m_T(l+1) + H_T(l+1) = m_T(l) + H_T(l) + 2$ ;
- if  $u_{l+2}(T)$  is not the first son of  $u_{l+1}(T)$  then it is necessarily a son of one of the ancestor  $v$  of  $u_{l+1}(T)$ . Therefore  $H_T(l+1) = |v| + 1$  and  $m_T(l+1) = m_T(l) + d_T(u_{l+1}(T), v) + 1 = m_T(l) + |u_T(l)| - |v| + 1$ ; hence  $m_T(l+1) + H_T(l+1) = m_T(l) + |u_T(l)| + 2$ .  $\square$

Here is a second lemma, that controls the difference between both processes; it will also be useful later on, for asymptotic considerations.

**Lemma 3** For  $l \in \{1, \dots, n\}$  and for any  $k \in \llbracket m_T(l), m_T(l+1) \rrbracket$  :

$$\begin{aligned} V_T(m_T(l+1)) - 1 &\leq V_T(k) \leq V_T(m_T(l)) \\ H_T(l+1) - 1 &\leq V_T(k) \leq H_T(l). \end{aligned}$$

**Proof .** The contour process is a kind of interpolation of the height process :

$$V_T(k) = \begin{cases} H_T(l) & \text{if } k = m_T(l) \text{ for a given } l \\ H_T(l) - (k - m_T(l)) & \text{if } k \in \llbracket m_T(l) + 1, m_T(l+1) - 1 \rrbracket \text{ for a given } l \end{cases} \quad \square$$

### 3.4.2 Link between the DFQ and the height process

Let  $u_1(T), \dots, u_n(T)$  be the nodes of a tree  $T$  of  $\mathcal{T}_n$ , sorted according to the lexicographical order; let  $S_T$  be the corresponding DFQ. We say that  $j$  (for a  $j < l$ ) is a right minimum for  $S_T$  on  $\llbracket 0, l \rrbracket$  if

$$\min_{j \leq k \leq l} S_T(k) = S_T(j).$$

Let also denote by  $RM_T(l) = \{j \mid \text{and right minima of } S_T \text{ on } \llbracket 0, l \rrbracket\}$ .

**Lemma 4** (Le Gall & Le Jan) For any  $l \in \llbracket 0, n-1 \rrbracket$  and  $j \in \llbracket 0, l-1 \rrbracket$ ,  $j$  is a right minimum for  $S_T$  on  $\llbracket 0, l \rrbracket$  if and only if  $u_{j+1}(T)$  is an ancestor of  $u_{l+1}(T)$ . Therefore

$$\#RM_T(l) = H_T(l). \quad (6)$$

Formula (6) will appear to be fundamental ; it will be the cornerstone of the convergence of the contour process.

**Proof .** We interpret  $S_T$  as the size of a queue in a LIFO (last in first out) system. Consider a LIFO server, where one client per unit of time gets out of the queue. The nodes of  $T$  are the clients. The root  $u_1(T)$  arrives at time 0; set  $queue(0) := \{u_1(T)\}$ . The size of the queue is then

$$S'_T(0) = 1.$$

The evolution of the queue is then as follows : at time  $i$ , the children of  $u_i(T)$  arrive in the queue,

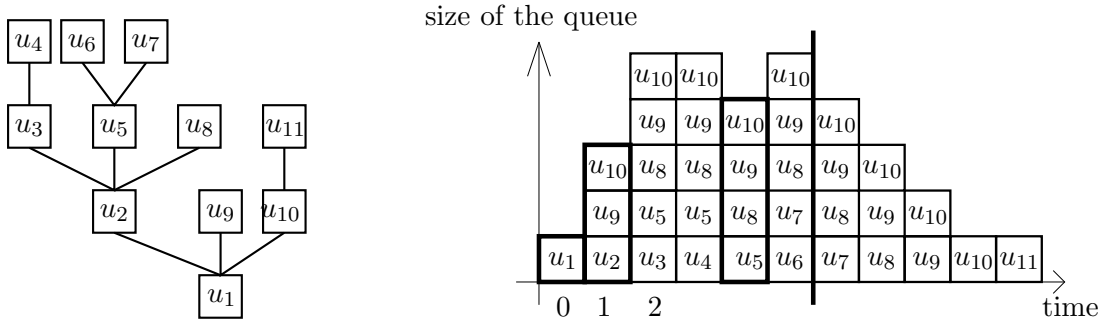


Figure 16: : LIFO queue size associated with a tree. Observe the node 6. The node 6 is present in the queue at time 5, and no more at time 6, by construction. Its depth in the tree is 3 : its ancestors are  $u_5$ ,  $u_2$  and  $u_1$ . The right minima of  $S_T$  on  $\llbracket 0, 5 \rrbracket$  are exactly  $1-1=0$ ,  $2-1=1$  and  $5-1=4$ .

from the rightmost one (first) to the leftmost one (last), and then in the queue, the leftmost one will get out of the queue at first. Simultaneously,  $u_i(T)$  leaves the queue

$$\text{queue}(i) = (\text{queue}(i-1) \setminus \{u_i(T)\}) \cup \{u_i(T)1, \dots, u_i(T)c_{u_i(T)}\}.$$

The size of the queue is then

$$S'_T(i) = S'_T(i-1) + c_{u_i(T)} - 1.$$

Hence the queue size process and the DFQ are related by

$$S'_T(i) = S_T(i) + 1, \quad \text{for any } i \in \llbracket 0, n \rrbracket.$$

Hence the right minima of  $S'_T$  and  $S_T$  coincide.

In the LIFO procedure, the parents are served before their children. Consider the node  $u_k(T)$ ; it is in  $\text{queue}(k-1)$ , not in  $\text{queue}(k)$ ; let  $T_k$  be the subtree of  $T$  rooted at  $k$ . At time  $k$ , appear the children of  $u_k(T)$  in the queue; it is easy to see that the first nodes that are served after  $u_k(T)$  are the nodes from  $T_k$  (apart  $u_k(T)$ ,  $|T_k| - 1$  nodes are concerned). Therefore, it is immediate that at times  $(k-1) + 1, (k-1) + 2, \dots, (k-1) + |T_k| - 1$  the queue contains the same nodes as that at time  $k-1$  (except  $u_k(T)$ ) plus some nodes from  $T_k$ ; thus

$$S'_T(k-1+i) \geq S'_T(k-1) \text{ for any } i \in \{0, 1, \dots, |T_k| - 1\}.$$

At time  $(k-1) + |T_k|$  the queue contains no element of  $T_k$ , no more  $u_k(k)$  and no new elements. Hence

$$S'_T(k-1+|T_k|) = S'_T(k-1) - 1;$$

Hence,  $k-1$  is a right minima on  $\llbracket 0, (k-1) + |T_k| - 1 \rrbracket$  but not in  $\llbracket 0, (k-1) + |T_k| \rrbracket$ . This suffices to conclude:  $j$  is a right minima on  $[0, l]$  if  $u_{l+1}(T)$  is a descendant of  $u_{j+1}(T)$ , and it is not the case if  $u_{l+1}(T)$  is a not a descendant of  $u_{j+1}(T)$ .  $\square$

### 3.5 Random walks

We call *path* (or *walk*) a sequence (finite or infinite)  $(S_k)_{k \in \llbracket 0, N \rrbracket}$  of elements of  $\mathbb{R}$  (or of any set, more generally).  $S_0$  is called the initial position and  $N$  the length of the path. In the sequel we will assume that  $S_0 = 0$ . Denote by  $X_k = S_k - S_{k-1}$  the  $k$ th increment of the path  $S$ . We have, for any  $k \in \llbracket 0, N \rrbracket$

$$S_k = \sum_{j=1}^k X_j.$$

We represent usually such a path by a sequence of points  $(k, S_k)_{k \in \llbracket 0, N \rrbracket}$  in  $\mathbb{R}^2$ , sometimes interpolated linearly.

**Definition :** A *random walk* with length  $N$  is a path  $(S_k)_{k \in \llbracket 0, N \rrbracket}$  where the increments are i.i.d. random variables.

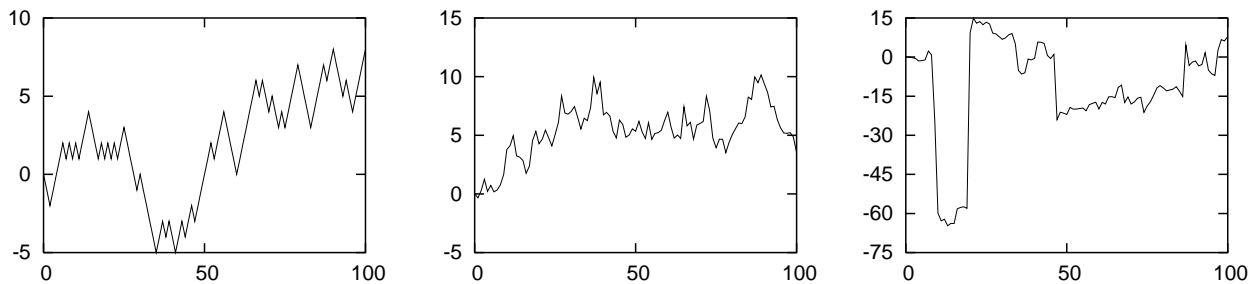
When  $S_0 = 0$ , and if all the  $X_i$  are  $\pm 1$ , independently with probability  $1/2$ , the  $2^N$  paths with length  $N$  have the same probability. This model is called the model of simple random walk.

When the  $X_i$  have a discrete distribution  $\mu$ , that is,  $\mathbb{P}(X_i = k) = \mu_k$  for  $k \in \mathbb{Z}$ , then

$$\begin{aligned} \mathbb{P}(S_1 = s_1, S_2 = s_2, \dots, S_n = s_n) &= \mathbb{P}(X_1 = s_1, X_2 = s_2 - s_1, \dots, X_n = s_n - s_{n-1}) \\ &= \mu_{s_1} \mu_{s_2 - s_1} \cdots \mu_{s_n - s_{n-1}} \end{aligned} \quad (7)$$

**Exercise 6 :** When the increments have a density, how reformulate (7)?

Clearly, a distribution  $\mu$  of the increments, induces a distribution on the corresponding random walk; usually a random walk is described via the distribution of its increments. The distribution



**Figure 17:** Simulation of three random walks with length 100. The first one has increments  $X_i = \pm 1$  with probability  $1/2$ , for the second one  $X_i \sim \mathcal{N}(0, 1)$ , for the third one,  $X_i \sim \text{Cauchy } C(1)$  (distribution with density  $\frac{1}{\pi} \frac{1}{1+x^2}$  on  $\mathbb{R}$ ).

of the final position  $S_N$  is given by the  $N$ th self convolution of  $\mu$  (that is  $\mu^{*N}$ ). The asymptotic behavior of the whole trajectory  $(S_k)_{k=0, \dots, N}$  will be done further in the course; one may first say that the asymptotic behavior of  $S_n/n$  is given by the law of large numbers (depending on the

existence or not of  $\mathbb{E}(X_1)$ , and of its value); more thinly  $(S_n - \mathbb{E}(S_n))/\sqrt{n}$  will have an asymptotic Gaussian behavior (if  $\text{var}(X_1)$  exists).

### 3.6 A probability distribution on the set of trees

A Galton-Watson (GW) process is a Markov chain  $(Z_n)_{n=0,1,2,\dots}$  on  $\mathbb{N}$  such that  $Z_0 = 1$  and such that, for  $k \geq 0$ ,  $Z_{k+1}$  can be written :

$$Z_{k+1} = \sum_{l=1}^{Z_k} Y_l^{(k)}$$

where the random variables  $Y_l^{(k)}$  are i.i.d. copies of  $Z_1$ , having distribution  $\mathbf{p} := (p_j)_{j \geq 0}$ . In more concrete terms :

- At time  $T = k$ ,  $Z_k$  individuals are alive.
- At time  $T = k + 1$ , each of the  $Z_k$  individual dies, but produces  $j$  descendants with probability  $p_j$ , independently from the other individuals. The distribution  $\mathbf{p}$  is called the *offspring distribution* (noted OD for short). The genealogical tree of this population is called *Galton-Watson tree*. These trees are planar trees: there is a root (the first individual who was there at time 0), and the different children of a nodes are ordered.

Denote by  $f$ , the generating function of  $\mathbf{p}$  :

$$f(s) = \sum_{k \geq 0} p_k s^k = \mathbb{E}(s^{Y_1}) \text{ for } |s| \leq 1.$$

Let  $f_n$  be the generating function of  $Z_n$ . For  $n \geq 2$

$$\begin{aligned} f_n(s) &= \mathbb{E}(s^{Z_n}) = \mathbb{E}_{\mathbf{p}}(s^{\sum_{i=1}^{Z_{n-1}} Y_i}) = \mathbb{E}\left(\sum_{k \geq 0} s^{\sum_{i=1}^{Z_{n-1}} Y_i} \mathbb{I}_{Z_{n-1}=k}\right) \\ &= \sum_{k \geq 0} \mathbb{E}_{\mathbf{p}}(s^{\sum_{i=1}^k Y_i}) \mathbb{E}_{\mathbf{p}}(\mathbb{I}_{Z_{n-1}=k}) \\ &= \sum_{k \geq 0} \mathbb{E}_{\mathbf{p}}(s^{Y_1})^k \mathbb{P}_{\mathbf{p}}(Z_{n-1} = k) \\ &= \sum_{k \geq 0} \mathbb{P}_{\mathbf{p}}(Z_{n-1} = k) f(s)^k \\ &= f_{n-1}(f(s)) \end{aligned}$$

Hence, the function  $f_n$  is simply  $f^{\circ n}$ . We have

$$\begin{aligned} \mathbb{E}_{\mathbf{p}}(Z_1) &= f'(1) \stackrel{\text{def}}{=} m \\ \mathbb{E}_{\mathbf{p}}(Z_n) &= f'_n(1) = f'_{n-1}(1) f'(1) = (f'(1))^n = m^n. \end{aligned}$$

Then, when  $m > 1$  (resp.  $m < 1$ ) the mean number of nodes of the  $n$  level increases (or decreases) exponentially fast. When  $m = 1$ , it is constant equal to 1.

Assume that  $p_0 + p_1 \neq 1$ . The *extinction time*  $T$  of the Galton-Watson process  $(Z_n)_{n \geq 0}$  is the first  $t$  for which  $Z_t = 0$  :

$$T = \inf\{t, t \in \mathbb{N}, Z_t = 0\};$$

(note that if  $Z_t = 0$ , then  $Z_{t+1} = 0$ . In term of the GW family tree, it is height  $+1$ . When  $p_0 > 0$ , the *extinction probability*  $q$ , defined by

$$q = \mathbb{P}_{\mathbf{p}}(T < +\infty),$$

is strictly positive ( $Z_1 = 0$  with probability  $p_0$ ). The extinction probability  $q$  equals 1 if  $m \leq 1$  (if  $p_1 \neq 1$ ) and is the smallest positive solution to  $f(t) = t$  if  $m > 1$ .

**Exercise 7 :** Prove this result. First, using  $f_n$  and  $f_{n+1}$ , show that  $q = f(q)$  thanks to a passage to the limit. Draw  $f$  on  $[0, 1]$ , and indicate the slope of  $f$  in 1. How many solution to this equation on  $[0, 1]$  when  $m < 1$ , when  $m > 1$ ? Conclude.

The *total population*  $X$  of the GW process is the random variable defined by

$$X = Z_0 + Z_1 + Z_2 + \dots .$$

Since the  $Z_1$  subtrees of the root are independent and have same size distribution as  $X$ , the following representation of  $X$  holds :

$$X \stackrel{(d)}{=} 1 + \sum_{k=1}^{Z_1} X_k$$

where the  $X'_k$ 's are independent copies of  $X$ , and are independent of  $Z_1$ . Let  $g(x) = \mathbb{E}_{\mathbf{p}}(x^X)$  be the probability generating function of  $X$ ;  $g$  is solution of the functional equation :

$$g(x) = x f(g(x)).$$

If  $\mathbb{E}(Z_1) < 1$ , one may deduce of this the mean size of the total population :

$$\mathbb{E}_{\mathbf{p}}(X) = g'(1-) = \frac{1}{1 - f'(1)} = \frac{1}{1 - \mathbb{E}_{\mathbf{p}}(Z_1)}.$$

If  $\mathbb{E}(Z_1) \geq 1$ , the mean size of the total population is  $+\infty$ .

### 3.7 Galton-Watson trees conditioned by the size

**Lemma 5** Denote by  $\mathbb{P}_{\mathbf{p}}$  the distribution of the family tree of a Galton-Watson tree with OD  $\mathbf{p}$ , satisfying  $m = \sum_{i \geq 0} ip_i \leq 1$  (this is the mean number of children of a node). Then  $\mathbb{P}_{\mathbf{p}}$  is a distribution on the set of finite trees : for any finite tree  $t \in \mathcal{T}$ ,

$$\mathbb{P}_{\mathbf{p}}(\{t\}) = \prod_{u \in t} p_{c_u(t)} = \prod_i p_i^{N_i(t)} \tag{8}$$

where  $N_i(t)$  is the number of nodes in  $t$  having  $i$  children.

**Sketch of proof .** Let  $(Y_u, u \in U)$  be a family of i.i.d. random variables indexed by the complete infinite tree  $U$  and having  $\mathbf{p}$  has distribution. Consider

$$\mathbf{T} := \{\epsilon\} \cup \{u = u^1 \dots u^n \in U \mid u^j \leq Y_{u^1 \dots u^{j-1}} \text{ for } j \in \llbracket 1, n \rrbracket\}.$$

One can check that  $\mathbf{T}$  is the family tree of a GW process, with OD  $\mathbf{p}$  (introduce  $Z_k$ , has above, and check that  $(Z_k, k \geq 0)$  has the same representation as that given at the beginning of Section 3.6). By what is explained above, the probability of extinction of this GW process is 1 a.s., and then the corresponding tree is finite with probability 1 (if  $p_1 \neq 1$ ). Now, take a finite tree  $t$ . When computing  $\mathbb{P}_{\mathbf{p}}(\mathbf{T} = t)$  one sees that only the variable  $Y_u$ , indexed by  $u \in t$  need to satisfy  $Y_u = c_u(t)$  in order than  $\mathbf{T} = t$ , the “others” being any integers. It follows that (by summing on all possible values of the “others”), that  $\mathbb{P}_{\mathbf{p}}(\mathbf{T} = t) = \prod_{u \in t} p_{c_u(t)}$ .  $\square$ .

Notice that when  $m > 1$ , as explained also in the exercise above,  $\mathbb{P}_{\mathbf{p}}$  is not a distribution on the set of finite trees, but on the set of all trees (leading to some difficulties, since the weight of the set of infinite trees is the probability of non extinction, which is  $> 0$  (but the probability of each infinite tree is 0)).

### 3.8 From random walk to GW trees

Consider a random walk with an infinite length :  $S(0) = 0$  and for any  $k \geq 1$ ,  $S(k) = \sum_{i=1}^k (Y_i - 1)$  where the  $Y_i$  are i.i.d. r.v. with distribution  $\mathbf{p}$ . Let

$$\text{Hit}_{-1}(S) = \inf\{k \mid S(k) = -1\}$$

the first hitting time of  $-1$  by  $S$ . Consider  $\bar{S} = (S(0), \dots, S(\text{Hit}_{-1}))$  the random walk killed at time  $\text{Hit}_{-1}$ .

For any path  $s$  in  $\mathbf{L}_n$  (for some  $n$ ), we have

$$\mathbb{P}_{\mathbf{p}}(\bar{S} = s) = \prod_{i=1}^n \mathbb{P}_{\mathbf{p}}(Y_i - 1 = s_i - s_{i-1}, i \in \llbracket 1, n \rrbracket) = \prod_{i=1}^n p_{s_i - s_{i-1} + 1}.$$

This distribution is the same as  $\Gamma_{lex}(\mathbf{T})$  (see Proposition 2) for  $\mathbf{T}$  the family tree of a GW process having OD  $\mathbf{p}$ .

It follows that

**Proposition 4** *If  $\mathbf{T}$  is a GW tree with OD  $\mathbf{p}$ , then the trajectory  $\mathbf{S} := \Gamma_{lex}(\mathbf{T})$  is distributed as a random walk with increment  $Y_i - 1$  (i.i.d. where  $Y_i$  has distribution  $\mathbf{p}$ ) killed at its first hitting time of  $-1$ . For any  $n \geq 1$ ,*

$$\mathbb{P}_{\mathbf{p}}(\text{Hit}_{-1} = n) = \mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n).$$

To compute the value of  $\mathbb{P}(\text{Hit}_{-1} = n)$ , we use the following lemma :

**Lemma 6 The rotation principle** (Dvoretzky & Motzkin) For  $S$  defined as above, we have for  $k \geq 1$ ,

$$\mathbb{P}_{\mathbf{p}}(\text{Hit}_{-1} = k) = \frac{1}{k} \mathbb{P}(S_k = -1).$$

This result needs that the increments have their support in  $\{-1, 0, 1, \dots\}$ .

**Proof .** In the proof, all the considered paths have increments in  $\{-1, 0, 1, 2, \dots\}$ . The proof we provide is essentially combinatorial. Recall the definition of  $\mathbf{L}_k$  and let

$$\mathbf{B}_k = \{(s_0, \dots, s_k), s_k = -1\}$$

be the set of walk ending at  $-1$ . We will associate with each trajectory  $t$  of  $\mathbf{L}_k$ , a set of  $k$  trajectories  $\Phi(t) = \{t_1, \dots, t_k\}$  included in  $\mathbf{B}_k$ .

First, note that  $\mathbf{L}_k$  is a subset of  $\mathbf{B}_k$ .

Let  $t$  in  $\mathbf{B}_k$ ; let  $\Delta_i t = t_i - t_{i-1}$  for  $i$  from 1 to  $k$  the  $k$  increments of  $t$ . For any  $j \in \{0, \dots, k-1\}$ , denote by  $t^{(j)}$  the trajectory having as increments  $\Delta_j t, \dots, \Delta_k t, \Delta_1 t, \dots, \Delta_{j-1} t$  (geometrically, cut  $t$  at position  $j$ , and exchange the left and right part of the trajectory as done on Fig. 18).

– Start with  $t \in \mathbf{L}_k$ . All the trajectories  $t^{(j)}$ ,  $j \in \{0, \dots, k-1\}$  are different and only one of them is in  $\mathbf{L}_k$ : the reason is that  $t^{(j)}$  reaches its minimum for the first time at time  $k-j$ . Hence,  $t$  is the only element in  $\Phi(t)$  being also in  $\mathbf{L}_k$ .

– take an element on  $t \in \mathbf{B}_k$ : then there exists an only pair  $(t', j)$  such that  $t = (t')^{(j)}$  and  $t'$  is in  $\mathbf{L}_k$ . Indeed, assume that  $t$  reaches its minimum for the first time at position  $j$ ; then  $t^{(j)}$  is in  $\mathbf{L}_k$  (and  $t^{(j')} \in \mathbf{L}_k$  iff  $j' = j$ );

Hence each rotation class  $(t^{(j)}, j = 0, \dots, k-1)$  for  $t \in \mathbf{L}_k$  are distinct, have cardinality  $k$ , and  $\cup_{t \in \mathbf{L}_k} \Phi(t) = \mathbf{B}_k$  (that is each element of  $\mathbf{B}_k$  is in a rotation class).

Now, in a rotation class, all the trajectories have the same weight under  $\mathbb{P}_{\mathbf{p}}$ , since they have the same increments (in a different order). Hence, to each trajectory  $t$  from  $\mathbf{L}_k$  with probability  $\mathbb{P}_{\mathbf{p}}(t)$ , we have associated a set of  $k$  trajectories  $t^{(0)}, \dots, t^{(k-1)}$  from  $\mathbf{B}_k$  having the same weight. It remains just to sum on all these trajectories to end the proof.  $\square$

**Note 4 :** A direct consequence of the arguments developed in the proof is

$$\#\mathbf{L}_k = \frac{\#\mathbf{B}_k}{k}.$$

Consider  $h := \gcd\{i \mid \mathbb{P}_{\mathbf{p}}(Y_1 = i) > 0 \text{ and } i > 0\}$ , and the set of possible trees size under  $\mathbb{P}_{\mathbf{p}}$  :

$$\text{PTS}_{\mathbf{p}} := \{n \mid \mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n) > 0\}.$$

If  $h = 1$  it is not difficult to see that there exists  $N$  such that  $m \in \text{PTS}_{\mathbf{p}}$  for any  $m \geq N$ ; if  $h > 1$ , it is not the case : an infinite number of integers are lacking in  $\text{PTS}_{\mathbf{p}}$  (think to the example of binary trees: the size is odd,  $\text{PTS}_{\mathbf{p}} = 1 + 2\mathbb{N}$ ).

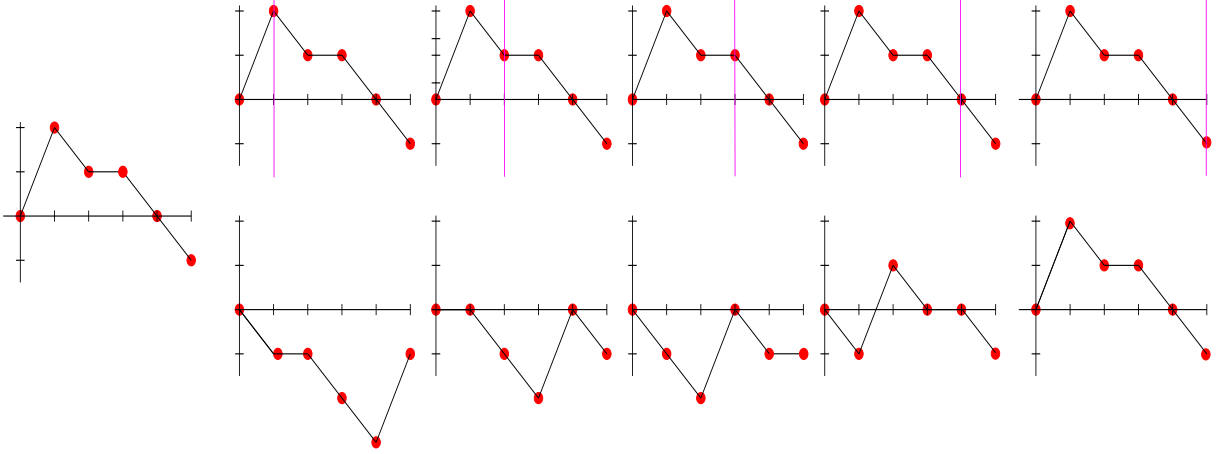


Figure 18: A path  $t$  from  $EX(5)$ , and its 5 conjugates  $t^{(1)}, \dots, t^{(5)}$  on the line below.

**Lemma 7** Let  $\mathbf{p}$  be a critical OD (with mean 1 and variance  $\sigma_{\mathbf{p}}^2 \in (0, +\infty)$ ). For  $n$  in  $\text{PTS}_{\mathbf{p}}$ ,

$$\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n) \sim \frac{hn^{-3/2}}{\sqrt{2\pi}\sigma_{\mathbf{p}}}. \quad (9)$$

This means that if  $(x_n, n \geq 1)$  is a sequence of elements of  $\text{PTS}_{\mathbf{p}}$  going to  $+\infty$ , then  $\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = x_n) \sim \frac{hx_n^{-3/2}}{\sqrt{2\pi}\sigma_{\mathbf{p}}}$ .

This is a consequence of Proposition 4, Lemma 6, and of a local version of the central limit theorem, here recalled (or given) :

**Theorem 4** Local central limit theorem (see Port, *Theoretical probability for applications*). Let  $X_1, X_2, \dots$  be some i.i.d. r.v. with mean  $m$ , and variance  $\sigma^2 \in (0, +\infty)$ . Let  $S_n = X_1 + \dots + X_n$ . Assume that the distribution  $X_1$  has its support in the set  $\{a + hk \mid k \in \mathbb{Z}\}$ , where  $h$  is maximal. Then, when  $n \rightarrow +\infty$

$$\frac{\sqrt{n}}{h} \mathbb{P}(S_n = an + hk) - \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(hk + an - nm)^2}{2\sigma^2 n}\right) \rightarrow 0 \text{ uniformly on } k.$$

Here  $m = 0$ ; one then has  $\sqrt{n}\mathbb{P}(S_n = -1) - h(\sqrt{2\pi}\sigma_{\mathbf{p}})^{-1}e^{-1/(2\sigma_{\mathbf{p}}^2 n)}$  tends to 0, and then  $\mathbb{P}(S_n = -1) \sim h(\sqrt{2\pi n}\sigma_{\mathbf{p}})^{-1}$ .

In the sequel we will restrict ourselves to the case where  $h = 1$ . The case  $h > 1$ , that arises in important applications as in binary trees, can be treated with taking care of  $\text{PTS}_{\mathbf{p}}$  and by introducing  $h$  where it is needed.

### GW tree under the conditioning $\{|\mathbf{T}| = n\}$

Let  $\mathbf{p} = (p_k)_{k \geq 0}$  be an OD. Let  $(y_1, \dots, y_n) \in \mathbf{D}_n$  (see page 20) be the description of a tree with  $n$  node. A natural question is the following :



Conditionally on  $\{|\mathbf{T}| = n\}$ , what is the probability that the description of a GW with offspring distribution  $p$  is  $(y_i)$ ?

Without any conditioning we have :

$$\mathbb{P}_{\mathbf{p}}((Y_1, \dots, Y_n) = (y_1, \dots, y_n)) = \prod_{i=1}^n p_{y_i}. \quad (10)$$

The conditioning does not change too much the things: the probability being proportional to that before conditioning :

$$\begin{aligned} \mathbb{P}_{\mathbf{p}}((Y_1, \dots, Y_n) = (y_1, \dots, y_n) | |\mathbf{T}| = n) &= \frac{\mathbb{P}_{\mathbf{p}}((Y_1, \dots, Y_n) = (y_1, \dots, y_n))}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)} \\ &= \frac{\prod_{i=1}^n p_{y_i}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)} \end{aligned} \quad (11)$$

since  $(y_i, i = 1, \dots, n)$  is the description of a size  $n$  tree.

### 3.9 Example of Galton-Watson tree conditioned by the size

#### Binary tree

A binary GW tree is the family tree of a GW process when the OD is  $p_2 = \alpha$ ,  $p_0 = 1 - \alpha$  (with  $0 < \alpha < 1$ ). Of course, a.s. the family tree of this GW process is a binary tree. Let us condition by  $\{|\mathbf{T}| = 2n + 1\}$ .

Take a tree  $t$  with  $2n + 1$  nodes: it has  $n + 1$  leaves (nodes with 0 children) and  $n$  internal nodes (nodes with 2 children); hence

$$\mathbb{P}_{\mathbf{p}}(\mathbf{T} = t \mid |\mathbf{T}| = 2n + 1) = \frac{p_2^{N_2(t)} p_0^{N_0(t)}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = 2n + 1)} = \frac{\alpha^n (1 - \alpha)^{n+1}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = 2n + 1)}.$$

This weight does not depend on the detail of the tree: all binary trees with  $2n + 1$  nodes have the same weight! This is the uniform distribution on the set of binary trees with  $2n + 1$  nodes (of course, we used also that all binary trees, and only them, have a positive weight). Notice that what we said does not depend on the choice of  $\alpha$ .

**Exercise 8 :** How many binary trees with  $2n + 1$  nodes? Ternary trees?

#### Plane trees (also called geometrical trees)

Assume that  $\mathbf{p}$  is the geometrical distribution with parameter  $\alpha$ , for an  $\alpha \in (0, 1)$ . That is  $p_i = (1 - \alpha)^i \alpha$  for any  $i \geq 0$ . Take  $t$  a planar tree with  $n$  nodes ( $t \in \mathcal{T}_n$ ). We have

$$\mathbb{P}_{\mathbf{p}}(\mathbf{T} = t \mid |\mathbf{T}| = n) = \frac{\prod_{u \in t} p_{c_u(t)}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)} = \frac{(1 - \alpha)^{\sum_{u \in t} c_u(t)} \alpha^{|t|}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)} = \frac{(1 - \alpha)^{n-1} \alpha^n}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)}$$

since each node, except the root, is the child of another node; again, the probability is the same for all tree having a weight. But all trees in  $\mathcal{T}_n$  has a positive weight, that is here, all trees in  $\mathcal{T}_n$ : hence the conditional distribution  $\mathbb{P}_{\mathbf{p}}(\cdot | \mathbf{T} = n)$  is the uniform distribution on  $\mathcal{T}_n$ , and this whatever the choice of  $\alpha$ .

### Cayley trees

(Remark due to Spencer (97)) A Cayley tree with  $n$  nodes is a connected graph with set of vertices  $\{1, 2, \dots, n\}$  with no cycle. The vertex 1 is chosen to be the root. According to Cayley (1889) there are  $n^{n-2}$  Cayley trees (see also Section 3.11). We are willing to talk about the shape of Cayley trees. Cayley trees are not given with an order between the brothers of a given nodes

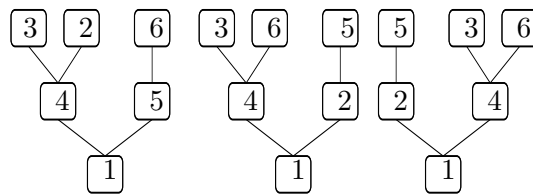


Figure 19: Three Cayley trees.

(recall that a graph is just a set of vertices, and a set of pairs of vertices, called edges). Then, a priori, the depth first order is not well defined on Cayley trees. But using the labels of the nodes, there is a simple canonical ordering of the children of the nodes : first 1 is the ancestor. Then order the children of 1, according to the standard order in  $\mathbb{N}$ , and do the same thing for each node in the tree. This provides a canonical representation of the nodes of Cayley trees (one can represent Cayley trees on the plane by respecting this order). On the above picture, the two first Cayley trees are different, and the two trees 2 and 3 are the same tree. The third tree has its labels sorted, and is then the canonical representative. When dealing with canonical representation, a Cayley tree is defined by a  $n$ -tuples of disjoint sets  $(A_1, A_2, \dots, A_n)$  where  $\cup_{i \geq 1} A_i = \{2, \dots, n\}$ . To build the tree, use an order, for example the breadth first order. Gives to the  $i$ th node visited  $\#A_i$  children with labels the element of  $A_i$ , sorted according to the increasing order (to children of the root are the vertices with label in  $A_1$ , ...). For example, the two trees 2 and 3 on the figure, are defined by :  $(\{2, 4\}, \{5\}, \{3, 6\}, \emptyset, \emptyset, \emptyset)$ . This is deeply related to the BFQ considerations explained above in the course. For a  $n$ -tuple  $(A_1, A_2, \dots, A_n)$  coming from a tree, the sequence  $(y_i, i = 1..n)$  satisfying  $y_i = \#A_i$  in in  $\mathbf{D}_n$ .

The question is : given a sequence  $(y_1, \dots, y_n)$  in  $\mathbf{D}_n$ , how many  $n$ -tuples  $(A_1, \dots, A_n)$  corresponding to a Cayley tree exist ? (we mean, such that  $(\#A_1, \dots, \#A_n) = (y_1, \dots, y_n)$ ). In other words, how many Cayley tree with a canonical description given by the  $y$ 's ?

We have 1 choice for the root label (it is 1, by convention), for the labels in  $A_1$  we have  $\binom{n-1}{y_1}$  choices, for  $A_2$  we have  $\binom{n-1-y_1}{y_2}$  choices... finally, the number of Cayley tree with a prescribed

vector  $(\#A_1, \dots, \#A_n) = (y_1, \dots, y_n)$  is :

$$\binom{n-1}{y_1} \prod_{k=2}^n \binom{n-1-(y_1+\dots+y_{k-1})}{y_k} = \frac{(n-1)!}{y_1! \cdots y_n!};$$

the probability to observe this shape, when the set of Cayley tree is equipped with the uniform distribution is then proportional to this number.

Take now a GW process with OD  $\mathbf{p}$ , the Poisson distribution with parameter  $\lambda$  (that is  $p_i = \frac{\lambda^i e^{-\lambda}}{i!}$ ). We have

$$\begin{aligned} \mathbb{P}_{\mathbf{p}}(D(\mathbf{T}) = (y_1, \dots, y_n) \mid |\mathbf{T}| = n) &= \frac{\prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)} \\ &= \frac{e^{-\lambda n} \lambda^{n-1}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)} \prod_{i=1}^n \frac{1}{y_i!}. \end{aligned}$$

Hence, the two distribution are proportional, then equal, since they give a positive weight to the same trees. Therefore the shape of Cayley tree is distributed as a GW tree with OD the Poisson distribution, conditioned to have  $n$  nodes.

**Exercise 9 :** 1) A unary-binary tree is a tree in which the nodes have 0, 1, or 2 nodes. Build an OD  $\mathbf{p}$  for which  $\mathbb{P}_{\mathbf{p}}(\cdot \mid |\mathbf{T}| = n)$  is the uniform law on the set of unary-binary trees with  $n$  nodes.

2) Same question, if the set of allowed degrees is given by a subset  $I$  include in  $\mathbb{N}$  (containing  $\{0\}$ ).

### 3.10 An exponential family and a critical parameter

We saw that according whether  $m = \sum ip_i > 1$  or  $m \leq 1$  a GW process with distribution  $\mathbf{p}$  eventually a.s dies out or not. Above, we saw also that under the conditioning by  $|\mathbf{T}| = n$ , the distribution of binary trees – and the same happened just after for geometrical trees, or Poisson trees – does not depend on the parameter  $\alpha$  which entered in the definition of the OD, and which tuning of course, change the mean number of children of a given node (when no conditioning takes place). This phenomenon occurs similarly for any OD, as been shown by Kennedy (in his paper The Galton-Watson process conditioned on the total progeny).

More precisely, consider  $\mathbf{p} = (p_k)_{k \geq 0}$  an OD as usual, and consider

$$f(x) = \mathbb{E}_{\mathbf{p}}(x^Y) = \sum_{k \geq 0} p_k x^k$$

its probability generating function. Now, consider  $\tilde{\mathbf{p}} := (\tilde{p}_k)_{k \geq 0}$  defined by

$$\tilde{p}_k \stackrel{def}{=} \frac{a^k p_k}{f(a)},$$

for a certain  $a > 0$  fixed and smaller than the radius of convergence of  $f$  (which is a power series, with radius 1, at least). We have

$$\mathbb{P}_{\mathbf{p}}(Y_1 = y_1, \dots, Y_n = y_n \mid |\mathbf{T}| = n) = \mathbb{P}_{\tilde{\mathbf{p}}}(Y_1 = y_1, \dots, Y_n = y_n \mid |\mathbf{T}| = n);$$

that is, the distributions on  $\mathcal{T}_n$  of family trees of GW processes with OD  $\mathbf{p}$  and  $\tilde{\mathbf{p}}$  under the conditioning  $\{|\mathbf{T}| = n\}$  coincide.

Indeed,

$$\begin{aligned} \mathbb{P}_{\tilde{\mathbf{p}}}(Y_1 = y_1, \dots, Y_n = y_n \mid |\mathbf{T}| = n) &= \frac{\prod_{k=1}^n a^{y_k} p_{y_k}}{f(a)^n \mathbb{P}_{\tilde{\mathbf{p}}}(|\mathbf{T}| = n)} \\ &= \frac{\prod_{k=1}^n p_{y_k}}{\mathbb{P}_{\tilde{\mathbf{p}}}(|\mathbf{T}| = n) f(a)^n a^{-n+1}} \end{aligned}$$

which is indeed proportional to

$$\mathbb{P}_{\mathbf{p}}(Y_1 = y_1, \dots, Y_n = y_n \mid |\mathbf{T}| = n) = \frac{\prod_{k=1}^n p_{y_k}}{\mathbb{P}_{\mathbf{p}}(|\mathbf{T}| = n)}$$

These probability having clearly the same support, and being proportional, are equal. (One can pass from a Bernoulli distribution to another one, or a geometrical distribution to another one, or a Poisson... by the choice of right  $a$  in what is done above).

The value of  $a$  (when it exists) such that  $af'(a) = f(a)$  provides the so-called *critical case*, that is the case for which  $\mathbb{E}_{\tilde{\mathbf{p}}}(Y) = \tilde{m} = 1$ . The corresponding variance  $\tilde{\sigma}^2 = \mathbb{E}_{\tilde{\mathbf{p}}}(Y^2) - 1$  of the number of children of a given node under  $\tilde{\mathbf{p}}$  can be expressed in terms of  $f$  :

$$\tilde{\sigma}^2 = \frac{a^2 f''(a)}{f(a)}.$$

It will play a crucial role, later on.

In the sequel, we will suppose that the OD is critical: it has mean 1 and its variance  $\sigma_{\mathbf{p}}^2 \in (0, +\infty)$

**Exercise 10 :** Find an exponential family which does not contain any critical distribution. (Under such distributions, tree behave asymptotically radically differently than those under critical distributions: even the order of their height is not the same; the study of these trees is principally due to Le Gall & Duquesne)

### 3.11 Problem: Analysis of hashing with linear probing

Hashing with linear probing is an algorithm used to store and retrieved data in a computer. The data belong to a certain set  $E$ . A function  $h$  from  $E$  to  $\llbracket 0, n-1 \rrbracket$  is then used to place some data  $x_1, \dots, x_k$  from  $E$  (with  $k \leq n$ ) in an array with  $n$  cases, labeled  $0, 1, \dots, n-1$ .

The data are inserted successively, as follows :

- 1) data  $x_i$  **chooses** the place  $h(x_i)$ ,
- 2)  $x_i$  is stored in the first empty case among  $h(x_i), h(x_i) + 1 \bmod n, h(x_i) + 2 \bmod n$ , etc. If  $x_i$  is stored in  $h(x_i) + j \bmod n$  (with  $j$  taken in  $\llbracket 0, n-1 \rrbracket$ ) we say that  $x_i$  probed the  $j+1$  cases  $h(x_i), \dots, h(x_i) + j \bmod n$ . The cost of insertion of  $x_i$  is  $j+1$  (definition).

[This problem appears also in the literature under the "parking allegory"; some cars try to park on a circular parking with  $n$  places: car  $x_k$  (or  $k$ ) chooses a place, say  $j$ , and then park on the first empty place starting from  $j$ .]

In order to analyze the cost of insertion (in term of displacement of the items), the following hypothesis is made on the successive values  $h(x_1), h(x_2)$ , etc of choices: they are assume to be i.i.d. uniform on the set  $\llbracket 1, n \rrbracket$ . (Of course, as explained in the book of Knuth, from a practical point of view, the construction of an efficient hashing function – which doesn't send all the data on the same case! – is quite difficult.)

We propose in this problem to analyze this algorithm;<sup>3</sup> the asymptotic cost will follows some considerations of the last section of this course.

Denote by  $A_i(k) = \{j \mid h(x_i) = j, \text{ for } i = 1, \dots, k\}$ , the number of data that have chosen case  $k$ .

- 1) For a fixed  $k$  smaller than  $n$ , give the joint law of  $(\#A_0(k), \dots, \#A_{n-1}(k))$ .
- 2) For  $m \in \llbracket 0, m-1 \rrbracket$ , let  $Y_m$  be the number of data that probed case  $m$ .
  - Why  $Y_m = \#A_m + (Y_{m-1} - 1)_+$ ? (for  $m = 0$ ,  $Y_0 = \#A_0 + (Y_{n-1} - 1)_+$ )
- 3) Assume now that  $k = n - 1$  (that is  $n - 1$  data stored in  $n$  cases);
  - Give a set of necessary and sufficient condition on  $(\#A_1, \dots, \#A_n)$  in order than the last case of the array stays empty at the end.
  - Simplify the description of the process  $Y$  given in (2), in this case.
  - What is the distribution of the process  $Y$ ?
- 4) Express the total cost of insertion of  $n - 1$  items in terms of the process  $Y$ ?
- 5) In the case where  $n - 1$  items are inserted, what is the importance of the first assumption: "the last case is empty"?
- 6) How change the total cost, if a last data  $h(x_n)$  is inserted?

---

<sup>3</sup>The analysis of hashing with linear probing has been done at first by Flajolet, Viola and Pobleto. Here, as an exercise, we give another analysis, published later on in collaboration with Chassaing (Parking functions, empirical processes and the width of rooted labeled trees).

## 4 Asymptotic comparison between the height process and DFQ

Above in the course, the DFQ of a GW tree with  $n$  nodes appeared to be a random walk conditioned to hit  $-1$  at time  $n$ . We will see here that some properties valid for non conditioned random walks stay true for conditioned one (see Figure 20). We will deduce from this some elements on trees with  $n$  nodes, but what is much important for the aim of this course, we will deduce the asymptotic “equality” of the two processes: height process and DFQ.

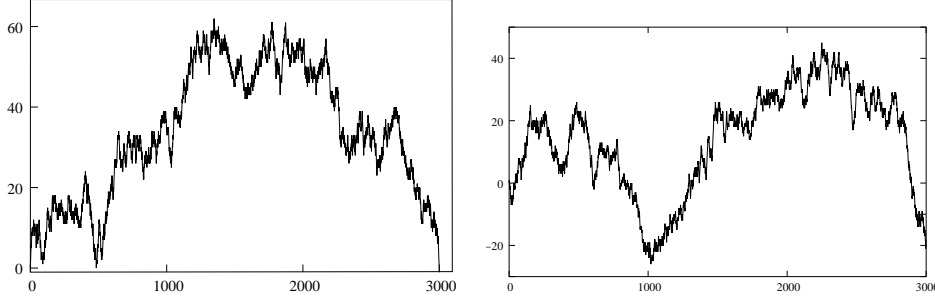


Figure 20: *Two random walks with increments  $\pm 1$  with probability  $1/2$ ; the first is conditioned by  $\text{Hit}_{-1} = 3001$*

### 4.1 Concentration principle : from random walks to Lukaciewish paths

Let  $S_0 = 0$ ,  $S_j = \sum_{i=1}^j Y_i - 1$  for  $j \geq 1$ , be a random walk where the  $Y_i$  are i.i.d. and are  $\mathbf{p}$  distributed. We assume from now on that  $\mathbb{E}(Y_i - 1) = 0$  and  $\sigma_{\mathbf{p}}^2 = \text{var}(Y_i - 1) = \text{var}(Y_i) \in (0, +\infty)$ .

Let  $\mathcal{W}_n$  be the set of all paths of size  $n$  (this is  $\mathbb{R}^n$  equipped with the Borelian  $\sigma$ -field); and let  $A_n$  be a subset of  $\mathcal{W}_n$  (a Borelian); denote by  $\mathbb{P}_n$  the distribution of size  $n$  random walk. We have

$$\mathbb{P}_n(S \in A_n \mid S \in \mathbf{L}_n) = \frac{\mathbb{P}_n((S \in A_n) \cap (S \in \mathbf{L}_n))}{\mathbb{P}_n(S \in \mathbf{L}_n)} \leq \frac{\mathbb{P}_n(S \in A_n)}{\mathbb{P}_n(S \in \mathbf{L}_n)}.$$

Hence since by Lemma 6 and Lemma 7 (when  $h = 1$ ),

$$\mathbb{P}_n(S \in \mathbf{L}_n) = \frac{1}{n} \mathbb{P}(S_n = -1) \sim \frac{n^{-3/2}}{\sqrt{2\pi\sigma_{\mathbf{p}}}},$$

we get

$$\mathbb{P}_n(S \in A_n \mid S \in \mathbf{L}_n) = O(n^{3/2}) \mathbb{P}_n(S \in A_n); \tag{12}$$

the right hand side is simple, since it deals with non conditioned random walks. Then, immediately

**Corollary 1** *For any  $c > 0$  and  $\beta > 0$*

$$\mathbb{P}_n(S \in A_n) = o(n^{-\beta}) \implies \mathbb{P}_n(S \in A_n \mid S \in \mathbf{L}_n) = o(n^{-\beta+3/2}).$$

$$\mathbb{P}_n(S \in A_n) = o(\exp(-cn^\beta)) \implies \mathbb{P}_n(S \in A_n \mid S \in \mathbf{L}_n) = o(\exp(-c'n^\beta)).$$

In other words, very rare events for random walk are still very rare even on Lukaciewish paths. (An event can be true with proba.  $1/2$  (or much more) on random walk and have proba  $0$  (or close to  $0$ ) on Lukaciewish paths. Can you find examples?).

To use this concentration principle, we focus on rare events that describe the wanted phenomenon (for example, instead of saying that something converges to  $1$ , we say that “the probability it stay far from  $1$ ” is very small).

## 4.2 The theorem of Petrov

We refer to Petrov (Sums of independent random variables) for the (not difficult) proof of the following theorem and for much more!

**Lemma 8** (Petrov) *Let  $z_1, z_2, \dots$ , be i.i.d. r.v. with mean  $0$ , such that*

$$\mathbb{E}(e^{tz_1}) < +\infty \text{ for } |t| \leq H,$$

*for a constant  $H > 0$ . Let  $Z_n = z_1 + \dots + z_n$ . Then there exists constant  $g > 0$  and  $T > 0$  s.t.*

$$\mathbb{P}(|Z_n| \geq x) \leq 2 \exp\left(-\frac{x^2}{2gn}\right) \quad \text{for } |x| \leq gnT \quad (13)$$

$$\mathbb{P}(|Z_n| \geq x) \leq 2 \exp(-Tx/2) \quad \text{for } |x| \geq gnT \quad (14)$$

This Lemma is admitted. ( valid constants  $g, T$  can be computed). (the proof is simple, but a bit long; it is an application of the Markov inequality  $\mathbb{P}(X \geq a) \leq a^{-1}\mathbb{E}(X)$  valid for non negative r.v.; for  $X$  positive or not  $\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta})$  for  $t > 0$ ; now  $e^{tX}$  is a positive random variable; by Markov,  $\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq \mathbb{E}(e^{tX})/e^{ta}$ . Now take  $X = Z_n$ , use that  $\mathbb{E}(e^{tZ_n}) = \mathbb{E}(e^{tz})^n$  An ingredient used is the following: if  $X$  is a centered r.v. such that, for some  $H > 0$ ,  $\mathbb{E}(e^{tX}) < +\infty$  for  $|t| \leq H$ , then there exists  $g > 0, T > 0$  s.t.  $\mathbb{E}(e^{tX}) \leq e^{gt^2/2}$  for  $|t| \leq T$ . To end, an optimization on  $t$  is done).

Notice that the Bienaymé-Tchebichev theorem says that

$$\mathbb{P}(|Z_n| \geq x) \leq \text{var}(Z_n)/x^2 = n \text{var}(z_1)/x^2.$$

Petrov bound are much better, since the bound are exponentially small in  $x$ . Using Petrov bounds one can prove

**Lemma 9** *Let  $n \in \mathbb{N}^*$  and let  $Z_n = z_1 + \dots + z_n$  satisfying the hypothesis of the previous Lemma.*

• *For  $\nu \in ]0, \frac{1}{2}[$ , there exists  $c > 0$ ,*

$$\mathbb{P}(|Z_n| \geq n^{1/2+\nu}) = O\left(\exp\left(-\frac{n^{2\nu}}{2g}\right)\right) \quad (15)$$

• *For  $c > 0$ ,*

$$\mathbb{P}(|Z_n| \geq c\sqrt{n \ln n}) = O\left(n^{-c^2/2g}\right) \quad (16)$$

**Proof .** Formula (13) says that  $\mathbb{P}(|Z_n| \geq n^{1/2+\nu}) \leq 2 \exp(-n^{2\nu}/(2g))$  for  $n$  large enough. Same proof for the second formula.  $\square$

### 4.3 Application : How many nodes with degree $k$ in a GW tree with $n$ nodes?

The application  $\Gamma_{lex}$  sends a tree with  $n_l$  nodes of degree  $l$  to a Lukaciewish path with  $n_l$  increments  $l - 1$ . If the OD is  $\mathbf{p} = (p_i)_{i \geq 0}$ , then the increments have distributions  $\mu := (\mu_j)_{j \geq -1}$ . Let  $S$  be a r.w. with i.i.d. increments with law  $\mu$ . Let

$$M_l(n) = \sum_{k=1}^n \mathbb{I}_{S_k - S_{k-1} = l-1},$$

be the number of increments on  $S$  equal to  $l - 1$ . Since the increments are i.i.d.  $M_l(n)$  is a random walk with i.i.d increments with distribution Bernoulli( $\mu_{l-1}$ ), then

$$M_l(n) - \mu_{l-1}n = \sum_{k=1}^n (\mathbb{I}_{S_k - S_{k-1} = l-1} - \mu_{l-1}),$$

is a sum of centered increment, having some exponential moments in a neighborhood of zero (exercise). By Petrov (Lemma 9) for  $\nu \in (0, 1/2)$ , for some  $g > 0$

$$\mathbb{P}(|M_l(n) - \mu_{l-1}n| \geq n^{1/2+\nu}) = O(e^{-n^{2\nu}/2g}),$$

By Corollary 1

$$\mathbb{P}(|M_l(n) - \mu_{l-1}n| \geq n^{1/2+\nu} \mid S \in \mathbf{L}_n) = O(n^{3/2}e^{-n^{2\nu}/2g}) = O(e^{-cn^{2\nu}}). \quad (17)$$

for some  $c > 0$ . Hence, in a Lukaciewish path, with a huge probability  $M_l(n)$  is close to  $\mu_{l-1}n = p_l n$ , and then of course, this can be translated in term of nodes of degree  $l$  in a GW with OD  $\mathbf{p}$  and  $n$  nodes. One can show also by the same method, using (16)

$$\mathbb{P}(|M_l(n) - \mu_{l-1}n| \geq d\sqrt{n \log n}) = O(n^{-d^2/2g})$$

$$\text{and then } \mathbb{P}(|M_l(n) - p_l n| \geq d\sqrt{n \log n} \mid S \in \mathbf{L}_n) = O(n^{3/2}n^{-d^2/(2g)}) = O(n^{3/2-d^2/(2g)}). \quad (18)$$

Hence, the probability that  $|N_l(n) - p_l n| \geq \sqrt{cn \log n}$  goes to 0, if  $c$  is chosen large enough.

### 4.4 Asymptotic comparison between the height process and the DFQ

From now on, we assume that a variable  $w$  having distribution  $\mathbf{p}$ , is centered and has some exponential moments in a neighborhood of zero :

$$(A) : \exists a > 0 \text{ s.t. } \mathbb{E}(e^{a|w|}) < +\infty.$$



### Relation between the height process and the DFQ

Let  $T$  be a tree in  $\mathcal{T}_n$ ,  $S_T$  its DFQ and  $H_T$  its height process. In Lemma 4 we saw that

$$RM_T(l) = \left\{ j \mid 0 \leq j < l, \min_{j \leq k \leq l} S_T(k) = S_T(j) \right\}. \quad (19)$$

then

$$H_T(l) = \#RM_T(l), \text{ for } l \geq 1.$$

Our aim now is to show that

**Theorem 5** *Assume (A). There exists a constant  $c > 0$  such that for any  $\nu \geq 0$*

$$\mathbb{P}_{\mathbf{p}} \left( \sup_{0 \leq l \leq n} \left| S_{\mathbf{T}}(l) - H_{\mathbf{T}}(l) \frac{\sigma_{\mathbf{p}}^2}{2} \right| \geq n^{1/4+\nu} \mid |\mathbf{T}| = n \right) = O(\exp(-cn^\nu)). \quad \square$$

Since  $S_{\mathbf{T}}(l)$  will appear to have order  $n^{1/2}$  for most  $l \in \llbracket 0, n \rrbracket$ , this result indeed says that  $H_{\mathbf{T}}(l) \frac{\sigma_{\mathbf{p}}^2}{2}$  is close to  $S_{\mathbf{T}}(l)$ .

### Passage via non-conditioned random walk

Consider a random walk  $W$  with increment a sequence of i.i.d. r.v.  $(w_i, i \geq 1)$ , with distribution  $\mu = (\mu_j)_{j \geq -1} = (p_{j+1})_{j \geq -1}$ . From what it said in the previous subsection, this random walk, conditioned to hit  $-1$  at first at time  $n$  has the same law as the DFQ  $S_{\mathbf{T}}$  of a GW tree  $\mathbf{T}$  with OD  $\mathbf{p}$  conditioned to have size  $n$  :

$$\mathbb{P}(W \in A \mid \text{Hit}_{-1}(W) = n) = \mathbb{P}_{\mathbf{p}}(S_{\mathbf{T}} \in A \mid |\mathbf{T}| = n).$$

In other words,  $W$  is the unconditioned version of  $S_{\mathbf{T}}$ . The set of right minima of  $W$  on  $\llbracket 0, l \rrbracket$  is:

$$RM(l) = \left\{ j \mid 0 \leq j < l, \min_{j \leq k \leq l} W(k) = W(j) \right\}.$$

In probability we prefer variable that depends on the past of the random walk, instead of the future. Since  $j$  belongs to  $RM(l)$  if the walk behaves properly between time  $j$  and  $l$ , we prefer to turn over the walk.

### Turning back of $W$

We associate with the path  $(W(j))_{j \in \llbracket 0, l \rrbracket}$ , the path  $(W^\bullet(j))_{j \in \llbracket 0, l \rrbracket}$  defined by :

$$W^\bullet(j) = W(l) - W(l - j) \text{ for every } j \in \llbracket 0, l \rrbracket. \quad (20)$$

The graph of  $W^\bullet$  is obtained by taking the symmetric of that of  $W$  with respect to the center  $(l/2, W(l)/2)$ . It is easy to note that the increments of  $(W^\bullet(i))_{i \geq 0}$  have the same distribution of

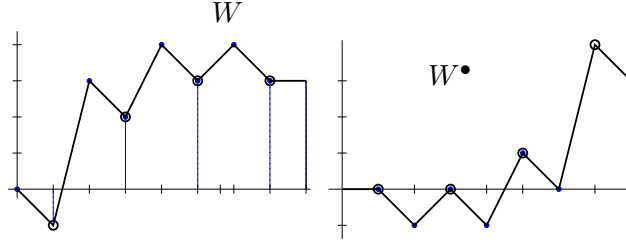


Figure 21: : Construction of  $W^\bullet$  from  $W$ . The right minima of  $W$  are marked on  $W$ , the record of  $W^\bullet$  on the graph of  $W^\bullet$ .

those of  $W$ , namely  $(\mu_j)_{j \geq -1}$ , and are independent. Let

$$R^\bullet(l) = \{j \mid 1 \leq j \leq l, \max_{0 \leq k \leq j} W^\bullet(k) = W^\bullet(j)\},$$

be the (weak) records time of the path  $(W^\bullet(j))_{j \in \llbracket 0, l \rrbracket}$ . We have immediately by (20) :

$$\#R^\bullet(l) = \#RM(l), \quad (21)$$

$$\max_{0 \leq i \leq l} W^\bullet(i) = W(l) - \min_{0 \leq i \leq l} W(i). \quad (22)$$

$$\left| \max_{0 \leq j \leq l} \{W^\bullet(j)\} - \#R^\bullet(l) \frac{\sigma_{\mathbf{p}}^2}{2} \right| = \left| W(l) - \min_{0 \leq i \leq l} \{W(i)\} - \#RM(l) \frac{\sigma_{\mathbf{p}}^2}{2} \right|. \quad (23)$$

Hence, instead of working with right minima, we are now dealing with records. That's more easy!

#### 4.5 Study of the number of records of $W^\bullet$ : ladder variables

Let  $(w(i), i \geq 1)$  be the i.i.d. increments of  $W^\bullet$  :

$$W^\bullet(0) = 0, \quad W^\bullet(k) = \sum_{j=1}^k w(j), \quad \text{for } k \geq 1;$$

recall that  $\mathbb{P}(w(1) = i) = \mu_i$  for  $i \geq -1$ , that  $\mathbb{E}(w(1)) = 0$ , and  $\text{var}(w(1)) = \sigma_{\mathbf{p}}^2$ .

The weak records time of  $W^\bullet$  increase are the  $0 = \tau_0 < \tau_1 < \tau_2 < \dots$  defined by :

$$\tau_k = \inf \{j > \tau_{k-1}, W^\bullet(j) \geq W^\bullet(\tau_{k-1})\}, \quad \text{for } k \geq 1. \quad (24)$$

**Lemma 10** For any  $k \in \mathbb{N}$ ,  $\mathbb{P}(W^\bullet(\tau_1) = k) = \mathbb{P}(w(1) \geq k)$ . Moreover

$$\mathbb{E}(W^\bullet(\tau_1)) = \sigma_{\mathbf{p}}^2/2. \quad (25)$$

Moreover if  $w(1)$  has some exponential moment then  $W^\bullet(\tau_1)$  too.

**Proof .** (Taken from Feller 1). Let  $\lambda_r = \mathbb{P}(W^\bullet(\tau_1) = r)$ ; for any  $r \geq 0$ , let us establish

$$\lambda_r = \mu_r + \frac{\mu_{-1}}{1 - \lambda_0} \lambda_{r+1}. \quad (26)$$

Recall that we are talking about weak records. We have by a clear decomposition according to the first step

$$\lambda_r = \mathbb{P}(w(1) = r) + \mathbb{P}(w(1) = -1, W^\bullet(\tau_1) = r)$$

A path satisfying  $\{w(1) = -1, W^\bullet(\tau_1) = r\}$  can be decomposed without ambiguity as follows: A step -1, then a finite sequence of paths where  $W^\bullet(\tau_1) = 0$ , then a last path with  $W^\bullet(\tau_1) = r + 1$ , (relatively to level  $-1$ , the pieces in the decomposition coincides with weak records time) In term of probabilities, this becomes

$$\lambda_r = \mu_r + \mu_{-1} \left( \sum_{j=0}^{+\infty} \lambda_0^j \right) \lambda_{r+1},$$

(which is (26)) where  $j$  counts the number of passage at level  $-1$  between time 2 and  $\tau_1$ . Then

$$\sum_{r \geq 0} r \lambda_r = \sum_{r \geq 0} r \left( \mu_r + \frac{\mu_{-1}}{1 - \lambda_0} \lambda_{r+1} \right).$$

Which simplify to

$$\mathbb{E}(W^\bullet(\tau_1)) \left( 1 - \frac{\mu_{-1}}{1 - \lambda_0} \right) = 0,$$

Since  $\mathbb{E}(W^\bullet(\tau_1)) > 0$ , one finds  $\lambda_0 = 1 - \mu_{-1}$  and then, eventually

$$\lambda_r = \mu_r + \lambda_{r+1} = \mathbb{P}(W^\bullet(1) \geq r) \text{ for } r \geq 1; \quad (27)$$

Now the computation of the mean runs as follows :

$$\begin{aligned} \mathbb{E}(W^\bullet(\tau_1)) &= \sum_{k \geq 0} k \lambda_k \\ &= \sum_{k \geq 0} k \mathbb{P}(W^\bullet(1) \geq k) = \sum_{k \geq 0} k \sum_{l \geq k} \mathbb{P}(W^\bullet(1) = l) \\ &= \sum_{l=0}^{+\infty} \mathbb{P}(W^\bullet(1) = l) \left( \sum_{k=0}^l k \right) = \sum_{l=0}^{+\infty} \mathbb{P}(W^\bullet(1) = l) (l(l+1)/2) \\ &= (\mathbb{E}(W^\bullet(1))^2 - \mu_{-1}) + \mathbb{E}(W^\bullet(1)) + \mu_{-1} = \mathbb{E}(W^\bullet(1)^2) = \sigma_{\mathbf{p}}^2 \end{aligned}$$

Now, a r.v.  $X$  (or a measure  $\nu$ ) has some exponential moments around 0 iff there exists  $a > 0$  such that  $\mathbb{P}(X \geq k) \leq ce^{-ak}$  for any  $k > 0$  for some  $c > 0$ , and  $\mathbb{P}(X \leq -k) \leq ce^{-ak}$  for any  $k > 0$  for some  $c > 0$ . If  $\mu$  has exponential moments, then it is not difficult to see that also does  $\lambda$ .  $\square$

#### 4.5.1 Concentration of the number of records

What follows is a bit technical: the probabilities are cut into some small pieces; but only simple tools are used. We first start by showing that  $\frac{\sigma_{\mathbf{p}}^2}{2} R^\bullet(l)$  and  $\max_{0 \leq j \leq l} W^\bullet(j)$  are close :

**Proposition 5** *Assume (A). For each  $\nu > 0$ , there exists  $\gamma > 0$  and  $N > 0$  such that for all  $n \geq N$  and every  $l \in \llbracket 0, n \rrbracket$ ,*

$$\mathbb{P} \left( \left| \max_{0 \leq j \leq l} W^\bullet(j) - \frac{\sigma_{\mathbf{p}}^2}{2} R^\bullet(l) \right| \geq n^{1/4+\nu} \right) \leq \exp(-\gamma n^\nu). \quad (28)$$

Again since the order of  $\max_{0 \leq j \leq l} W^\bullet(j)$  is  $n^{1/2}$  this gives a rich information if  $\nu \in ]0, 1/4[$ .

**Proof .** Here  $W^\bullet$  is considered as a random walk with infinite length. Denote for short, for each  $k \geq 1$ ,  $\Delta_k = W^\bullet(\tau_k) - W^\bullet(\tau_{k-1})$  the  $k$ th records growth. By Markovianity  $\Delta_k \stackrel{(d)}{=} W^\bullet(\tau_1)$ , and the sequence  $(\Delta_k, k \geq 1)$  as well as the sequence  $(\tau_k - \tau_{k-1}, k \geq 1)$  is a sequence of i.i.d. r.v.. In addition

$$\max_{0 \leq j \leq l} W^\bullet(j) = \sum_{k=1}^{R^\bullet(l)} \Delta_k. \quad (29)$$

From here, a result as (28) is quite intuitive :  $\max_{0 \leq j \leq l} W^\bullet(j)$  is a sum of  $R^\bullet(l)$  variables having mean  $\mathbb{E}(\Delta_k) = \sigma_{\mathbf{p}}^2/2$ . Then  $\max_{0 \leq j \leq l} W^\bullet(j)$  should be close to  $R^\bullet(l)\sigma_{\mathbf{p}}^2/2$ . We will show that this is the case.

We will use the Petrov's Lemma : under (A),  $\Delta_1$  has some exponential moments. The left hand side (LHS) of (28) according to(29), is

$$\mathbb{P}\left(\left|\sum_{k=1}^{R^\bullet(l)} \left(\Delta_k - \frac{\sigma_{\mathbf{p}}^2}{2}\right)\right| \geq n^{1/4+\nu}\right) \leq A(l, n) + B(l, n)$$

where

$$\begin{aligned} A(l, n) &= \mathbb{P}(R^\bullet(l) \geq n^{1/2+\nu}), \\ B(l, n) &= \mathbb{P}\left(\left|\sum_{k=1}^{R^\bullet(l)} \left(\Delta_k - \frac{\sigma_{\mathbf{p}}^2}{2}\right)\right| \geq n^{1/4+\nu}, R^\bullet(l) \leq n^{1/2+\nu}\right). \end{aligned}$$

Let us see why  $A(l, n)$  and  $B(l, n)$  are small; define the random walk  $(U_k)_{k \geq 0}$  :

$$U_k = \sum_{j=1}^k (\Delta_j - \sigma_{\mathbf{p}}^2/2);$$

its increments are the centered r.v.  $(\Delta_j - \frac{\sigma_{\mathbf{p}}^2}{2}, j \geq 1)$ , and own exponential moments. The following Lemma 11 (that follows) allows to end the proof of Proposition 5.

**Lemma 11** 1) For  $\nu > 0$ , there exist constants  $c_1 > 0$  and  $N_1 > 0$  s.t. for any  $n \geq N_1$  and all  $l \in \llbracket 0, n \rrbracket$ ,

$$A(l, n) \leq \exp(-c_1 n^\nu)$$

2) For  $\nu > 0$ , there exist two constants  $\tilde{c}_1 > 0$  and  $N_2 > 0$  s.t. for every  $n \geq N_2$  and all  $l \in \llbracket 0, n \rrbracket$ ,

$$B(l, n) \leq \exp(-\tilde{c}_1 n^\nu).$$

**Proof .** (2) To control  $B(l, n)$ , write

$$\begin{aligned} B(l, n) &\leq \sum_{j=1}^{n^{1/2+\nu}} \mathbb{P}(|U_{R^\bullet(l)}| \geq n^{1/4+\nu} \mid R^\bullet(l) = j) \mathbb{P}(R^\bullet(l) = j) \\ &\leq \max_{j \in \llbracket 1, n^{1/2+\nu} \rrbracket} \mathbb{P}(|U_j| \geq n^{1/4+\nu}) \end{aligned}$$

since  $U_k$  is a sum of centered r.v. having exponential moments, we can use Petrov's lemma. Each term is smaller than  $2e^{-Tn^{\nu+1/4}}$  or  $2e^{-n^\nu/2g}$  according to the value of  $j$ , as says Petrov's Lemma. In any case, there exists  $c > 0$ , which does not depend on  $l$ , such that this is smaller than  $2e^{-cn^\nu}$ .

(1) In the sequel we note  $n^{1/2+\nu}$  instead of  $\lfloor n^{1/2+\nu} \rfloor$ . For all  $l \in \llbracket 0, n \rrbracket$ ,

$$\begin{aligned} A(l, n) &\leq \mathbb{P}(\tau_{n^{1/2+\nu}} \leq l) \\ &\leq \mathbb{P}(\tau_{n^{1/2+\nu}} \leq n) \\ &\leq \mathbb{P}\left(\tau_{n^{1/2+\nu}} \leq n, \max_{0 \leq j \leq n} W^\bullet(j) \leq n^{1/2+\nu/2}\right) + \mathbb{P}\left(\max_{0 \leq j \leq n} W^\bullet(j) \geq n^{1/2+\nu/2}\right) \end{aligned}$$

By the union bound, the second term is smaller than  $n$  terms all smaller than  $\exp(-cn^\nu)$ . The first one is bounded by

$$\mathbb{P}\left(\max_{0 \leq j \leq \tau_{n^{1/2+\nu}}} W^\bullet(j) \leq n^{1/2+\nu/2}\right).$$

Using  $\max_{0 \leq j \leq \tau_{n^{1/2+\nu}}} W^\bullet(j) = U_{n^{1/2+\nu}} + \lfloor n^{1/2+\nu} \rfloor \frac{\sigma_{\mathbf{P}}^2}{2}$ , this term rewrites

$$\begin{aligned} &= \mathbb{P}(U_{n^{1/2+\nu}} + \lfloor n^{1/2+\nu} \rfloor \frac{\sigma_{\mathbf{P}}^2}{2} \leq n^{1/2+\nu/2}) \\ &\leq \mathbb{P}(-U_{n^{1/2+\nu}} \geq \lfloor n^{1/2+\nu} \rfloor \frac{\sigma_{\mathbf{P}}^2}{2} - n^{1/2+\nu/2}) \\ &\leq \exp(-c_2 n^{1/2+\nu}). \end{aligned}$$

by Petrov again, for  $n$  large enough; Lemma 11 follows.  $\square$

#### 4.5.2 Back to right minima considerations on $W$

Proposition 5 and (23) says that for any  $\nu > 0$  there exist constants  $\gamma > 0$  and  $N > 0$  s.t. for all  $n > N$  and  $l \in \llbracket 0, n \rrbracket$

$$\mathbb{P}\left(\left|W(l) - \min_{0 \leq i \leq l} W(i) - RM(l) \frac{\sigma_{\mathbf{P}}^2}{2}\right| \geq n^{1/4+\nu}\right) \leq \exp(-\gamma n^\nu); \quad (30)$$

Since this is true for each  $l \in \llbracket 0, n \rrbracket$ , by the union bound :

**Corollary 2** For all  $\nu > 0$  there exist two constants  $\gamma > 0$  and  $N > 0$  s.t. for all  $n > N$

$$\mathbb{P}\left(\sup_{0 \leq l \leq n} \left|W(l) - \min_{0 \leq i \leq l} W(i) - RM(l) \frac{\sigma_{\mathbf{P}}^2}{2}\right| \geq n^{1/4+\nu}\right) = O(\exp(-\gamma n^\nu)).$$

**Proof of Theorem 5 :** As already said  $S_{\mathbf{T}}$  has the same law as the random walk  $W$  conditioned by  $\{\text{Hit}_{-1}(W) = n\}$ . (Notice that for  $l \leq |\mathbf{T}|$ ,  $\min_{0 \leq i \leq l} S_{\mathbf{T}}(i) = 0$ ). Write

$$\mathbb{P}_{\mathbf{P}}\left(\sup_{0 \leq l \leq n} \left|S_{\mathbf{T}}(l) - H_{\mathbf{T}}(l) \frac{\sigma_{\mathbf{P}}^2}{2}\right| \geq n^{1/4+\nu} \mid |\mathbf{T}| = n\right)$$

$$= \mathbb{P} \left( \sup_{0 \leq l \leq n} \left| W(l) - \min_{0 \leq i \leq l} W(i) - \#RM(l) \frac{\sigma_{\mathbf{P}}^2}{2} \right| \geq n^{1/4+\nu} \mid W \in \mathbf{L}_n \right).$$

According to Corollaries 1 and 2, there exists a constant  $C > 0$  such that

$$\mathbb{P}_{\mathbf{P}} \left( \sup_{0 \leq l \leq n} \left| S_{\mathbf{T}}(l) - H_{\mathbf{T}}(l) \frac{\sigma_{\mathbf{P}}^2}{2} \right| \geq n^{1/4+\nu} \mid |\mathbf{T}| = n \right) = O(n^{5/2}) \exp(-Cn^{\nu}). \quad \square$$

#### 4.6 Relation between the contour process and the DFQ

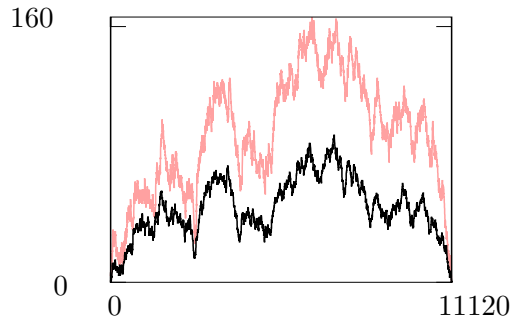
The aim now is to establish the following theorem

**Theorem 6** *For any  $\nu > 0$ , there exist some constants  $\gamma > 0$  and  $N > 0$  such that*

$$\forall n \geq N, \quad \mathbb{P}_{\mathbf{P}} \left( \sup_{t \in [0,1]} \left| S_{\mathbf{T}}(nt) - \frac{\sigma_{\mathbf{P}}^2}{2} V_{\mathbf{T}}(2nt) \right| \geq n^{1/4+\nu} \mid |\mathbf{T}| = n \right) = O(\exp(-\gamma n^{\nu})).$$

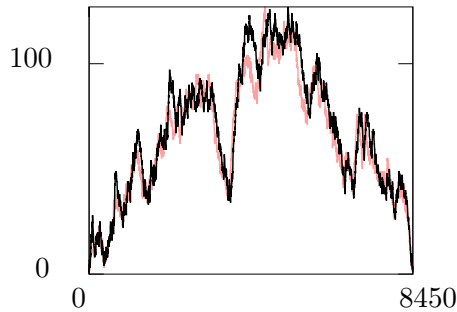
Theorem 6 possesses very visual interpretations, as can be seen on the following simulations.

1) On the following figure, we have simulated a tree with size 5560 and with lattice offspring distribution  $p_0 = 13/18, p_2 = 1/6, p_6 = 1/9$ ; the corresponding value of  $\sigma^2/2$  is  $11/6$ . In black, the contour process, in grey the DFQ. One can observe the quasi proportionality of the two processes.



Note also that since  $2l \geq m(l)$  (for any  $l$ ), the contour process “lags” from the DFQ.

2) On the following figure, we have simulated a size 4208 tree with a non lattice offspring distribution  $p_0 = 8/15, p_1 = 4/15, p_3 = 2/15, p_5 = 1/15$ , (in this case  $\sigma^2/2 = 16/15$ )



**Proof .** Recall Lemmas 2 and 3 page 24. For each  $t \in [0, 1[$ , we note  $\zeta(nt)$  the integer such that

$$[2nt] \in [m(\zeta(nt)), m(\zeta(nt) + 1)]. \quad (31)$$

Now, decompose...

$$\mathbb{P}_{\mathbf{p}}\left(\sup_{t \in [0,1]} |S_{\mathbf{T}}(nt) - \frac{\sigma^2}{2} V_{\mathbf{T}}(2nt)| \geq n^{1/4+\nu} \mid |\mathbf{T}| = n\right) \leq A + B + C + D$$

where

$$\begin{aligned} A &= \mathbb{P}_{\mathbf{p}}\left(\sup_{t \in [0,1]} |S_{\mathbf{T}}(nt) - S_{\mathbf{T}}(\zeta(nt))| \geq \frac{n^{1/4+\nu}}{3}, \sup_t |\zeta(nt) - nt| < n^{1/2+\nu} \mid |\mathbf{T}| = n\right) \quad (32) \\ B &= \mathbb{P}_{\mathbf{p}}\left(\sup_{t \in [0,1]} \left|S_{\mathbf{T}}(\zeta(nt)) - \frac{\sigma_{\mathbf{p}}^2}{2} H_{\mathbf{T}}(\zeta(nt))\right| \geq \frac{n^{1/4+\nu}}{3} \mid |\mathbf{T}| = n\right) \\ C &= \mathbb{P}_{\mathbf{p}}\left(\frac{\sigma_{\mathbf{p}}^2}{2} \sup_{t \in [0,1]} |H_{\mathbf{T}}(\zeta(nt)) - V_{\mathbf{T}}(2nt)| \geq \frac{n^{1/4+\nu}}{3} \mid |\mathbf{T}| = n\right) \\ D &= \mathbb{P}_{\mathbf{p}}\left(\sup_{t \in [0,1]} |\zeta(nt) - nt| > n^{1/2+\nu} \mid |\mathbf{T}| = n\right) \quad (33) \end{aligned}$$

To control  $A$ , again, we remove the condition:

$$\begin{aligned} A &= O(n^{3/2}) \mathbb{P}\left(\sup_{k \leq n^{1/2+\nu}} \sup_{0 \leq j \leq n-k} |W(j) - W(j+k)| \geq \frac{n^{1/4+\nu}}{3}\right) \\ &= O(n^{3/2}) \sum_{j=1}^n \mathbb{P}\left(\sup_{k \leq n^{1/2+\nu}} |W(k)| \geq \frac{n^{1/4+\nu}}{3}\right) = O(e^{-c_5 n^\nu}), \end{aligned}$$

for a constant  $c_5 > 0$ , if  $n$  is large enough according to Lemma 8.

For  $B$ , by Theorem 5, there exists  $c_6 > 0$  such that  $B = O(n^{5/2} \exp(-c_6 n^\nu))$ .

Since  $[2nt] \in [m(\zeta(nt)), m(\zeta(nt) + 1)[$ , we have according Lemma 3 :

$$\begin{aligned} C &\leq \mathbb{P}_{\mathbf{p}}\left(\frac{\sigma^2}{2} \sup_{l \in \{1, \dots, n\}} |H_{\mathbf{T}}(l) - H_{\mathbf{T}}(l+1) + 1| \geq \frac{n^{1/4+\nu}}{3} \mid |\mathbf{T}| = n\right) \\ &\leq \mathbb{P}_{\mathbf{p}}\left(\frac{\sigma^2}{2} \sup_{l \in \{1, \dots, n\}} |H_{\mathbf{T}}(l+1) - 1 - \frac{2}{\sigma_{\mathbf{p}}^2} S_{\mathbf{T}}(l)| \geq \frac{n^{1/4+\nu}}{6} \mid |\mathbf{T}| = n\right) \\ &\quad + \mathbb{P}_{\mathbf{p}}\left(\frac{\sigma^2}{2} \sup_{l \in \{1, \dots, n\}} |H_{\mathbf{T}}(l) - \frac{2}{\sigma_{\mathbf{p}}^2} S_{\mathbf{T}}(l)| \geq \frac{n^{1/4+\nu}}{6} \mid |\mathbf{T}| = n\right) \end{aligned}$$

We then conclude by the Theorem 5.

To bound  $D$ , we use (31) and Lemma 2; we have

$$\begin{aligned} \sup_{t \in [0,1]} |2\zeta(nt) - 2nt| &\leq \sup_{t \in [0,1]} |2\zeta(nt) - m(\zeta(nt))| + \sup_{t \in [0,1]} |m(\zeta(nt)) - 2nt| \\ &\leq 2 \sup_{0 \leq l \leq n} H_{\mathbf{T}}(l). \end{aligned}$$

since  $H_{\mathbf{T}}(l)$  is the number of right minima of  $S_{\mathbf{T}}$ , use (21) and Corollary 1 we get

$$\mathbb{P}_{\mathbf{P}} \left( \sup_{0 \leq l \leq n} H_{\mathbf{T}}(l) \geq n^{1/2+\nu} \mid |\mathbf{T}| = n \right) = O(n^{3/2}) \mathbb{P} \left( \sup_{0 \leq l \leq n} R^{\bullet}(l) \geq n^{1/2+\nu} \right).$$

We conclude by Lemma 11, and the union bound.  $\square$



## 5 Continuous processes and weak convergence in $C[0, 1]$

In this part we provide some elements allowing one to understand the convergence in  $C[0, 1]$ , the set of continuous functions from  $[0, 1]$  with values in  $\mathbb{R}$ ; we then evoke the convergence in distribution of random walk and excursion type paths, properly rescaled. To begin, we give some elements to understand what is a probability measure on  $C[0, 1]$ , how it is characterized.

### 5.1 Probability measure on a Polish set

Let  $\mathcal{S}$  be a set and  $\rho$  a distance on  $\mathcal{S}$ . For  $x$  in  $\mathcal{S}$ ,  $r \geq 0$ , the ball  $B(x, r)$  of center  $x$  and radius  $r$ , is the set

$$B(x, r) = \{y \in \mathcal{S}, \rho(x, y) < r\};$$

an open set of  $\mathcal{S}$  is any union of balls. The family of open subsets of a space forms the standard topology (a topology is a collection of subsets of a space  $E$ , stable by union, by finite intersection, containing  $E$  and the empty set).

A set equipped with a distance  $(\mathcal{S}, \rho)$  is called a **Polish space**, if it is separable and complete:

- separable: a countable subset  $D$  of  $\mathcal{S}$  is dense ( $D$  is dense in  $\mathcal{S}$ , if for any  $s \in \mathcal{S}$ , there exists an element of  $D$  as close as wanted to  $s$ : for any  $\varepsilon > 0$  there exists  $d \in D$  such that  $\rho(s, d) < \varepsilon$ ),
- complete: a space  $\mathcal{S}$  is complete if any Cauchy sequence of  $\mathcal{S}$  converges in  $\mathcal{S}$  (has a limit, and the limit is inside  $\mathcal{S}$ ). A sequence  $(s_n)$  is a Cauchy sequence, if for any  $\varepsilon > 0$ , there exists  $N$  such that if  $n \geq N$  and  $m \geq N$  then  $|x_n - x_m| \leq \varepsilon$ .

Hence  $\mathbb{R}$  is a Polish space when equipped with the distance  $d(x, y) = |x - y|$ , since:

- $\mathbb{Q}$  is countable and dense,
- The Cauchy sequences converge: a Cauchy sequence  $(x_n)$  is clearly bounded, then as some accumulations points. The set of accumulation points is necessarily of cardinality 1 (since accumulation points are at distance smaller than  $\varepsilon$ , for any  $\varepsilon$ , clearly).

The **Borelian**  $\sigma$ -field  $\mathcal{B}(\mathcal{S})$  is the  $\sigma$ -field induced by the standard topology: this is the set of subsets of  $\mathcal{S}$  that are obtained by a (at most) countable sequence of unions, intersections, complementations, applied on the set of open sets of  $\mathcal{S}$ .

Hence, the open sets, closed sets, singleton, sets with a countable cardinality... are all Borelian sets. In fact, what is difficult is to find a non Borelian set (even on  $\mathbb{R}$ , it is not at all a simple exercise).

### 5.2 Measure on Polish space. Characterization

A function  $f$  defined on  $\mathcal{S}$ , taking its value in  $\mathbb{R}$  is continuous, if for any open set  $\mathcal{O}$  of  $\mathbb{R}$ ,  $f^{-1}(\mathcal{O})$  is an open set of  $\mathcal{S}$ . The function  $f$  is said to be measurable (or Borelian) if for any  $B$  in

$\mathcal{B}(\mathbb{R})$ ,  $f^{-1}(B) \in \mathcal{B}(\mathcal{S})$ . It is easy to see, that a sufficient condition, is that  $f^{-1}(\mathcal{O}) \in \mathcal{B}(\mathcal{S})$  for any open set  $\mathcal{O}$  of  $\mathbb{R}$ . In particular, the continuous functions from  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  to  $\mathbb{R}$  are measurable.

The measure  $\mathbb{P}$  is sufficient to define a notion of integral for measurable non negative functions  $f$  :

$$\int f d\mathbb{P};$$

this is done in several steps : if  $f = \mathbb{I}_A$  (is a indicator function of a set) where  $A \in \mathcal{B}(\mathcal{S})$ , set

$$\int_S \mathbb{I}_A(\omega) d\mathbb{P}(\omega) = \int_A d\mathbb{P}(\omega) = \mathbb{P}(A).$$

Prolong by linearity : let  $A_1, \dots, A_n$  a family of Borelian (disjoint or not) from  $\mathcal{B}(\mathcal{S})$  and any real numbers  $\lambda_1, \dots, \lambda_n$ ; for  $f$  defined by  $f = \sum_{i=1}^n \lambda_i \mathbb{I}_{A_i}$ , set

$$\int_S f(\omega) d\mathbb{P}(\omega) = \int \sum_{i=1}^n \lambda_i \mathbb{I}_{A_i}(\omega) d\mathbb{P}(\omega) = \sum_{i=1}^n \lambda_i \int_{A_i} d\mathbb{P}(\omega) = \sum_{i=1}^n \lambda_i \mathbb{P}(A_i).$$

This construction is valid, since it can be checked that the result does not depend on the representation of a given function  $f$ .

The notion of integral is then defined for the family of step functions (positive or not). Then one shows that one may build an increasing “function integral” and linear (that is,  $f \longrightarrow \int f d\mathbb{P}$  is increasing and linear), on (first) the set of measurable non negative functions. For this, we use that any measurable function  $f$ , non negative is a limit of an increasing sequence of non negative scale functions  $f_n$  : for example,

$$f_n(x) = \begin{cases} [10^n f(x)]/10^n & \text{for } f(x) \leq n \\ n & \text{for } f(x) > n \end{cases}$$

which is the below approximation at  $10^{-n}$  of  $f(x)$  for  $f(x) < n$ . It is easy to check that  $f_n$  is measurable, is a step function, and that  $f_n$  is increasing.

One then sets 
$$\int f d\mathbb{P} = \lim_n \int f_n d\mathbb{P}. \tag{34}$$

The integral of a positive bounded function  $f$  is always finite (since the measure is finite,  $\mathbb{P}(\mathcal{S}) = 1$ ). The integral of a measurable function  $f$  is defined in the case where  $\int f^+ d\mathbb{P}$  and  $\int f^- d\mathbb{P}$  are finite, by

$$\int f d\mathbb{P} = \int f^+ d\mathbb{P} - \int f^- d\mathbb{P};$$

if only one of the integral  $\int f^+ d\mathbb{P}$  or  $\int f^- d\mathbb{P}$  diverges, we say accordingly that  $\int f d\mathbb{P} = +\infty$  or  $-\infty$ . The words “one then sets” in (34) is a hidden theorem: one has to show that the limit does not depend on the chosen sequence  $f_n$ .

**Lemma 12** Any probability measure on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  is regular, that is, for  $A$  in  $\mathcal{S}$  and  $\varepsilon > 0$ , there exists a closed set  $F$  and an open set  $\mathcal{O}$  such that  $F \subset A \subset \mathcal{O}$  and  $\mathbb{P}(\mathcal{O} - F) < \varepsilon$ .

**Proof .** Let  $E$  be the set of subsets of  $A$  in  $\mathcal{S}$  satisfying the property stated in the Lemma. Let  $\rho$  denote the metric on  $\mathcal{S}$ ;  $\rho(x, A)$  is the distance between  $x$  and the set  $A$ . If  $A$  is closed, take  $F = A$  and  $\mathcal{O} = \mathcal{O}_\delta = \{x, \rho(x, A) < \delta\}$  for  $\delta$  chosen small enough (indeed,  $\mathcal{O}_{1/n}$  is decreasing, and  $\lim_n \mathbb{P}(\mathcal{O}_{1/n}) = \mathbb{P}(\lim \mathcal{O}_{1/n}) = \mathbb{P}(\overline{A}) = \mathbb{P}(F)$ ). Hence, every closed set belongs to  $E$ . Let us show that the family of sets  $E$  having this property is a  $\sigma$ -field (since it has already been shown that it contains the closed sets, it would contain the Borelian sets). Let  $A_n$  be a sequence of elements of  $E$ . Consider some open sets  $\mathcal{O}_n$  and closed set  $F_n$  such that

$$F_n \subset A_n \subset \mathcal{O}_n \text{ and } \mathbb{P}(\mathcal{O}_n - F_n) < \varepsilon/2^{n+1}.$$

Then, take  $\mathcal{O} = \cup_n \mathcal{O}_n$  and  $F = \cup_{n \leq n_0} F_n$  ( $n_0$  is chosen such that  $\mathbb{P}(\cup_n F_n - F) < \varepsilon/2$ ). One then has  $A = \cup_n A_n \in E$ . Since  $E$  is stable by taking the complementary, this ends the proof.  $\square$

**Lemma 13** *Two measures on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  which coincide on the closed sets of  $\mathcal{S}$  (or on the open sets of  $\mathcal{S}$ ) are equal.*

**Proof .** Two measures  $\mathbb{P}$  and  $\mathbb{P}'$  on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  are equal if they are equal on all Borelian sets. The closed set generates the Borelian  $\sigma$ -field, but this does not provide a proof of the Lemma.

First, by taking the complementary, it suffices to prove this lemma for closed sets. Let us see why this is a consequence of the previous Lemma. Let  $A$  be a Borelian and  $F$  and  $\mathcal{O}$  as said in the previous Lemma. One has

$$\mathbb{P}(F) \leq \mathbb{P}(A) \leq \mathbb{P}(\mathcal{O}) = 1 - \mathbb{P}(\mathcal{O}^c).$$

Hence, to know the probability on the closed sets is sufficient to determine the probability of any Borelian.  $\square$

To go faster, one can also use a Theorem about Dynkin systems, which states the following result. (A  $\pi$ -system of Dynkin is a set of subsets stable by finite intersection).

**Lemma 14** *If two measures on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  coincides on a  $\pi$ -system which generates  $\mathcal{B}(\mathcal{S})$ , then, they coincide on  $\mathcal{B}(\mathcal{S})$ .*

(Billingsley, Probability and measure, Theorem 3.3). Open sets forms a  $\pi$ -system, and generate the Borelian  $\sigma$ -field.

Associated with  $\mathbb{P}$ , a notion of integral has been defined. If  $\mathbb{P}$  and  $\mathbb{P}'$  are two measures on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  such that for any measurable function  $f$ , one has  $\int f d\mathbb{P} = \int f d\mathbb{P}'$  then  $\mathbb{P} = \mathbb{P}'$  (that is, for all  $A$  in  $\mathcal{B}(\mathcal{S})$ , one has  $\mathbb{P}(A) = \mathbb{P}'(A)$ ). This is clear, since it suffices to take  $f = \mathbb{I}_A$  to check it. Which is less obvious is the following proposition:

**Proposition 6** *Let  $\mathbb{P}$  and  $\mathbb{P}'$  be two probability measures on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  such that  $\int f d\mathbb{P} = \int f d\mathbb{P}'$  for all bounded continuous function  $f$  (with values in  $\mathbb{R}$ ). Under these conditions, one has  $\mathbb{P} = \mathbb{P}'$ .*

**Proof .** To prove this proposition, we will show that the two measures coincide on closed sets (and then, we use Lemma 13). For this, we approximate the function  $\mathbb{I}_F$  of a closed set  $F$  by bounded continuous functions. We make here the same work as on  $\mathbb{R}$  : let  $\psi$  defined on  $\mathbb{R}$  taking its values in  $\mathbb{R}$  defined by

$$\psi(t) = \begin{cases} 1 - t & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if } t \geq 1 \end{cases} .$$

The function  $f_n$  defined by  $f_n(x) = \psi(n\rho(x, F))$  is equal to 1 if  $x \in F$ , 0 if  $\rho(x, F) > 1/n$  and is  $[0,1]$  in the other case (it is uniformly continuous).

The sequence of functions  $f_n$  is decreasing and converge point-wise toward  $\mathbb{I}_F$ . Since all the  $f_n$  are dominated by the constant function  $g = 1$  on  $\mathcal{S}$  (which is integrable), by the Theorem of Lebesgue (of dominated convergence) one has

$$\lim_n \int_{\mathcal{S}} f_n(x) d\mathbb{P}(x) = \int_{\mathcal{S}} \mathbb{I}_F(x) d\mathbb{P}(x) = \mathbb{P}(F).$$

Then, one gets  $\mathbb{P}'(F) = \lim_n \int_{\mathcal{S}} f_n(x) d\mathbb{P}'(x)$ . Hence  $\mathbb{P}(F) = \mathbb{P}'(F)$ .  $\square$

### 5.3 Random variables with a values in a Polish space

Let  $(\Omega, \mathcal{A}, \mathbb{P})$  a probability space, and  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  a Polish space, equipped with its Borelian  $\sigma$ -field.

A map  $X : \Omega \rightarrow \mathcal{S}$  is a random variable if  $X$  is measurable from  $(\Omega, \mathcal{A})$  to  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  (that is,  $X^{-1}(b) \in \mathcal{A}$  for any  $b \in \mathcal{B}(\mathcal{S})$ ). The distribution of  $X$ , denoted by  $\mathbb{P}_X$  is the probability measure defined on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  by

$$\mathbb{P}_X(b) = \mathbb{P}(X^{-1}(b)) = \mathbb{P}(\{\omega \mid X(\omega) \in b\}).$$

Let  $f$  be a measurable function from  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  with values in  $\mathbb{R}$  (equipped with  $\mathcal{B}(\mathbb{R})$ ). The expectation of  $f(X)$  is defined by

$$\mathbb{E}(f(X)) = \int f(x) d\mathbb{P}_X(x).$$

### 5.4 Convergence in distribution in a Polish space

**Definition :** Let  $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$  be a sequence of probability measure defined on the same Polish space  $(\mathcal{S}, \mathcal{S})$ . The sequence  $(\mathbb{P}_n, n \geq 1)$  weakly converges to  $\mathbb{P}$  (we write  $\mathbb{P}_n \xrightarrow[n]{weak} \mathbb{P}$ ), if for any bounded continuous function  $f$  (from  $\mathcal{S}$  to  $\mathbb{R}$ ),

$$\int f d\mathbb{P}_n \longrightarrow \int f d\mathbb{P}. \tag{35}$$

A sequence of r.v.  $(X_n)$  converges in distribution to  $X$  if  $\mathbb{P}_{X_n} \xrightarrow[n]{weak} \mathbb{P}_X$ , in other words, if

$$\mathbb{E}(f(X_n)) = \int f(x) d\mathbb{P}_{X_n}(x) \longrightarrow \int f(x) d\mathbb{P}_X(x) = \mathbb{E}(f(X)),$$

for bounded continuous functions  $f$ . We note  $X_n \xrightarrow[n]{d} X$ .

We have

**Proposition 7** *Let  $(X_n, n \geq 1)$  be a sequence of r.v. taking their values in a Polish space  $(S, \mathcal{S})$ . If  $X_n \xrightarrow[n]{weak} X$  then for any continuous function  $f$  from  $S$  to a Polish space  $(S', \mathcal{S}')$ , then  $g(X_n) \xrightarrow[n]{d} g(X)$ .*

**Proof .** It suffices to show that for  $h$  bounded continuous, from  $S'$  to  $\mathbb{R}$ , we have  $\mathbb{E}(h(g(X_n))) \rightarrow \mathbb{E}(h(g(X)))$ . Since  $h \circ g$  is bounded continuous from  $S$  to  $\mathbb{R}$ , it is indeed the case since  $X_n \xrightarrow[n]{weak} X$ .  $\square$

Other criterion: the Portmanteau Theorem (recall Theorem 2 page 15).

### 5.5 Probability measures on $(C[0, 1], \mathcal{B}(C[0, 1]))$

The norm  $\|\cdot\|_\infty$  induce a metric ( $d(f, g) = \|f - g\|_\infty$ ) on  $C[0, 1]$ ; this space equipped with this metric is a Polish space (exercise). The Borelian  $\sigma$ -field of  $C[0, 1]$  is the  $\sigma$ -field generated by the open balls for the norm  $\|\cdot\|_\infty$ ; a ball  $B(f, r) = \{g \in C[0, 1] \mid \|f - g\|_\infty < r\}$  where  $r \in \mathbb{R}^{+\ast}$ , and  $f \in C[0, 1]$ . According to Lemma 13, a probability measure  $\mathbb{P}$  on  $C[0, 1]$  is determined by its value on the open sets (or closed sets).

The map  $\pi_t$

$$\begin{aligned} \pi_t : C[0, 1] &\longrightarrow \mathbb{R} \\ f &\longmapsto f(t) \end{aligned}$$

associates with each function  $f$  from  $C[0, 1]$  the value  $f(t)$ . The function  $\pi_t$  is called coordinate function. For any  $t$ ,  $\pi_t$  is continuous and thus measurable. To see the continuity, we have to check that  $\pi_t^{-1}(] - \infty, \alpha[)$  is an open set for any  $\alpha$ ; to see this take  $f$  in  $\pi_t^{-1}(] - \infty, \alpha[)$ ; therefore, by definition  $f(t) < \alpha$ ; clearly the ball  $B(f, |f(t) - \alpha|/2)$  is also included in  $\pi_t^{-1}(] - \infty, \alpha[)$ . Hence, around each point of  $\pi_t^{-1}(] - \infty, \alpha[)$ , there is an open ball entirely included in  $\pi_t^{-1}(] - \infty, \alpha[)$ . This set is thus open.

Therefore  $\pi_t$  is a random variable on  $(C[0, 1], \mathcal{B}(\mathbb{R}), \mathbb{P})$  with values in  $\mathbb{R}$ , its distribution being as usual  $\mathbb{P} \circ \pi_t^{-1}$  : for any  $b \in \mathcal{B}(\mathbb{R})$ ,

$$\mathbb{P}_{\pi_t}(b) = \mathbb{P}(\pi_t^{-1}(b)) = \mathbb{P}(\{f \in C[0, 1] \mid f(t) \in b\}).$$

Further, the family of maps  $\pi_{t_1, \dots, t_k}$  (indexed by the possible  $k$ 's and  $t_i$ 's) are called the finite dimensional coordinates functions :

$$\begin{aligned} \pi_{t_1, \dots, t_k} : C[0, 1] &\longrightarrow \mathbb{R}^k \\ f &\longmapsto (f(t_1), \dots, f(t_k)) \end{aligned}$$

for  $(t_1, \dots, t_k) \in [0, 1]^k$ . The map  $\pi_{t_1, \dots, t_k}$  is a random variable with values in  $\mathbb{R}^k$  (equipped with the Borelian  $\sigma$ -field). The family of distributions  $\mathbb{P} \circ \pi_{t_1, \dots, t_k}^{-1}$  for  $k$  in  $\mathbb{N}^\ast$  and  $(t_i)_{i \in [1, k]}$  in  $[0, 1]^k$ , are called the finite dimensional distribution (FDD) of  $\mathbb{P}$ .

The family of random variables  $(\pi_t)$  is called canonical process.

Again, there exist also random variables with values in  $C[0, 1]$  (they are called stochastic processes); let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space. A map  $\mathbf{x}$  from  $(\Omega, \mathcal{A}, \mathbb{P})$  with values in  $(C[0, 1], \mathcal{B}(C[0, 1]))$  is a random variable if it is measurable.

The distribution  $\mathbb{P}_{\mathbf{x}}$  of  $\mathbf{x}$  is defined as usual by  $\mathbb{P}_{\mathbf{x}}(A) = \mathbb{P}(\{\omega \mid \mathbf{x}(\omega) \in A\}) = \mathbb{P}(\mathbf{x}^{-1}(A))$  for all elements  $A$  in  $\mathcal{B}(C[0, 1])$ . One has also  $\pi_t(\mathbf{x}) = \mathbf{x}(t)$  is a r.v. having distribution  $\mathbb{P}_{\mathbf{x}} \circ \pi_t^{-1}$ . Further,  $\pi_{t_1, \dots, t_k}(\mathbf{x}) = (\mathbf{x}(t_1), \dots, \mathbf{x}(t_k))$  is a random vector with distribution  $\mathbb{P}_{\mathbf{x}} \circ \pi_{t_1, \dots, t_k}^{-1}$ .

We saw that the FDD of  $\mathbb{P}$  is the family  $\mathbb{P} \circ \pi_{t_1, \dots, t_k}^{-1}$  for all  $k$  and  $t_i$ 's (then the FDD are defined given  $\mathbb{P}$ ). Conversely

**Proposition 8** *Let  $\mathbb{P}$  and  $\mathbb{P}'$  be two probability distributions on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ . If the FDD of  $\mathbb{P}$  and  $\mathbb{P}'$  are equal then  $\mathbb{P} = \mathbb{P}'$ .*

(In other words, if  $\mathbf{x}$  and  $\mathbf{y}$  are two processes with values in  $C[0, 1]$ . If for any  $k$  in  $\mathbb{N}^*$  and  $(t_i)_{i \in \llbracket 1, k \rrbracket}$  in  $[0, 1]^k$ ,  $(\mathbf{x}(t_1), \dots, \mathbf{x}(t_k)) \stackrel{(d)}{=} (\mathbf{y}(t_1), \dots, \mathbf{y}(t_k))$  then  $\mathbf{x}$  and  $\mathbf{y}$  have the same distribution).

**Proof .** We call finite dimensional set, a set of the form  $\pi_{t_1, \dots, t_k}^{-1}(H)$  where  $H$  is a Borelian from  $\mathbb{R}^k$ . Finite dimensional sets form a  $\pi$ -system of Dynkin (stable by finite intersections). Hence, two probabilities  $\mathbb{P}$  and  $\mathbb{P}'$  coinciding on finite dimensional sets, coincides on the  $\sigma$ -field they generate. It remains to see that this  $\sigma$ -field is the Borelian  $\sigma$ -fields, which will allow one to deduce  $\mathbb{P} = \mathbb{P}'$ . For this, consider  $\overline{B}(f, r) = \{g \mid \|g - f\|_{\infty} \leq r\}$  be the closed ball with center  $f \in C[0, 1]$  and  $r \geq 0$ . We have

$$\overline{B}(f, r) = \bigcap_{t \in [0, 1] \cap \mathbb{Q}} \{g \mid f(t) - r \leq g(t) \leq f(t) + r\}$$

Write the rational number in  $[0, 1]$  as a list,  $t_1, t_2, \dots$ , and let

$$A_k = \bigcap_{t \in \{t_1, \dots, t_k\}} \{g \mid f(t) - r \leq g(t) \leq f(t) + r\} = \pi_{t_1, \dots, t_k}^{-1} \left( \prod_{i=1}^k [f(t_i) - r, f(t_i) + r] \right)$$

where  $\prod_{i=1}^k [f(t_i) - r, f(t_i) + r]$  is the parallelepiped in  $\mathbb{R}^k$ . Hence,  $A_k$  is a finite dimensional set and we see that  $\cap A_k = \overline{B}(f, r)$ . Hence the  $\sigma$ -field generated by the finite dimensional set contained the closed balls and then Borelian sets. Hence,  $\mathbb{P} = \mathbb{P}'$ .  $\square$

### Examples of probability measure on $C[0, 1]$

A) Consider the random function  $W := (t \mapsto Xt)$ ,  $t \in [0, 1]$ , where  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is a real random variable. This random function, is of course a random process with values in  $C[0, 1]$ . Let us be more formal, to see what happens. The map

$$\Phi : \begin{array}{ccc} (\mathbb{R}, \mathcal{B}(\mathbb{R})) & \longrightarrow & (C[0, 1], \mathcal{B}(C[0, 1])) \\ x & \longmapsto & (t \mapsto xt) \end{array} ;$$

is measurable (why?). Hence,  $W = \Phi \circ X$  is a r.v. from  $(\Omega, \mathcal{A})$  taking its values in  $(C[0, 1], \mathcal{B}(C[0, 1]))$ . Let  $\mathbb{P}_X$  be the law of  $X$ ; the distribution of  $W$  on  $(C[0, 1], \mathcal{B}(C[0, 1]))$  is given by  $\mathbb{P}_W = \mathbb{P}(W \in b) = \mathbb{P}(\{\omega : \Phi(X(\omega)) \in b\}) = \mathbb{P}(\{\omega : X(\omega) \in \Phi^{-1}(b)\}) = \mathbb{P}_X(\Phi^{-1}(b))$ , for  $b \in \mathcal{B}(C[0, 1])$ . One can compute the probability that  $W \in B$  for any ball  $B$  in  $\mathcal{B}(C[0, 1])$  if  $\mathbb{P}_X$  is known. How ?

B) Let  $X = (X_1, \dots, X_n) \in \mathbb{R}^n$  a random vector with distribution  $\mu_n$ . Associate with  $X$  the path  $Z := Z(X)$  defined by  $Z_i = X_1 + \dots + X_i$  for  $i \in \llbracket 1, n \rrbracket$ , and  $Z_0 = 0$ , and next consider the normalized version of  $Z$ , the random function  $z \in C[0, 1]$ , defined by

$$z_t = \frac{1}{c_n} \left( Z_{\lfloor nt \rfloor} + \{nt\} X_{\lfloor nt \rfloor + 1} \right) = \frac{1}{c_n} \left( Z_{\lfloor nt \rfloor} + \{nt\} (Z_{\lfloor nt \rfloor + 1} - Z_{\lfloor nt \rfloor}) \right) \text{ for } t \in [0, 1]. \quad (36)$$

The random function  $z$  interpolates piecewise linearly the sequence  $(Z_k)_{k \in \llbracket 0, n \rrbracket}$ ; there have been two rescaling: the time as been divided by  $n$ , the space by  $c_n$  (where  $(c_n)$  is a deterministic sequence in  $\mathbb{R}^*$ ). The random function  $z$  belongs to  $C[0, 1]$ . The map  $\Psi$ , which sends  $(x_1, \dots, x_n)$  to the function  $z$  defined as in (36), is continuous and is then measurable. The distribution of  $z$  is  $\mu_n \circ \Psi^{-1}$ .

• In the case where  $\mu_n = \mu \otimes \dots \otimes \mu$  (that is, the increments are i.i.d.) then for  $c_n = \sqrt{n\sigma_\mu^2}$  (where  $\sigma_\mu^2$  is the variance of  $\mu$ ), the corresponding  $z$ , will be called the normalized random walk.

We now, come back to our processes  $S_{\mathbf{T}}, H_{\mathbf{T}}, V_{\mathbf{T}}$  under the conditioning  $|\mathbf{T}| = n$ , and define some normalized version of them.

• Again  $\mathbf{p}$  is the offspring distribution; denote by

$\mathbf{s}_n$  the normalized version of  $S_{\mathbf{T}}$  under the conditioning by  $|\mathbf{T}| = n$ ; time rescaling  $=n$ ; space rescaling  $=\sqrt{n}$ ;

$\mathbf{h}_n$  the normalized version of  $H_{\mathbf{T}}$  under the conditioning by  $|\mathbf{T}| = n$ ; time rescaling  $=n - 1$ ; space rescaling  $=\sqrt{n}$ ;

$\mathbf{v}_n$  the normalized version of  $V_{\mathbf{T}}$  under the conditioning by  $|\mathbf{T}| = n$ ; time rescaling  $=2(n - 1)$ ; space rescaling  $=\sqrt{n}$ ;

$\mathbf{s}_n, \mathbf{h}_n, \mathbf{v}_n$  are called normalized DFQ, height process, contour process. Their distribution will be denoted as usual,  $\mathbb{P}_{\mathbf{s}_n}, \mathbb{P}_{\mathbf{h}_n}$  and  $\mathbb{P}_{\mathbf{v}_n}$ .

In the three last examples, we have associated with a vector in  $\mathbb{R}^j$  a piecewise linear path; the process was then just another representation of this vector, and then we were still in  $\mathbb{R}^j$ , in some sense. The next example reveals much more the deep richness of the distribution in  $C[0, 1]$ ;

B) The Wiener measure is a probability measure on  $(C[0, 1], \mathcal{B}(C[0, 1]))$  denoted by  $\mathbb{W}$  such that : if  $\mathbf{x}$  has distribution  $\mathbb{W}$ , the process  $(\mathbf{x}_t)_{t \in [0, 1]} = (\pi_t(\mathbf{x}))_{t \in [0, 1]}$  (called Brownian motion) satisfies :

•  $\mathbf{x}_t \sim \mathcal{N}(0, t)$ , for each  $t \in [0, 1]$ ,

• if  $0 \leq t_1 < t_2 < \dots < t_n \leq 1$ , then  $\mathbf{x}_{t_1} - \mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_n} - \mathbf{x}_{t_{n-1}}$  are i.i.d., and  $\mathbf{x}_t - \mathbf{x}_s$  has distribution  $\mathcal{N}(0, t - s)$  for any  $t \geq s$ .

The FDD of the Brownian motion  $\mathbb{P}_n \circ \pi_{t_1, \dots, t_k}^{-1}$  are defined by the second point. The density of  $\mathbb{W} \circ \pi_{t_1, \dots, t_k}^{-1}$  with respect to the Lebesgue measure on  $\mathbb{R}^k$  is given

$$f_{t_1, \dots, t_k}(y_1, \dots, y_k) = \prod_{i=1}^k \frac{\exp\left(-\frac{(y_i - y_{i-1})^2}{2(t_i - t_{i-1})}\right)}{\sqrt{2\pi(t_i - t_{i-1})}}$$

where  $t_0 = y_0 = 0$ .

Notice that this time, the knowledge of a finite number of coordinates  $\pi_t(\mathbf{x}) = \mathbf{x}_t$  is not sufficient to reconstitute  $\mathbf{x}$  as it was the case in the previous cases. The fact that such a measure  $\mathbb{W}$  on  $C[0, 1]$  having these FDD exist, is not at all obvious. There are several methods to show that it is indeed the case; we chose to prove that it exists as a *limit in distribution* of a sequence of measures in  $C[0, 1]$ .

## 5.6 Weak convergence in $C[0, 1]$

The aim of this part is to explain some elements about the weak convergence in  $C[0, 1]$ ; in particular, we discuss some criterions of convergence. We then show that rescaled random walks converge in distribution to  $\mathbb{W}$  in  $C[0, 1]$  (normalized random walk converges in distribution to the Brownian motion); then also that the other processes  $\mathbf{s}_n, \mathbf{h}_n, \mathbf{v}_n$  have a limit, the so called, normalized Brownian excursion (up to some scale constant).

The first idea is to try to understand the link between the weak convergence of  $\mathbf{x}_n$  to  $\mathbf{x}$  in  $C[0, 1]$ , and the convergence of FDD.

**Lemma 15** *Let  $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$  be probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$  such that  $\mathbb{P}_n \xrightarrow[n]{weak} \mathbb{P}$ . Then for any  $k, k \in \mathbb{N}^*$  any  $(t_1, \dots, t_k) \in [0, 1]^k$ ,  $\mathbb{P}_n \circ \pi_{t_1, \dots, t_k}^{-1} \xrightarrow[n]{d} \mathbb{P} \circ \pi_{t_1, \dots, t_k}^{-1}$*

**Proof .** The map  $\pi_{t_1, \dots, t_k}$  is continuous from  $C[0, 1]$  in  $\mathbb{R}^k$  (see Proposition 7 page 53).  $\square$

What about the converse? Let  $(\mathbb{P}_n, n \geq 1)$  be a sequence of probability measure on  $C[0, 1]$  such that for any  $k \in \mathbb{N}^*$  and any  $(t_1, \dots, t_k) \in [0, 1]^k$ ,  $\mathbb{P}_n \circ \pi_{t_1, \dots, t_k}^{-1}$  converges. Can we deduce that the sequence  $(\mathbb{P}_n)$  converges? The answer is NO! (see the exercise below).

**Exercise 11 :** For any  $n$ , take  $\mathbb{P}_n = \delta_{f_n}$  the Dirac measure on a deterministic function  $f_n \in C[0, 1]$ .

– Find an example where  $(f_n)$  converges point-wise to a non continuous function  $f$ . Shows that the FDD converges to those of the Dirac measure on  $f$ . But that  $\delta_{f_n}$  does not weakly converge to  $\delta_f$  in  $C[0, 1]$ , since  $\delta_f$  has not its support in  $C[0, 1]$ !

– find an example where  $(f_n)$  converges point-wise to  $f \equiv 0$  but that  $(f_n)$  does not converge uniformly to  $f$ . Again show that the FDD of  $\mathbb{P}_n$  converge to those of  $\delta_f$  but that the sequence  $(\mathbb{P}_n)$  does not converge weakly in  $C[0, 1]$  to  $\delta_f$ .



**Exercise 12 :** a) Let  $X_n$  be a sequence of real r.v., such that  $X_n \xrightarrow{\frac{d}{n}} X$ . Consider the process  $z_n$  defined by  $z_n(t) = \cos(X_n t)$  and  $z$  the process  $z(t) = \cos(X t)$  for  $t \in [0, 1]$ .

Is it true that  $z_n \xrightarrow{\frac{d}{n}} z$  in  $C[0, 1]$ ?

b) Let  $X^{(n)} = (X_1^{(n)}, \dots, X_k^{(n)})$  a r.v. in  $\mathbb{R}^k$ . Assume that  $X^{(n)} \xrightarrow{\frac{d}{n}} X = (X_1, \dots, X_k)$ . Let

$$\begin{aligned} z^{(n)}(t) &= Z_{[kt]}^{(n)} + \{kt\}(Z_{[(k+1)t]}^{(n)} - Z_{[kt]}^{(n)}) \text{ for } t \in [0, 1] \\ z(t) &= Z_{[kt]} + \{kt\}(Z_{[(k+1)t]} - Z_{[kt]}) \text{ for } t \in [0, 1] \end{aligned}$$

where  $Z^{(n)}(0) = Z(0) = 0$ ,  $Z^{(n)}(j) = X_1^{(n)} + \dots + X_j^{(n)}$  and  $Z(j) = X_1 + \dots + X_j$ ,  $1 \leq j \leq k$ .

Do we have  $z^{(n)} \xrightarrow[\frac{n}{n}]{\text{weak}} z$ ?

FDD characterize the measures on  $C[0, 1]$  but not the convergence of measures.

The problem is that convergence of FDD distributions corresponds (in a stochastic form) to the point-wise convergence, when the topology we are working with is the topology of uniform convergence. What may also happen, is that the FDD converges to a non-continuous limit (there exists some sequences of continuous functions converging point-wise to non continuous limits). Hence, to get the convergence of distributions in  $(C[0, 1], \mathcal{B}(C[0, 1]))$ , something else is needed.

**Definition :** Let  $\Pi$  be a family of probability measures on a Polish space  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ , equipped with its Borelian  $\sigma$ -field. We say that  $\Pi$  is relatively compact if any sequence of elements of  $\Pi$  contains a subsequence converging weakly (to a measure  $\mathbb{P}$  possibly not in  $\Pi$ ).

(More generally, relatively compact means with compact closure).

The limit of this subsequence (in the closure on  $\Pi$ ) is necessarily a probability measure on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ , as one can see in integrating the bounded continuous function equals to 1 on  $\mathcal{S}$  (this gives the weight of the total space).

**Proposition 9** Let  $(\mathbb{P}_n, n \geq 1)$  be a sequence of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ . If the FDD of  $\mathbb{P}_n$  converge to those of  $\mathbb{P}$  and if the family  $(\mathbb{P}_n, n \geq 0)$  is relatively compact, then  $\mathbb{P}_n \xrightarrow[\frac{n}{n}]{\text{weak}} \mathbb{P}$  (in the space of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ ).

[The convergence of  $\mathbb{P}_n \circ \pi_{t_1, \dots, t_k}^{-1}$  is defined as usual, that is, it is the convergence of a sequence of probability measures on  $\mathbb{R}^k$ ].

**Proof .** From any subsequence  $(\mathbb{P}_{n'})$  of  $\mathbb{P}_n$ , one can extract a converging subsequence  $\mathbb{P}_{n''}$ , converging to a measure  $\mathbb{Q}$  (possibly depending of this extracted subsequence). By hypothesis, the FDD of  $\mathbb{Q}$  are those of  $\mathbb{P}$ ; thus  $\mathbb{P} = \mathbb{Q}$  by Proposition 8. Hence  $\mathbb{P}_{n''} \xrightarrow[\frac{n}{n}]{\text{weak}} \mathbb{P}$ . One deduces of this that  $\mathbb{P}_n \xrightarrow[\frac{n}{n}]{\text{weak}} \mathbb{P}$  (indeed: if  $(x_n, n \geq 1)$  is a sequence of real numbers such that from any subsequence

$(x_{n'})$  one can extract a converging subsequence  $(x_{n''})$  which converges to  $x$  (for an  $x$  which does not depend on  $(x_{n'})$ ), then  $x_n \rightarrow x$  (such a sequence can not have two accumulation points in  $\overline{\mathbb{R}}$ ). Then, apply this to the sequence of real numbers  $x_n = \int f d\mathbb{P}_n$  for  $f$  bounded continuous: the limit is  $\int f d\mathbb{P}$ .  $\square$ .

**Proposition 10** *Let  $(\mathbb{P}_n, n \geq 1)$  be a sequence of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ . Assume that  $(\mathbb{P}_n, n \geq 1)$  is relatively compact, and that for any  $k \in \mathbb{N}^*$  and  $t_1 < t_2 < \dots < t_k$ ,  $\mathbb{P}_n \circ \pi_{t_1, \dots, t_k}^{-1}$  converges to a measure  $\mu_{t_1, \dots, t_k}$  on  $\mathbb{R}^k$ ; then  $\mathbb{P}_n$  weakly converges to a measure  $\mathbb{P}$  on  $(C[0, 1], \mathcal{B}(C[0, 1]))$  such that  $\mathbb{P} \circ \pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$ .*

**Proof .** From any subsequence  $(\mathbb{P}_{n'})$  of  $(\mathbb{P}_n)$ , extract a converging subsequence  $(\mathbb{P}_{n''})$ ; it then converges weakly to a certain  $\mathbb{P}$ , whose FDD are necessarily given by  $\mu$ . Then such a  $\mathbb{P}$  is unique, since the FDD characterize probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ . Hence  $\mathbb{P}_n \xrightarrow[n]{weak} \mathbb{P}$ .  $\square$

Hence, to prove that a sequence of probability measures on  $C[0, 1]$  weakly converges, two things have to be done :

- One has to prove the convergence of the FDD; the limiting FDD will be those of the limit (if the weak convergence indeed holds true).
- the relative compactness of the family  $(\mathbb{P}_n, n \geq 1)$  must be proved.

In most cases, the second point is much more difficult to handle than the first one.

**Definition :** A family  $\Pi$  of probability measures on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$  is tight, if for any  $\varepsilon > 0$ , there exists a compact  $K \subset \mathcal{S}$  such that for any  $\mathbb{P} \in \Pi$ ,  $\mathbb{P}(K) > 1 - \varepsilon$ .

First, let us state a simple result (which proof is a bit more complex than one could expect).

**Lemma 16** *If  $\Pi = \{\mathbb{P}\}$  is reduced to a single Borelian probability measure on a Polish space  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ , then  $\Pi$  is tight.*

**Proof .** Since  $\mathcal{S}$  is Polish, it is separable : there exists a dense subsequence  $(x_i, i \geq 1)$  in  $\mathcal{S}$ . For  $p \in \mathbb{N}^*$  and  $\varepsilon > 0$ , we have  $\bigcup_{n \geq 1} \overline{B}(x_n, 1/p) = \mathcal{S}$ . To do this we use here a property of probability measures: if  $A_N$  is a non-decreasing sequences of sets (for inclusion) such that  $\bigcup_n A_n = A$  then  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ ; indeed for disjoint measurable sets  $B_n$ ,  $\mathbb{P}(\bigcup B_n) = \sum \mathbb{P}(B_n)$ . Thus taking  $B_n = A_n \setminus \bigcup_{i < n} A_i$ , the  $B_i$  are disjoint,  $\bigcup_{i \leq n} B_i = \bigcup_{i \leq n} A_i$ , and thus  $\mathbb{P}(A_n) = \mathbb{P}(\bigcup_{i \leq n} B_i) \rightarrow 1$  since  $\mathbb{P}(\bigcup_i B_i) = \mathbb{P}(\Omega) = 1$ .

Take  $A_N(p) = \bigcup_{n=1}^N \overline{B}(x_n, 1/p)$ . The sequence  $A_N(p)$  converges to  $\mathcal{S}$ ; then if  $N := N_p$  is large enough  $\mathbb{P}(A_{N_p}(p) \geq 1 - \varepsilon/2^p)$ . And thus

$$\mathbb{P}(\mathcal{S} \setminus \bigcup_{n \leq N_p} \overline{B}(x_n, 1/p)) \leq \varepsilon/2^p.$$

Let  $K_\varepsilon = \bigcap_p \bigcup_{n \leq N_p} \overline{B}(x_n, 1/p)$ . One has  $\mathbb{P}(K_\varepsilon) \geq 1 - \varepsilon$ . Moreover  $K_\varepsilon$  is compact : to see this, since we are in a metric space, it suffices to show that any sequence in  $K_\varepsilon$  owns a converging subsequence in  $K_\varepsilon$ . Take a sequence  $(y_n, n \geq 1)$  in  $K_\varepsilon$ . For  $p$  fixed, this sequence has an infinite number of terms in the  $N_p$  balls  $\bigcup_{n \leq N_p} \overline{B}(x_n, 1/p)$  (since  $\bigcup_{n \leq N_p} \overline{B}(x_n, 1/p)$  contains  $K_\varepsilon$ ); therefore, the sequence  $(y_n)$  has an infinite number of terms in one of these balls, say  $B(x_{n_p}, 1/p)$  (for  $n_p \leq N_p$ ). In other words, we can build a subsequence of  $y$  having all its terms in  $B(x_{n_p}, 1/p)$ . By the diagonal procedure, one constructs from there a sequence converging in all balls  $B(x_{n_p}, 1/p)$ . Thus its limit belongs to  $\bigcap_p B(x_{n_p}, 1/p)$  which is included in  $K_\varepsilon$  (this set is reduced to a single point).  $\square$

**Theorem 7 (Prohorov)** *Let  $\Pi$  be a family of probability measures on a Polish space  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ .*

- 1) *If  $\Pi$  is tight, then it is relatively compact.*
- 2) *If  $\Pi$  is relatively compact, then it is tight.*

Admitted (See Billingsley, convergence of probability measures).

We are working on  $C[0, 1]$  (equipped with the topology induced by the distance  $d_\infty(f, g) = \|f - g\|_\infty$ , topology often called, topology of uniform convergence). In order to prove tightness for a family of probability measures on this space, it will be useful to know the following classical theorem, characterizing the relatively compact subsets of this space.

**Theorem 8 (Arzelà-Ascoli)** *A subset  $H$  of  $C[0, 1]$  is relatively compact if and only if*

$$\begin{cases} \text{For any } t \in [0, 1], \{g(t) \mid g \in H\} \text{ is bounded in } \mathbb{R}. \\ H \text{ is equicontinuous.} \end{cases}$$

(admitted) In the first point, “for any” can be replaced by, “there exists”.

The second point means  $\forall \varepsilon > 0, \forall t \in [0, 1]$ , there exists  $\delta > 0$  such that for all  $t' \in [0, 1]$ ,

$$|t - t'| \leq \delta \Rightarrow \sup\{|g(t) - g(t')| \mid g \in H\} \leq \varepsilon.$$

Hence, this condition, looks like to simple continuity, except that the same  $\delta$  is valid for all  $g$  of the family. The modulus of continuity of a function  $g$  in in  $C[0, 1]$  is

$$w_x(\delta) = \sup_{|s-t|<\delta} |g(s) - g(t)|.$$

A sufficient condition for equicontinuity is

$$\limsup_{\delta \downarrow 0} \sup_{g \in H} w_g(\delta) = 0, \tag{37}$$

which is equivalent to the uniform equicontinuity (continuity of  $f$  is  $\forall \varepsilon > 0$ , for all  $x \in [0, 1]$ ,  $\exists \delta > 0$ ,  $|x - y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \varepsilon$ . Here, the same  $\delta$  is valid for all  $x$ , and all  $g \in H$ .)

**Remark 1** Since  $[0, 1]$  is compact, a continuous function  $f$  on  $[0, 1]$  is also uniformly continuous. We say that  $f$  is uniformly continuous if,  $\forall \varepsilon > 0, \exists \delta > 0, |x - y| \leq \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ . The proof which proves that continuous implies uniform continuous, can be adapted easily to prove that equicontinuity implies uniform equicontinuity. Hence (37) is a criterion of equicontinuity.

**Theorem 9** Let  $(\mathbb{P}_n, n \geq 1)$  be a sequence of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ . If the two following conditions are satisfied then the sequence  $(\mathbb{P}_n, n \geq 1)$  is tight :

(i) For any  $\eta > 0$ , there exists  $a \in \mathbb{R}^+$  such that

$$\mathbb{P}_n(\{x \mid |x(0)| > a\}) \leq \eta \text{ for } n \geq 1.$$

(ii) For any  $\varepsilon > 0$  and  $\eta > 0$ , there exists  $\delta \in ]0, 1[$  and  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$ ,

$$\mathbb{P}_n(\{x \mid w_x(\delta) \geq \varepsilon\}) \leq \eta.$$

**Proof .** Let  $\varepsilon > 0$ , by (i) there exists  $a$  such that  $\mathbb{P}_n(\{x \mid |x(0)| > a\}) \leq \varepsilon/2$ .

By (ii), for every  $p \geq 1$ , one can find  $\delta_p > 0$  and  $n_p$  such that

$$\sup_{n \geq n_p} \mathbb{P}_n(\{x \mid w_x(\delta_p) \geq 1/p\}) \leq \varepsilon/2^{p+1}.$$

But also, for  $n \leq n_1$ , one can find  $\delta'_p > 0$  such that

$$\sup_{n < n_p} \mathbb{P}_n(\{x \mid w_x(\delta'_p) \geq 1/p\}) \leq \varepsilon/2^{p+1},$$

since, at  $n$  fixed,  $\mathbb{P}_n(\{x \mid w_x(1/m) \geq 1/p\})$  tends to 0 when  $m \rightarrow +\infty$ .

Hence, taking  $\tilde{\delta}_p = \inf\{\delta_p, \delta'_p\} > 0$ , we have

$$K_\varepsilon = \{x \mid |x(0)| < a, w_x(\tilde{\delta}_p) \leq 1/p \text{ for all } p \geq 1\}$$

satisfies  $\mathbb{P}_n(K_\varepsilon) > 1 - \varepsilon$  for every  $n \geq 0$  and  $K_\varepsilon$  is relatively compact; moreover, since  $K_\varepsilon$  is closed (intersection of closed sets  $\{x \mid |x(0)| < a\}$  and  $\{x \mid w_x(\tilde{\delta}_p) \leq 1/p\}$ ), it is then compact, and then according to Prohorov, the family  $(\mathbb{P}_n, n \geq 1)$  is tight.  $\square$

**Remark 2** (Modulus of continuity, via dyadic numbers)

We want to bound  $w_x(2^{-n})$ , by the help of some “paths on dyadic numbers”.

Let  $\Delta_n = \{q/2^n, 0 \leq q < 2^n\}$  and

$$U_n = \sup \left\{ \left| x\left(\frac{q}{2^n}\right) - x\left(\frac{q+1}{2^n}\right) \right|, 0 \leq q < 2^n \right\}.$$

Let  $t \in [0, 1]$ ;  $t_n = [t2^n]/2^n$  is the maximum element of  $\Delta_n$ , smaller than  $t$ . For  $t \in \Delta_m$  (with  $m > n$ ),

$$|x(t) - x(t_n)| \leq \sum_{n \leq p < m} |x(t_{p+1}) - x(t_p)| \leq \sum_{n \leq p < m} U_p.$$

(the  $t_{p+1}$ 's constitutes a path of dyadic approximation of  $t$ , becoming much thin each time; for  $p \geq m, t_{p+1} = t$ ). Let  $\Delta = \cup \Delta_n$  be the set of dyadic numbers in  $[0, 1]$ . For  $t \in \Delta$ ,

$$|x(t) - x(t_n)| \leq \sum_{p \geq n} U_p.$$

For  $\delta \leq 2^{-n}$  and for every  $s$  and  $t$  in  $\Delta$  such that  $|s - t| \leq \delta$ , we have

$$|x(s) - x(t)| \leq |x(s) - x(s_n)| + |x(s_n) - x(t_n)| + |x(t_n) - x(t)| \leq 3 \sum_{p \geq n} U_p.$$

Now, for a continuous function  $x$ , the modulus of continuity is

$$w_x(\delta) = \sup_{\substack{|s-t| \leq \delta \\ (s,t) \in \Delta^2}} |x(s) - x(t)|;$$

we then deduce that

$$w_x(2^{-n}) \leq 3 \sum_{p \geq n} U_p.$$

Next theorem gives the so-called ‘‘moment criterion’’ for tightness, often more easy to check than the previous one, since instead of dealing with modulus of continuity, it involved only a difference.

**Theorem 10** Consider  $(\mathbb{P}_n, n \geq 1)$  a sequence of probability measures on  $C[0, 1]$ . Let  $\mathbf{x}_n$  be  $\mathbb{P}_n$  distributed. If there exists  $\alpha > 1, \beta > 0$  and  $\gamma > 0$  such that for any  $s, t$  in  $[0, 1]$  and any  $n$

$$\mathbb{E}(|\mathbf{x}_n(s) - \mathbf{x}_n(t)|^\alpha) \leq \beta |s - t|^{\gamma+1}$$

and if moreover the sequence  $(\mathbf{x}_n(0))$  is tight, then  $(\mathbb{P}_n, n \geq 1)$  is tight.

(a reference: Theorem 12.3, Billingsley, convergence of probability measure).

**Proof .** We prove that assumption (ii) of the preceding theorem is fulfilled; this is sufficient, since (i) has been also assumed. Let  $U_p^{(n)} = \sup\{|x_n(\frac{q}{2^p}) - x_n(\frac{q+1}{2^p})|, 0 \leq q < 2^p\}$ ; we have

$$\mathbb{E}((U_p^{(n)})^\alpha) \leq 2^p \sup \left\{ \mathbb{E} \left( \left| x_n \left( \frac{q}{2^p} \right) - x_n \left( \frac{q+1}{2^p} \right) \right|^\alpha \right), 0 \leq q < 2^p \right\} \leq \beta 2^{-p\gamma}.$$

Since  $\mathbb{E}(\sup\{|X_1|, \dots, |X_k|\}) \leq \mathbb{E}(|X_1| + \dots + |X_k|) \leq k \sup_i \mathbb{E}(|X_i|)$ . Therefore,

$$\|U_p^{(n)}\|_\alpha \leq 2^{-p\gamma/\alpha} \beta^{1/\alpha}.$$

By remark 2, we have

$$w_{x_n}(2^{-p}) \leq 3 \sum_{q \geq p} U_q^{(n)}.$$

Hence

$$\|w_{x_n}(2^{-p})\|_\alpha \leq 3 \sum_{q \geq p} \|U_q^{(n)}\|_\alpha \leq 3 \sum_{q \geq p} 2^{-q\gamma/\alpha} \beta^{1/\alpha} \leq 3\beta^{1/\alpha} \frac{2^{-p\gamma/\alpha}}{1 - 2^{-\gamma/\alpha}}.$$

Let  $\varepsilon > 0$  and  $\eta > 0$  be fixed and  $\delta = 2^{-p}$  a parameter to be tuned. We have

$$\begin{aligned} \mathbb{P}_n(\{x, w_x(\delta) \geq \varepsilon\}) &\stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}(|w_x(\delta)|^\alpha)}{\varepsilon^\alpha} \\ &\leq \frac{(3\beta^{1/\alpha} \frac{2^{-p\gamma/\alpha}}{1-2^{-\gamma/\alpha}})^\alpha}{\varepsilon^\alpha}; \end{aligned}$$

we see that when  $p \rightarrow +\infty$ , this bound goes to 0, and then is smaller than  $\eta > 0$  for  $p$  large enough.  $\square$

**Proposition 11** *Assume that  $(\mathbf{x}_n)$  is a sequence of processes with values in  $(C[0, 1], \mathcal{B}(C[0, 1]))$  converging in distribution to  $\mathbf{x}$ . Let  $(\mathbf{y}_n)$  be a sequence of processes taking their values in  $(C[0, 1], \mathcal{B}(C[0, 1]))$  ( $\mathbf{y}_n$  is defined on the same probability space as  $\mathbf{x}_n$ ), such that*

$$\forall \varepsilon > 0, \quad \mathbb{P}(\|\mathbf{x}_n - \mathbf{y}_n\|_\infty \geq \varepsilon) \xrightarrow[n]{} 0.$$

*Then one has  $\mathbf{y}_n \xrightarrow[n]{d} \mathbf{x}$ .*

This result is true in all Polish spaces (it does not use the fact that we are working on  $C[0, 1]$ ).

**Proof .** According to the Portmanteau theorem, it suffices to show that for any bounded uniformly continuous function  $f$  (from  $C[0, 1]$  with values in  $\mathbb{R}$ ),  $\mathbb{E}(f(\mathbf{x}_n)) - \mathbb{E}(f(\mathbf{y}_n))$  tends to 0. Let  $\varepsilon > 0$  be fixed. Since  $f$  is uniformly continuous, there exists  $\delta > 0$  such that  $\|z_1 - z_2\|_\infty \leq \delta \Rightarrow |f(z_1) - f(z_2)| \leq \varepsilon$ . Hence

$$\begin{aligned} |\mathbb{E}(f(\mathbf{x}_n)) - \mathbb{E}(f(\mathbf{y}_n))| &\leq \mathbb{E}(|f(\mathbf{x}_n) - f(\mathbf{y}_n)| \mathbb{I}_{\|\mathbf{x}_n - \mathbf{y}_n\|_\infty \leq \delta}) + \mathbb{E}(|f(\mathbf{x}_n) - f(\mathbf{y}_n)| \mathbb{I}_{\|\mathbf{x}_n - \mathbf{y}_n\|_\infty > \delta}) \\ &\leq \varepsilon + \|f\|_\infty \mathbb{P}(\|\mathbf{x}_n - \mathbf{y}_n\|_\infty > \delta); \end{aligned}$$

the second term goes to 0, by hypothesis.  $\square$

**Note 5 :** This point is much important for the asymptotic behavior of  $\mathbf{h}_n, \mathbf{s}_n$  and  $\mathbf{v}_n$  : we saw that they were asymptotically at distance 0 (by Theorems 5 and 6). It will then be sufficient to obtain the convergence in distribution of one of them in  $C[0, 1]$  in order to get for free the convergence of the other ones.

**Lemma 17** *(A tightness criterion for the convergence of rescaled discrete processes) Consider  $(\mathbb{P}_n, n \geq 1)$  a sequence of probability measures on  $C[0, 1]$ , and let  $w_n \in \mathbb{P}_n$  distributed. If a.s.  $w_n$  is linear between the points  $(j/n, j = 0, \dots, n)$ , and if there exists some constants  $\alpha > 2, \beta > 0$  such that, for any  $n \geq 1$ , any  $s, t$  in  $[0, 1]$  such that  $ns$  and  $nt$  are integers,*

$$\mathbb{E}(|w_n(s) - w_n(t)|^\alpha) \leq \beta |s - t|^{\alpha/2}$$

*then, this is also true for all  $s, t$  in  $[0, 1]$ . Hence if moreover the family  $(\mathbb{P}_n \circ \pi_0^{-1}, n \geq 1)$  is tight, then so does  $(\mathbb{P}_n, n \geq 1)$ .*

(The proof is boring, and can be skipped).

**Proof .** Let  $s, t \in [0, 1]$ . Two cases arise :

- if  $\lfloor nt \rfloor = \lfloor ns \rfloor$ ,  $nt$  and  $ns$  are in the same piece  $[\lfloor nt \rfloor/n, \lfloor nt + 1 \rfloor/n]$ . One has

$$\begin{aligned} \mathbb{E}|w_n(s) - w_n(t)|^\alpha &= (n(t-s))^\alpha \mathbb{E}|w_n(\frac{\lfloor nt \rfloor}{n}) - w_n(\frac{\lfloor nt + 1 \rfloor}{n})|^\alpha \\ &= (n(t-s))^\alpha \max_k \mathbb{E}|w_n(k/n) - w_n((k+1)/n)|^\alpha \\ &\leq (n(t-s))^\alpha \beta |1/n|^{\alpha/2} \\ &\leq (n(t-s))^{\alpha/2} \beta | \frac{n(t-s)}{n} |^{\alpha/2} \\ &\leq \beta |t-s|^{\alpha/2} \end{aligned}$$

because  $n|t-s| \leq 1$ . The first inequality follows the fact that on the considered interval, the slope of  $w_n$  is  $n|w_n(\frac{\lfloor nt \rfloor}{n}) - w_n(\frac{\lfloor nt + 1 \rfloor}{n})|$

- if  $0 \leq \lfloor ns \rfloor \leq \lfloor nt \rfloor$ . We have

$$\begin{aligned} \mathbb{E}|w_n(s) - w_n(t)|^\alpha &= \mathbb{E}|w_n(s) - w_n(\frac{\lfloor ns \rfloor}{n})|^\alpha + \mathbb{E}|w_n(\frac{\lfloor ns \rfloor}{n}) - w_n(\frac{\lfloor nt \rfloor}{n})|^\alpha + \mathbb{E}|w_n(\frac{\lfloor nt \rfloor}{n}) - w_n(t)|^\alpha \\ &\leq \beta |t - \frac{\lfloor nt \rfloor}{n}|^{\alpha/2} + \beta |s - \frac{\lfloor ns \rfloor}{n}|^{\alpha/2} + \beta |\frac{\lfloor ns \rfloor}{n} - \frac{\lfloor nt \rfloor}{n}|^{\alpha/2} \\ &\leq C' |t-s|^{\alpha/2} \end{aligned}$$

for a constant  $C'$  which does not depend on  $n$ , neither on  $(s, t)$  (such a constant exists since  $\alpha/2 > 1$  : one wants to check that  $(|a|^{\alpha/2} + |b|^{\alpha/2} + |c|^{\alpha/2})^{2/\alpha} \leq C(|a| + |b| + |c|)$  for a constant  $C$ ; this is true since all norms are equivalent on  $\mathbb{R}^3$  and  $(a, b, c) \rightarrow |a| + |b| + |c|$  as well as  $(a, b, c) \rightarrow (|a|^{\alpha/2} + |b|^{\alpha/2} + |c|^{\alpha/2})^{2/\alpha}$  are norms).  $\square$

### 5.6.1 Definition and existence of the Brownian motion

**Proposition 12** *Let  $(X_i, i \geq 1)$  be a sequence of r.v. i.i.d, with distribution  $\mathcal{N}(0, 1)$ . Let  $W$  be the random walk defined by  $W_0 = 0$  and  $W_k = X_1 + \dots + X_k$  (for  $k \geq 1$ ). Let  $w_n$  be the rescaled interpolated random walk, defined by*

$$w_n(t) = \frac{W_{\lfloor nt \rfloor} + \{nt\}(W_{\lfloor nt + 1 \rfloor} - W_{\lfloor nt \rfloor})}{\sqrt{n}} \text{ for } t \in [0, 1].$$

*The sequence of measures  $(\mathbb{P}_{w_n}, n \geq 1)$  weakly converges in the set of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ ; its limit is called the Wiener measure  $\mathbb{W}$ .*

A process having distribution  $\mathbb{W}$  is called the Brownian motion.

**Proof . Finite dimensional convergence :** Let  $t_1 < \dots < t_k \leq 1$  be  $k$  real numbers in  $[0, 1]$ .

First, notice that for any fixed  $t$  in  $[0, 1]$

$$w_n(t) = w_n\left(\frac{\lfloor nt \rfloor}{n}\right) + \{nt\} \left( w_n\left(\frac{\lfloor nt + 1 \rfloor}{n}\right) - w_n\left(\frac{\lfloor nt \rfloor}{n}\right) \right).$$

and then,  $w_n(t) - w_n(\lfloor nt \rfloor/n)$  tends to 0 in probability, because  $w_n(t) - w_n(\lfloor nt \rfloor/n)$  is  $\mathcal{N}(0, \frac{1}{n}(t - \frac{\lfloor nt \rfloor}{n})^2)$  distributed. Therefore, the same holds for

$$(w_n(t_1), \dots, w_n(t_k)) - (w_n(\lfloor nt_1 \rfloor/n), \dots, w_n(\lfloor nt_k \rfloor/n))$$

which goes to the null vector of  $\mathbb{R}^k$  in probability. We study the limit law of the second vector (the first will have the same limit, if any, by the remark following Proposition 11). Consider the random vector  $w(n) = (w_n(\frac{\lfloor nt_1 \rfloor}{n}), \dots, w_n(\frac{\lfloor nt_k \rfloor}{n}))$ . Let  $\lambda_1, \dots, \lambda_k$  be fixed real numbers. The Fourier transforms of  $w(n)$  is

$$\Phi_{w(n)}(\lambda_1, \dots, \lambda_k) = \mathbb{E} \left( \exp \left( i \sum_{j=1}^k \lambda_j w_n(\lfloor nt_j \rfloor/n) \right) \right).$$

In writing  $w_n(\frac{\lfloor nt_j \rfloor}{n}) = \sum_{l=1}^j (w_n(\frac{\lfloor nt_l \rfloor}{n}) - w_n(\frac{\lfloor nt_{l-1} \rfloor}{n}))$  (where by convention,  $t_0 = 0$ ), we get

$$\begin{aligned} \Phi_{w(n)}(\lambda_1, \dots, \lambda_k) &= \mathbb{E} \left( \exp \left( i \sum_{j=1}^k \lambda_j \sum_{l=1}^j (w_n(\frac{\lfloor nt_l \rfloor}{n}) - w_n(\frac{\lfloor nt_{l-1} \rfloor}{n})) \right) \right) \\ &= \mathbb{E} \left( \exp \left( i \sum_{l=1}^k (\lambda_l + \dots + \lambda_k) (w_n(\frac{\lfloor nt_l \rfloor}{n}) - w_n(\frac{\lfloor nt_{l-1} \rfloor}{n})) \right) \right) \\ &= \prod_{l=1}^k \mathbb{E} \left( \exp \left( i (\lambda_l + \dots + \lambda_k) (w_n(\frac{\lfloor nt_l \rfloor}{n}) - w_n(\frac{\lfloor nt_{l-1} \rfloor}{n})) \right) \right) \end{aligned}$$

The last equality comes from the independence of the “pieces” of path, used in each sum (each sum using some distinct increments  $X_i$ ). The factor  $\mathbb{E} \left( \exp \left( i (\lambda_l + \dots + \lambda_k) (w_n(\frac{\lfloor nt_l \rfloor}{n}) - w_n(\frac{\lfloor nt_{l-1} \rfloor}{n})) \right) \right)$  converges to the Fourier transform of the Gaussian  $\mathcal{N}(0, t_l - t_{l-1})$  distribution taken in  $\lambda_l + \dots + \lambda_k$  (by the limit central theorem). Eventually, we get

$$(w_n(t_1), \dots, w_n(t_k)) \xrightarrow{d} (w(t_1), \dots, w(t_k))$$

where  $w(t_i) \sim N(0, t_i)$  and where the r.v.  $w(t_{i+1}) - w(t_i)$ 's are independent, and have law  $\mathcal{N}(0, t_{i+1} - t_i)$ ; therefore, we have proved that the FDD converge to those of  $\mathbb{W}$  (introduced page 55).

**Tension :** From Lemma 17, it suffices to bound  $\mathbb{E}|w_n(t) - w_n(s)|^\alpha$  for  $t = k/n$  and  $s = k'/n$  for  $c|t - s|^{\alpha/2}$  for a certain  $c > 0$ . We have

$$\begin{aligned} \mathbb{E}|w_n(t) - w_n(s)|^\alpha &= \frac{\mathbb{E}(|N(|k' - k|)|^\alpha)}{n^{\alpha/2}} = \frac{|k - k'|^{\alpha/2}}{n^{\alpha/2}} \mathbb{E}(|N(0, 1)|^\alpha) \\ &= |t - s|^{\alpha/2} \mathbb{E}(|N(0, 1)|^\alpha) \end{aligned}$$

where we have denoted by  $N(a)$  a r.v. with distribution  $N(0, a)$ . According to Theorem 10, choosing  $\alpha > 2$ , the sequence  $(\mathbb{P}_{w_n}, n \geq 1)$  is tight.  $\square$



**Theorem 11** (Donsker) Let  $(X_i, i \geq 1)$  be a sequence of i.i.d. r.v. having mean 0 and variance  $\sigma^2 \in ]0, +\infty[$ . Let  $W$  be the r.v. defined by  $W_0 = 0$  and  $W_k = X_1 + \dots + X_k$  (for  $k \geq 1$ ). Let

$$w_n(t) = \frac{W_{\lfloor nt \rfloor} + \{nt\}(W_{\lfloor nt+1 \rfloor} - W_{\lfloor nt \rfloor})}{\sigma\sqrt{n}} \text{ for } t \in [0, 1]$$

the normalized random walk and let  $\mathbb{P}_{w_n}$  its distribution. We have

$$\mathbb{P}_{w_n} \xrightarrow[n]{(weak.)} \mathbb{W},$$

in the space of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ .

**Proof .** A complete proof use the tightness control using the modulus of continuity (it can be found in Billingsley, convergence of probability measure).

We give a proof under the additional assumption that increments own some finite moments of order  $p$ , for a  $p > 2$  (hypothesis used to prove the tightness). The proof of the convergence of the FDD is identical to the previous one (try!) .

For the tightness, we will admit the following Lemma, very useful to bound the sum of i.i.d. random variables : (See Petrov's book, Sums of independent random variables, p68-69). If  $(X_i, i \geq 1)$  is a sequence of i.i.d. r.v. with mean 0, and having some moments order  $p \geq 2$ ; we have

$$\mathbb{E}|X_1 + \dots + X_m|^p \leq C(p)m^{p/2}$$

where  $C(p)$  is a constant which depends only on  $p$  and of the distribution of the  $X_i$ 's.

With this result, the end of the proof is easy. For  $s$  and  $t$  such that  $ns$  and  $nt$  are integer :

$$\mathbb{E}|w_n(t) - w_n(s)|^p = \frac{\mathbb{E}|W_{n(t-s)}|^p}{n^{p/2}} \leq C(p) \frac{(n(t-s))^{p/2}}{n^{p/2}} \leq C(p)(t-s)^{p/2}$$

By Lemma 17, the proof is finished (provided that  $p > 2$ ).  $\square$

## 5.7 Weak convergence of Lukaciewish walks

Consider  $W$  be a random walk whose increments are i.i.d., have distribution  $(\mu_i, i \geq -1)$  with  $\mu_i = p_{i+1}$  for all  $i \geq -1$ . As usual  $\mathbf{p}$  is assumed to have mean 1, variance  $\sigma_{\mathbf{p}}^2 \in (0, +\infty)$  (this is an unconditioned version of the DFQ associated with a GW tree with offspring distribution  $\mathbf{p}$ ). We denote now  $W^+$  the random variable  $W$  conditioned by  $\text{Hit}_{-1}(W) = n$  (for  $n$  in  $\text{PTS}_{\mathbf{p}}$  as defined in page 31). Denote

$$e_n(t) = \frac{W^+(\lfloor nt \rfloor) + \{nt\}(W^+(\lfloor nt+1 \rfloor) - W^+(\lfloor nt \rfloor))}{\sigma_{\mathbf{p}}\sqrt{n}} \text{ for } t \in [0, 1],$$

the normalized discrete Lukaciewish path. This is a.s. an element of  $C[0, 1]$ , and then  $\mathbb{P}_{e_n}$  is a probability measure on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ .

**Proposition 13** (Kaigh- 1976) When  $n$  goes to  $+\infty$  in  $\text{PTS}_{\mathbf{p}}$ , the sequence  $\mathbb{P}_{e_n}$  weakly converges to a measure  $\mathbb{P}_e$ , called the distribution of the normalized Brownian excursion, in the space of probability measures on  $(C[0, 1], \mathcal{B}(C[0, 1]))$ .

A process  $e$  having distribution  $\mathbb{P}_e$  is called the “normalized Brownian excursion”.

**Proposition 14** A.s.  $e(0) = e(1) = 0$ . For any  $0 < t_1 < \dots < t_k < 1$ , the distribution of  $(e(t_1), \dots, e(t_k))$  owns a density  $f_{t_1, \dots, t_k}$  with respect to the Lebesgue measure on  $\mathbb{R}^k$ . They are given as follows. First, for  $t \in (0, 1)$  fixed, the density of  $e(t)$  is

$$f_t(x) = 2 \frac{x^2 \exp(-\frac{x^2}{2t(1-t)})}{\sqrt{2\pi t^3(1-t)^3}} \mathbb{I}_{x \geq 0}.$$

Set

$$p_0(t, x) = \frac{x e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t^3/2}} \mathbb{I}_{x \geq 0} \quad \text{and} \quad p(t_1, x_1, t_2, x_2) = \frac{e^{-\frac{(x_1-x_2)^2}{2(t_2-t_1)}} - e^{-\frac{(x_1+x_2)^2}{2(t_2-t_1)}}}{\sqrt{2\pi(t_2-t_1)}} \mathbb{I}_{x_1 \geq 0} \mathbb{I}_{x_2 \geq 0};$$

the density of  $(e(t_1), \dots, e(t_k))$  is

$$f_{t_1, \dots, t_k}(x_1, \dots, x_k) = 2\sqrt{2\pi} p_0(t_1, x_1) \left( \prod_{i=1}^{k-1} p(t_i, x_i, t_{i+1}, x_{i+1}) \right) p_0(1-t_k, x_k)$$

Admitted.

Let see where these formula come from; we will see this on the following particular case : take for  $W$ , the random walk with i.i.d. increments, taking values  $\pm 1$  with proba.  $1/2$ . We will condition afterward by  $\text{Hit}_{-1}(W) = n$ , for an odd  $n$ . Notice that this will provide the uniform distribution on the set of Lukaciewish paths of size  $n$ , having step  $\pm 1$ .

Let us compute  $\mathbb{P}(W(\lfloor nt_1 \rfloor) = x_1, W(\lfloor nt_1 \rfloor) = x_2 \mid \text{Hit}_{-1}(W) = n)$ ; the computation with more points is very similar because random walks conditioned to be Lukaciewish paths are Markov processes. Under the uniform distribution, we only have to count the trajectories such that  $\{W(\lfloor nt_1 \rfloor) = x_1, W(\lfloor nt_1 \rfloor) = x_2, \text{Hit}_{-1}(W) = n\}$  and divide by the total number of trajectories. For this, we decompose the trajectories from  $\{W(\lfloor nt_1 \rfloor) = x_1, W(\lfloor nt_1 \rfloor) = x_2, \text{Hit}_{-1}(W) = n\}$  into three parts, on the intervals  $\llbracket 0, \lfloor nt_1 \rfloor \rrbracket$ ,  $\llbracket \lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor \rrbracket$ , and  $\llbracket \lfloor nt_2 \rfloor, n \rrbracket$ . Set by convention  $\binom{n}{k} = 0$  if  $k \notin \llbracket 0, n \rrbracket$ . On the following points we count the number of (restriction of paths) valid on each of the pieces described above.

- On  $\llbracket 0, \lfloor nt_1 \rfloor \rrbracket$ , the number of good trajectories is  $\frac{x_1 + 1}{\lfloor nt_1 \rfloor + 1} \binom{\lfloor nt_1 \rfloor + 1}{\frac{\lfloor nt_1 \rfloor + x_1 + 1}{2}}$ .
- on  $\llbracket \lfloor nt_2 \rfloor, n \rrbracket$ , the number of good trajectories is  $\frac{x_2 + 1}{n - \lfloor nt_2 \rfloor} \binom{n - \lfloor nt_2 \rfloor}{\frac{n - \lfloor nt_2 \rfloor + x_2 + 1}{2}}$
- on  $\llbracket \lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor \rrbracket$ , there are  $\binom{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor}{\frac{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor + x_2 - x_1}{2}} - \binom{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor}{\frac{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor - x_2 - x_1 - 2}{2}}$  The two first results

come from the following facts : for  $k \in \mathbb{N}^*$

$$\mathbb{P}(\text{Hit}_{-k}(W) = n) = \frac{k}{n} \mathbb{P}(W_n = -k).$$

To get the results we proceed as follows : add a first step  $(-1, -1), (0, 0)$  to the piece on  $\llbracket 0, \lfloor nt_1 \rrbracket \rrbracket$ . By turning back the time (on starting from  $(\lfloor nt_1 \rfloor, x_1)$  we consider this piece as a random walk hitting position  $-x_1 - 1$  at time  $\lfloor nt_1 \rfloor + 1$ . The number of such trajectories is  $(x_1 + 1) / (\lfloor nt_1 \rfloor + 1)$  multiplied by the number of trajectories going from  $(\lfloor nt_1 \rfloor, x_1)$  to  $(-1, -1)$ . There are  $\binom{\lfloor nt_1 \rfloor + 1}{\frac{\lfloor nt_1 \rfloor + x_1 + 1}{2}}$  such path.

The second formula can be computed with the same method (except that it is not useful to add a step).

The last formula is computed thanks to the reflection principle (we search the number of trajectories having a non negative minimum on  $\llbracket \lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor \rrbracket$ ). The reflection principle says that there are as many trajectories from  $a > 0$  to  $b > 0$  having length  $n$  that hits  $-1$  than those from  $a$  to  $-2 - b$  (why?).

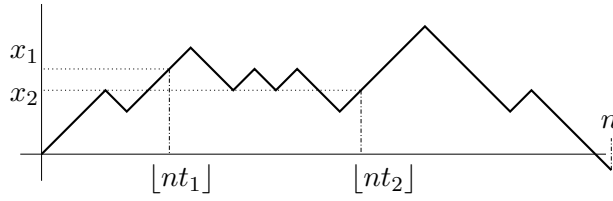


Figure 22: Decomposition of the excursion

Hence, we get

$$\begin{aligned} & \mathbb{P}(W(\lfloor nt_1 \rfloor) = x_1, W(\lfloor nt_2 \rfloor) = x_2 \mid \text{Hit}_{-1}(W) = n) \\ = & \frac{\frac{x_1+1}{\lfloor nt_1 \rfloor + 1} P_{\lfloor nt_1 \rfloor + 1}(x_1 + 1) (P_{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor}(x_2 - x_1 + 1) - P_{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor}(x_2 + x_1 + 2)) \frac{x_2+1}{n - \lfloor nt_2 \rfloor} P_{n - \lfloor nt_2 \rfloor}(x_2 + 1)}{\mathbb{P}(Z \in \mathbf{L}_n)} \end{aligned} \quad (38)$$

where the signs  $\lfloor \cdot \rfloor$  have been deleted so save some room. The sign  $P_k(l)$  is used as a notation for  $\mathbb{P}(W_k = l)$ . We already saw that  $\mathbb{P}(Z \in \mathbf{L}_n) \sim (\sqrt{2\pi})^{-1} n^{-3/2}$ . Observe the numerator. By the central local limit theorem, we have for  $t_1 \in ]0, 1]$  fixed,

$$|\sqrt{nt_1} P_{nt_1}(k) - \frac{1}{\sqrt{2\pi}} e^{-k^2/(2nt_1)}| \longrightarrow 0 \text{ unif. in } k.$$

Let  $t_1$  and  $t_2$  fixed,  $n \rightarrow +\infty$ ). We multiply the numerator and the denominator of (38) by  $\sqrt{nt_1} \sqrt{n(t_2 - t_1)} \sqrt{n(n - nt_2)}$  in order to use the local central limit theorem. We get the following approximation

$$\frac{(x_1 + 1) e^{-\frac{(x_1+1)^2}{2(nt_1+1)}} (e^{-\frac{(x_2-x_1+1)^2}{2n(t_2-t_1)}} - e^{-\frac{(x_2+x_1+2)^2}{2n(t_2-t_1)}}) (x_2 + 1) e^{-\frac{(x_2+1)^2}{2(n-nt_2)}}}{\sqrt{2\pi}^3 (\sqrt{2\pi})^{-1} n^2 (t_1(1-t_2))^{3/2} (t_2 - t_1)^{3/2}}$$

which is uniform in  $x_1$  and  $x_2$ .

In order to deduce the convergence from the local convergence we just obtained, we need a theorem (this is not so obvious, even if natural: we didn't prove the convergence of the cumulative distribution function... but, in some sense, the convergence of the density). We then use a Theorem (theorem 7.8 , Billingsley (Convergence of probability measures)) which states that the local convergence implies the convergence in distribution (under some mild hypothesis). The tightness can also be proved by bounding  $\mathbb{E}|x_n(s) - x_n(t)|^k$  (with integer power  $k$  larger than 3); there are other methods we admit it also.

**Exercise 13 :** Compute the distribution of the maximum  $M$  of the Brownian motion  $B$  on  $[0, a]$  (for  $a \in [0, 1]$ ), as well as the joint law of  $M$  and  $B_1$  (Brownian motion at time 1), using a discrete approach.

### 5.8 Limit of the processes $\mathbf{v}_n, \mathbf{h}_n, \mathbf{s}_n$

According to Kaigh's Theorem, we have for the normalized DFQ defined page 55 (the same holds for the normalized BFQ, of course);

$$\mathbf{s}_n \xrightarrow[n]{(d)} \sigma_{\mathbf{p}} e$$

here again, are assumed the criticality of the distribution of the increments, the existence of a variance, and this again, along subsequence in  $\text{PTS}_{\mathbf{p}}$ . We also used that if  $x_n \xrightarrow[n]{(d)} x$  in  $C[0, 1]$ , then for any deterministic constant  $\alpha$ ,  $\alpha x_n \xrightarrow[n]{(d)} \alpha x$  in  $C[0, 1]$ . If one assumes moreover the existence of exponential moments : according to Theorem 6 (page 46), we saw that

$$\mathbb{P} \left( \left\| \mathbf{s}_n - \frac{2}{\sigma_{\mathbf{p}}^2} \mathbf{v}_n \right\|_{\infty} \geq \varepsilon \right) \longrightarrow 0 \text{ for any } \varepsilon > 0;$$

we have much more than that, but this is sufficient by Proposition 11 to obtain

$$\mathbf{v}_n \xrightarrow[n]{(d)} \frac{2}{\sigma_{\mathbf{p}}} e \tag{39}$$

(this result, is due initially to Aldous (1993) who proved it with a very different method). By Theorem 5 (page 41) and Proposition 11, we get :

$$\mathbf{h}_n \xrightarrow[n]{(d)} \frac{2}{\sigma_{\mathbf{p}}} e \tag{40}$$

We have also as a direct consequence of Theorem 6, Theorem 5

$$\mathbf{h}_n - \mathbf{v}_n \xrightarrow[n]{(d)} 0, \quad \frac{\sigma_{\mathbf{p}}^2}{2} \mathbf{h}_n - \mathbf{s}_n \xrightarrow[n]{(d)} 0$$

where 0 is the null process on  $[0, 1]$ . One can also deduce from the work above that on  $C([0, 1]^2)$ ,  $(\mathbf{v}_n, \mathbf{s}_n) \xrightarrow[n]{(d)} (\frac{2}{\sigma_{\mathbf{p}}} e, \sigma_{\mathbf{p}} e)$  where the same excursion  $e$  appears as limit for  $\mathbf{v}_n$  and  $\mathbf{s}_n$ .

### 5.8.1 Consequences in terms of trees

In this section, we are talking about trees under  $\mathbb{P}_{\mathbf{p}}(\cdot \mid |\mathbf{T}| = n)$ . Even if not stated, we are taking limit in  $\text{PTS}_{\mathbf{p}}$ , and  $\mathbf{p}$  is assumed to have mean 1, a variance in  $(0, +\infty)$ , exponential moments.

#### Convergence of the height

The height of a tree  $T$  coincides with  $\max_k H_T(k)$ . Hence, let  $\mathbf{T}$  be a tree under  $\mathbb{P}_{\mathbf{p}}(\cdot \mid |\mathbf{T}| = n)$ ; let  $H_n := \max_k H_{\mathbf{T}}(k)$ . be its height. According to page 55,

$$H_n/\sqrt{n} \stackrel{(d)}{=} \max_{t \in [0,1]} \mathbf{h}_n(t).$$

Since the map

$$\begin{aligned} \max : C[0,1] &\longrightarrow \mathbb{R}^+ \\ f &\longmapsto \max\{f(x), x \in C[0,1]\} \end{aligned}$$

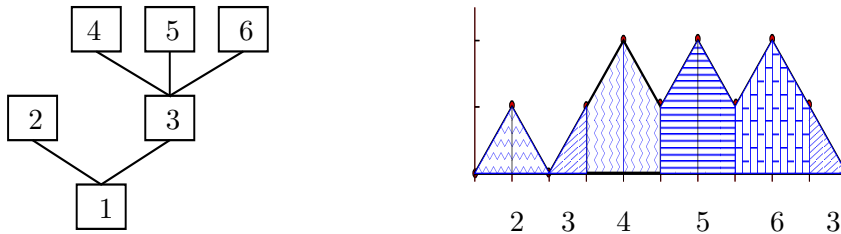
is continuous, and since  $\mathbf{h}_n \xrightarrow[n]{(d)} \frac{2}{\sigma_{\mathbf{p}}} e$ , therefore,

$$\frac{H_n}{\sqrt{n}} \stackrel{(d)}{=} \max \|\mathbf{h}_n\|_{\infty} \xrightarrow[n]{(d)} \frac{2}{\sigma} \max e.$$

(This does not give the limit, but the existence of a non degenerate limit, which is at the scale  $\sqrt{n}$ . The distribution of  $\max e$  is known.)

#### Total path length

It is defined for any tree  $T$  by  $\text{TPL}_T = \sum_{u \in T} |u|$ . In order to get the limit of  $\text{TPL}_{\mathbf{T}}$  we use the contour process (a similar work can be done using the height process). With any node  $u$  of  $\mathbf{T}$  different from the root  $\epsilon$ , associate the edge below  $u$ ; hence two steps of the contour process is associated with each node. If  $|u| = k$ , then the two steps are from position  $k - 1$  to  $k$ , and then from  $k$  to  $k - 1$ . Below this steps on the graph of  $V_{\mathbf{T}}$ , the area is  $2k - 1$ . Hence the area under the



contour process satisfies

$$\int_0^{2(|T|-1)} V_{\mathbf{T}}(x) dx = \sum_{u \in T, u \neq \epsilon} 2|u| - 1 = 2\text{TPL}(T) - (|T| - 1)$$

Take now  $\mathbf{T}$  under  $\mathbb{P}_{\mathbf{p}}(\cdot \mid |\mathbf{T}| = n)$ , and recall the normalization which defined  $\mathbf{v}_n$ , page 55. Since we have  $\int_0^{2(n-1)} V_{\mathbf{T}}(x) dx = 2(n-1)n^{1/2} \int_0^1 \mathbf{v}_n(t) dt$ ; therefore  $\frac{\int_0^{2(n-1)} V_{\mathbf{T}}(x) dx}{2(n-1)n^{1/2}} = \int_0^1 \mathbf{v}_n(t) dt$ , and since the map

$$\begin{aligned} \max : C[0, 1] &\longrightarrow \mathbb{R}^+ \\ f &\longmapsto \int_0^1 f(x) dx \end{aligned}$$

is continuous, we have  $\int_0^1 \mathbf{v}_n(t) dt \xrightarrow{\frac{d}{n}} \frac{2}{\sigma_{\mathbf{p}}} \int_0^1 e(t) dt$ ; this says that

$$\frac{2\text{TPL}(\mathbf{T}) + 1}{2(n-1)n^{1/2}} \xrightarrow{\frac{d}{n}} \frac{2}{\sigma_{\mathbf{p}}} \int_0^1 e(t) dt.$$

We deduce from this (exercise) that

$$\frac{\text{TPL}(\mathbf{T})}{n^{3/2}} \xrightarrow{\frac{d}{n}} \frac{2}{\sigma_{\mathbf{p}}} \int_0^1 e(t) dt.$$

Again, the distribution of the height under the Brownian excursion is known.

### Height of nodes

Take a random tree  $\mathbf{T}$  (with size  $n$ ) as above, and let  $u_{\mathbf{T}}(1), \dots, u_{\mathbf{T}}(n)$  be its nodes sorted according to the lexicographical order. Take  $0 < s_1 < \dots < s_k < 1$ ,  $k$  fixed. Consider the  $k$ -tuple

$$(|u(\lfloor ns_1 \rfloor)|, \dots, |u(\lfloor ns_k \rfloor)|)$$

giving the height of the nodes  $u(\lfloor ns_1 \rfloor), \dots, u(\lfloor ns_k \rfloor)$  in  $\mathbf{T}$ .

We have (using  $\mathbf{h}_n(t) = n^{-1/2} H_{\mathbf{T}}((n-1)t)$ , and  $H_{\mathbf{T}}(k) = |u(k+1)|$ ), we get

$$\frac{1}{\sqrt{n}} (|u(\lfloor ns_1 \rfloor)|, \dots, |u(\lfloor ns_k \rfloor)|) \stackrel{(d)}{=} \left( \mathbf{h}_n \left( \frac{\lfloor ns_1 \rfloor - 1}{n-1} \right), \dots, \mathbf{h}_n \left( \frac{\lfloor ns_k \rfloor - 1}{n-1} \right) \right).$$

We then have (why ?)

$$\frac{1}{\sqrt{n}} (|u(\lfloor ns_1 \rfloor)|, \dots, |u(\lfloor ns_k \rfloor)|) \stackrel{(d)}{\xrightarrow{\frac{d}{n}}} \frac{2}{\sigma_{\mathbf{p}}} (e(s_1), \dots, e(s_k)).$$

Further, assume that you are interested instead by the height of the random node  $u(\lfloor nU \rfloor)$  where  $U$  is uniform on  $[0, 1]$  and is chosen independently of  $\mathbf{T}$ , (or having another distribution) instead (or several of these kinds of nodes, with some i.i.d. (or not)  $U_i$ ). Then this time, use that the pair  $(U, \mathbf{h}_n)$  converges in distribution to  $(U, \frac{2}{\sigma_{\mathbf{p}}} e)$  in probability measures on  $[0, 1] \times C[0, 1]$  (with associated product Borelian  $\sigma$ -fields). From there, we get

$$\frac{|u(\lfloor nU \rfloor)|}{\sqrt{n}} \xrightarrow{\frac{(d)}{n}} \frac{2}{\sigma_{\mathbf{p}}} e(U)$$

where  $U$  is independent of  $e$  (why?).

### Asymptotic metric properties of trees

Consider each tree  $T$  as a metric space: the distance  $d_T(u, v)$  between two nodes  $u$  and  $v$  is as usual, the number of edges between them. Can we say something about the asymptotic distance between these points? Instead, of writing  $d_T(u, v)$  which is not nice from a functional point of view, consider the function  $D_T$  defined on  $\llbracket 1, n \rrbracket^2$  by

$$D_T(k, l) := d_T(u_T(k), u_T(l));$$

it simply gives the distance in  $T$  between the  $l$ th and  $k$  nodes, according to the lexicographical order. Take now a random tree  $\mathbf{T}$ , as above, and  $s := (s_1, \dots, s_k)$  with the  $s_i$ 's as above. The distance matrix

$$M_T(s) := (D_T(\lfloor ns_i \rfloor, \lfloor ns_j \rfloor))_{i=1..k, j=1..k}$$

gives all the relative distance of the nodes  $u(\lfloor ns_i \rfloor)$ . On the height process  $H_T$ , we see that

$$|D_T(k, l) - (H_T(k) + H_T(l) - 2 \min\{H_T(x), x \in \llbracket k \wedge l, k \vee l \rrbracket\})| \leq 1.$$

(where  $k \wedge l$  (resp  $k \vee l$ ) is the minimum, (resp. maximum) between  $k$  and  $l$  (why?). Hence the matrix,  $n^{-1/2}M_{\mathbf{T}}(s)$  has the same limit (if any) as the matrix

$$M_n(s) = \left( \mathbf{h}_n \left( \frac{\lfloor ns_i \rfloor - 1}{n - 1} \right) + \mathbf{h}_n \left( \frac{\lfloor ns_j \rfloor - 1}{n - 1} \right) - 2 \min\{\mathbf{h}_n(x) \mid x \in \llbracket i \wedge j, i \vee j \rrbracket\} \right)_{i=1..k, j=1..k}$$

But this matrix converges in distribution in the set of  $k \times k$  matrices  $\text{Mat}(k)$  with real coordinates (see these matrices as elements of  $\mathbb{R}^{k^2}$ , and equipped with the corresponding Borelian  $\sigma$ -field). The application  $\Psi_s$  from  $C[0, 1]$  onto  $\text{Mat}(k)$  defined by

$$\begin{aligned} \Psi_s : C[0, 1] &\longrightarrow \text{Mat}(k) \\ f &\longmapsto \Psi_s(f) := (f(s_i) + f(s_j) - 2 \min\{f(x) \mid x \in \llbracket s_i \wedge s_j, s_i \vee s_j \rrbracket\})_{i=1..k, j=1..k} \end{aligned}$$

is continuous. Therefore, it is not difficult to prove that

$$M_n(s) \xrightarrow[n]{d} \Psi_s \left( \frac{2}{\sigma_p} e \right),$$

in the set of Borelian probability measure on  $\text{Mat}(k)$ .

### 5.9 Convergence of trees

Here, we will present the main ideas, leading to a notion of convergence of trees. The reader must understand that something arbitrary must be bring into play. This is the general notion of trees that will be involved, together with the used topology of convergence : a notion that encompasses discrete trees and their limits is needed. Indeed, all objects have to be considered as element of the same topological space, in order to talk of convergence (and of course, it is simpler if this metric space is Polish, in order to have access to simple notions/tools related to weak convergence).

In the history, three embeddings have been used (as far as I know). The first one, historically used by Aldous (91) in his series of paper “continuum random tree 1,2, and 3” consists in an embedding in  $l^1$  (a tree is seen as a subset of  $l^1$ , the space of sequences, with a finite sum). Aldous deduces the convergence of the contour process from a nice but complicated construction, showing first the convergence of trees.

If a direct proof for the convergence of the contour of trees is known, another approach of the convergence to a continuum random tree can do provided (this is what we will do afterwards). More recently, for various reasons, another topology appeared : the Gromov-Hausdorff topology on compact metric spaces. The idea is to see trees as compact metric spaces (they are, as we saw in the last paragraph, compact metric spaces). Gromov defined a metric on the set of compact metric spaces (in fact, two compact metric spaces are considered to be equal if there exists an isometry which sends the first one on the second one; the distance given by Gromov, is a distance on equivalence classes of metric spaces). It is then possible to show that rescaled GW with  $n$  nodes, has a limit in distribution for the associated notion of weak convergence (this is also a consequence of the convergence of the contour process). We send the interested reader to the lecture notes “Random trees and applications” by Jean-François Le Gall (on its web page <http://www.dma.ens.fr/~legall/indexbis.html>) for complements. Here we will present a soft version of convergence to a continuum random trees, which is the richest one, when only ordered trees are considered.

This construction is due to JF Le Gall, and raises entirely on the convergence  $v_n \xrightarrow[n]{(d)} \frac{2}{\sigma_p} e$ . The idea is to define a tree from a function, as in the discrete case, the contour process encodes a tree.

Let  $C^+[0,1]$  be the set of continuous functions  $g$  from  $[0,1]$  with values in  $\mathbb{R}^+$ , such that  $g(0) = g(1) = 0$ . With each function  $g$  in  $C^+[0,1]$ , we define an equivalence relation in  $[0,1]$  by

$$x \underset{g}{\sim} y \iff g(x) = g(y) = \check{g}(x,y),$$

where

$$\check{g}(x,y) = \inf\{g(u) \mid u \in [x \wedge y, x \vee y]\}.$$

Let  $E_g = [0,1] / \underset{g}{\sim}$  be the set of equivalence classes modulo  $\underset{g}{\sim}$ ; this set is considered as a tree, its elements being considered as the nodes. We denote by  $F_g$  the canonical surjection from  $[0,1]$  into  $E_g$

$$\begin{aligned} F_g : [0,1] &\longrightarrow E_g \\ x &\longmapsto F_g(x) = \dot{x} := \{y \mid y \in [0,1], x \underset{g}{\sim} y\} \end{aligned}$$

which associates with each real number  $x$  the class  $\{y \mid y \in [0,1], x \underset{g}{\sim} y\}$ ; it is the continuous analogous of the function  $F_T$  (defined in page 23). The set  $E_g$  is totally ordered:  $\dot{x} < \dot{y}$  if  $\inf \dot{x} < \inf \dot{y}$  for the usual order on  $\mathbb{R}$ . The class  $\dot{0}$  is the root of  $E_g$ .

The set  $E_g$  is considered as a metric space. A distance on  $E_g$  is defined as follows. Let  $x$  and  $y$  be representative of  $\dot{x}$  and  $\dot{y}$ . The element  $\dot{z}$  in  $E_g$  defined by  $z \in [x,y]$  and  $g(z) = \check{g}(x,y)$  does not



depend on the chosen representative  $x$  and  $y$ . The point  $z$  is called the highest common ancestor of  $\dot{x}$  and  $\dot{y}$ . The distance between  $\dot{x}$  and  $\dot{y}$  is defined by

$$d_g(\dot{x}, \dot{y}) = d_{\mathcal{T}}(x, y) = g(x) + g(y) - 2g(z) = g(x) + g(y) - 2\check{g}(x, y).$$

The function  $g$  is considered as the contour process of  $E_g$  since

$$d_g(\dot{0}, \dot{x}) = g(x)$$

as in the discrete case. From there, we can define the notion of ancestor, descendent, subtrees branch, as in the discrete case (exercise). It can also be shown that  $E_g$  is also connected and contains no cycle.

Let  $K = \{t_g := (E_g, d_g) \mid g \in C^+[0, 1]\}$  be the set of such trees. We endow this space with the following metric which sends the topology from  $C[0, 1]$  onto  $K$  : the distance in  $K$  between  $t_g$  and  $t_f$  is  $d_K(t_g, t_f) = \|g - f\|_{\infty}$ . This makes of  $K$  a Polish space, and of course, if a sequence of function  $(f_n)$  converges in  $C[0, 1]$  to  $f$ , then  $t_{f_n}$  converges to  $t_f$  in  $K$  (it is the same metric on both spaces). Also, weak convergence in both space are equivalent : if  $x_n \xrightarrow[n]{weak} x$  on  $C^+[0, 1]$  then  $t_{x_n} \xrightarrow[n]{weak} t_x$  in the set of Borelian probability measures on  $K$ .

Why this is indeed a convergence of trees? What means that two trees are close? Assume that  $\|f - g\|_{\infty} < \varepsilon$ , and that both  $f$  and  $g$  are in  $C^+[0, 1]$ . Consider  $\dot{x}$ , the class of  $x$ , in the two trees  $E_g$  and  $E_f$ . We have

$$|d_g(\dot{0}, \dot{x}) - d_f(\dot{0}, \dot{x})| \leq \varepsilon;$$

also a short computation (or a small picture) shows that

$$\max_{u, v \in [0, 1]} |d_g(\dot{u}, \dot{v}) - d_f(\dot{u}, \dot{v})| \leq 3\varepsilon;$$

hence, seen as metric spaces, the two trees  $E_g$  and  $E_f$  are close.

**Theorem 12** Consider  $\mathbf{p}$  a critical GW OD, having exponential moments. For  $n \rightarrow +\infty$  in  $\text{PTS}_{\mathbf{p}}$ , we have for  $\mathbf{v}_n$  (the normalized contour process as defined on page 55),  $t_{\mathbf{v}_n} \xrightarrow[n]{weak} t_{\frac{2}{\sigma_{\mathbf{p}}}e} = \left(E_e, \frac{2}{\sigma_{\mathbf{p}}}d_e\right)$  in the space of probability measures on  $K$ .

[This is hence just a consequence of  $v_n \rightarrow \frac{2}{\sigma_{\mathbf{p}}}e$ ]. Notice that the nodes in  $E_g$  and in  $E_{cg}$  are the same for any  $c > 0$ , and that  $d_{cg} = cd_g$ . The limiting tree which appears in the theorem, is then defined to be the tree associated to the normalized Brownian excursion, up to some factors. Aldous, in his word, called continuum random tree (CRT) the tree associated with  $2e$ . Notice again that the limit is the same whatever is the starting distribution  $\mathbf{p}$  provided that it is critical and has exponential moments (in fact, Aldous showed the result under the existence of the variance, only).

**Remark 3** • We should also mention that the present construction can be completed as follows. The tree  $t_g$ , additionally to its branching structure, can be equipped with a measure. The aim of this

measure is to provide an information on the “density of nodes” in the different regions of the tree . The idea, here, is to following : in the discrete case, it was possible to choose one node uniformly in the tree, because there were only a finite number of nodes, and this had a sense. On a tree on  $t_g$  for a general  $g$ , it is still possible, starting from a measure on  $[0, 1]$ , and considering the image measure on  $E_g$ . Aldous, shows that starting from the uniform measure on the set of nodes in the discrete case, the limiting corresponding measure on  $[0, 1]$  is the Lebesgue measure. This allows one to take one node at random on the continuum random tree (we did similar thing in the previous subsection, when we saw clearly, using  $H$ , that choosing a node uniformly on the discrete case, corresponds, at the limit to taking a uniform position on  $[0, 1]$ ).

- the passage to the limit for the Gromov-Hausdorff topology has for aim to provide a maybe less arbitrary construction of the limiting object (because it relies on a pre-existing topology). The idea is to show, as we seen in the previous subsection, that indeed, seen as metric spaces, trees converge weakly. When dealing with metric spaces, it is reasonable to identify isometric metric spaces. Hence, the Gromov-Hausdorff topology is a topology on equivalence classes of isometric compact metric spaces. In the framework we are dealing with, the proof is similar to that we have provided on  $K$  (because, the topology of Gromov-Hausdorff is quite weak)... but here, the satisfaction to be on roads already followed by other people, disappoints a bit since the cost is the lost of the tree order (only the partial order “to be ancestor” is kept).

- A question remains, what have we done when we have said that  $(t_{v_n})$  converges, instead of saying that rescaled trees converges? In fact, if we consider the discrete trees, as a discrete metric space, the passage to  $t_{v_n}$  has made of it a continuous space where some points on “the edges” have been transformed into points in the space. This difference can be quantified if one works on a space of metric spaces, equipped with the Gromov-Hausdorff topology (it goes to 0, with this topology). If we don't want to work with this topology, it should be kept in mind that the trees which converges is  $t_{v_n}$ , and then, to be careful when one desires to show convergence of functional of trees: it has to be shown that they can be expressed simply in term of  $t_{v_n}$ .

## 6 Brownian motion

We will here give some properties of the standard Brownian motion. To learn more, we recommend the book of Revuz & Yor (Continuous Martingales and Brownian motion), Karatzas & Shreve (Brownian motion and Stochastic calculus) and Breiman (Probability).

### 6.1 Brownian motion on $\mathbb{R}$

First, notice that we have defined the Brownian motion on  $[0, 1]$ . The same construction can be done, more generally on  $\mathbb{R}^+$ ; for this (in order to see the Brownian motion as a limit of rescaled random walk), another topology should be used : this is the topology of uniform convergence on all compact sets. A metric is given by

$$d = \sum_{k \geq 1} \frac{d_k}{2^k(1 + d_k)},$$

where  $d_k(f, g) = \sup\{|f(x) - g(x)|, x \in [0, k]\}$ . (This approach is explained in a French book: Dacuhna-Castelle & Dufflo).

**Exercise 14 :** Let  $f, f_1, f_2 \dots$ , be some functions defined from  $\mathbb{R}^+$ , with values in  $\mathbb{R}$ . Show that  
 ( the sequence  $(f_n)$  converge to  $f$  on all compact subset of  $\mathbb{R}^+$  )  $\Leftrightarrow$  (  $d(f, f_n) \rightarrow 0$  )

Hence, there exists a probability measure  $\tilde{W}$  on  $C[0, +\infty)$ , for which the canonical process  $B = (B_t)_{t \in [0, +\infty[}$  has the two following properties :

- $B_0 = 0$  a.s..
- the increments of  $B$  are independant ( $B_{t_k} - B_{t_{k-1}} \perp \dots \perp B_{t_2} - B_{t_1} \perp B_{t_1}$  for all  $t_1 < t_2 < \dots < t_k$ ), and the law of  $B_{t_j} - B_{t_{j-1}}$  is  $\mathcal{N}(0, t_j - t_i)$ .

**Note 6 :** Often, we express the distribution of the vector  $(B_{t_1}, \dots, B_{t_k})$  as :

For all  $t_1, \dots, t_k \in [0, +\infty)^k$ ,  $(B_{t_1}, \dots, B_{t_k})$  is a centered Gaussian vector with covariance function

$$\text{cov}(B_{t_i}, B_{t_j}) = t_i \wedge t_j.$$

Indeed, if  $t_i = t_j$ , this says that the variance of  $B_{t_i}$  is  $t_i$ ; taking into account the other facts, this means  $B_{t_i} \sim \mathcal{N}(0, t_i)$ . In the other hand, if  $t_i < t_j$  using that  $B_{t_j} = B_{t_i} + (B_{t_j} - B_{t_i})$  and the fact that  $B_{t_j} - B_{t_i}$  is independent from  $B_{t_i}$  and has distribution  $\mathcal{N}(0, t_j - t_i)$ ,

$$\text{cov}(B_{t_j}, B_{t_i}) = \text{cov}(B_{t_i} + (B_{t_j} - B_{t_i}), B_{t_i}) = t_i = t_i \wedge t_j.$$

Conversely, using that the FDD characterizes the distribution of continuous processes (also true on  $C[0, +\infty)$ ), any process having these distributions has distribution  $\tilde{W}$ .

**Proposition 15** Let  $B$  be a Brownian motion, we have :

- (i) (Markov property) For any  $s > 0$ , the process  $(B_{t+s} - B_s)_{t \geq 0}$  is a Brownian motion (independent from the family  $(B_u)_{u \leq s}$ ).
- (ii) (Symmetry) The process  $(-B_s)_{s \geq 0}$  is a Brownian motion.
- (iii) (Scaling) For any  $c > 0$ , the process  $(cB_{t/c^2})_{t \geq 0}$  is a Brownian motion (equivalently  $(\frac{B_{ct}}{\sqrt{c}})_{t \geq 0}$  is a Brownian motion).

**Proof .** In each case, we are dealing with centered Gaussian processes. In the case (i) and (iii), the covariance function can be easily computed : for  $s < t$

$$\begin{aligned} \text{cov}(cB_{s/c^2}, cB_{t/c^2}) &= \text{cov}(cB_{s/c^2}, cB_{s/c^2} + (cB_{t/c^2} - cB_{s/c^2})) \\ &= \text{cov}(cB_{s/c^2}, cB_{s/c^2}) + \text{cov}(cB_{s/c^2}, cB_{t/c^2} - cB_{s/c^2}) \\ &= c^2 \frac{s}{c^2} + 0 = s = s \wedge t \end{aligned}$$

in this last line, we used that  $B$  is a Brownian motion, has independent increments, and the known covariance function.

For (ii), it suffices to say that the Gaussian distribution is symmetrical.  $\square$

**Exercise 15 :** Let  $(B_t)_{t \in [0,1]}$  a Brownian motion of  $[0,1]$ . Prove that the process  $(B_{1-s} - B_1)_{s \geq 0}$  is also a Brownian motion on  $[0,1]$  (time reversal formula).

## 6.2 Irregularities of Brownian trajectories

**Theorem 13** (Dvoretzki, Erdős et Kakutani) Almost surely, the Brownian motion is nowhere differentiable on  $[0,1]$ .

This result is also true on  $\mathbb{R}^+$ , since the Brownian motion is in some sense, the concatenation of Brownian motions on  $[0,1]$ .

**Proof .** Let  $\beta > 0$  fixed; let  $x$  a function defined on  $[0,1]$ , differentiable at a point  $s$  and such that  $|x'(s)| \leq \beta$ . Then, there exists  $n_0$  such that  $\forall n \geq n_0$

$$|x(t) - x(s)| < 2\beta|t - s| \text{ for } |t - s| < 2/n.$$

Hence, for

$$A_n = \{x \mid \exists s \text{ s.t. } |x(s) - x(t)| \leq 2\beta|t - s|, \text{ for } |t - s| < 2/n\} \tag{41}$$

we have : the sequence  $A_n$  is increasing and  $A = \cup_{n \geq 1} A_n$  contains every function having at least in one point, a derivative with an absolute value smaller than  $\beta$ .

If  $(x, s)$  satisfies  $|x(s) - x(t)| \leq 2\beta|t - s|$ , for  $|t - s| < 2/n$ , the function  $x$  at point  $s$  is between

the two straight lines with slope  $2\beta$  and  $-2\beta$  and passing via  $(s, x(s))$ ; by a small computation, or a small figure, one see that

$$\text{for } k = \max\{j, j/n \leq s\} = \lfloor ns \rfloor,$$

$$y_k = \max \left\{ \left| x\left(\frac{k+2}{n}\right) - x\left(\frac{k+1}{n}\right) \right|, \left| x\left(\frac{k+1}{n}\right) - x\left(\frac{k}{n}\right) \right|, \left| x\left(\frac{k}{n}\right) - x\left(\frac{k-1}{n}\right) \right| \right\} \leq 6\beta/n.$$

Hence, if  $D_n = \{x \mid \exists y_k, y_k \leq 6\beta/n\}$  then  $A_n \subset B_n$ . To show that  $\mathbb{P}(A) = \mathbb{P}(\lim A_n) = \lim_n \mathbb{P}(A_n) = 0$  it suffices to establish that  $\lim_n \mathbb{P}(D_n) = 0$ . For this, write :

$$D_n = \bigcup_{k=1}^{n-1} \{x \mid y_k \leq 6\beta/n\},$$

which leads to

$$\mathbb{P}(B \in D_n) \leq \bigcup_{k=1}^{n-1} \mathbb{P} \left( \max \left\{ \left| B\left(\frac{k+2}{n}\right) - B\left(\frac{k+1}{n}\right) \right|, \left| B\left(\frac{k+1}{n}\right) - B\left(\frac{k}{n}\right) \right|, \left| B\left(\frac{k}{n}\right) - B\left(\frac{k-1}{n}\right) \right| \right\} \leq 6\beta/n \right)$$

By using that the Brownian motion has independent increments with same distribution (on interval of the same size), the probabilities in the last formula are equal. Using the formula giving the distribution of the maximum of independent r.v. we get :

$$\begin{aligned} \mathbb{P}(B \in D_n) &\leq n \mathbb{P}^3(|B(1/n)| \leq 6\beta/n) \\ &= n \left( 2 \int_0^{6\beta/n} \frac{\sqrt{n} e^{-\frac{nx^2}{2}}}{\sqrt{2\pi}} dx \right)^3 \leq n \left( 2 \frac{6\beta}{n} \max \left\{ \sqrt{n} \frac{e^{-\frac{n(6\beta/n)^2}{2}}}{\sqrt{2\pi}} \right\} \right)^3 = O(n^{-1/2}). \end{aligned}$$

Hence,  $\mathbb{P}(A) = 0$  and thus, a.s. the Brownian motion admits no derivative smaller than  $\beta$ . We write now  $A^{(\beta)}$  instead of  $A$ . The set of functions which admits at least a derivative in a least one point on  $[0, 1]$  is thus contained in  $\bigcup_{n \in \mathbb{N}} A^{(n)}$ . But  $\mathbb{P}(\bigcup_n A^{(n)}) = 0$ , since it is a countable union of negligible sets. This ends the proof.  $\square$

**Proposition 16** Let  $M_1 = \max\{B_t, t \in [0, 1]\}$ . We have  $M_1 \stackrel{d}{=} |\mathcal{N}(0, 1)|$ .

**Proof .** Let  $(W_k)_{k \in [0, n]}$  be a random walk with i.i.d. increments  $\pm 1$  with probability  $1/2$ . For  $r \geq 0$ , by the reflection principle, we have :

$$\begin{aligned} \mathbb{P} \left( \max_{k \in [0, n]} W_k \geq r \right) &= \sum_{j \in \mathbb{N}} \mathbb{P} \left( \max_{k \in [0, n]} W_k \geq r, W_n = j \right) \\ &= \sum_{j < r} \mathbb{P} \left( \max_{k \in [0, n]} W_k \geq r, W_n = j \right) + \sum_{j \geq r} \mathbb{P} \left( \max_{k \in [0, n]} W_k \geq r, W_n = j \right) \\ &= \sum_{j < r} \mathbb{P}(W_n = 2r - j) + \sum_{j \geq r} \mathbb{P}(W_n = j) \\ &= \sum_{j > r} \mathbb{P}(W_n = j) + \sum_{j \geq r} \mathbb{P}(W_n = j) \\ &= 2\mathbb{P}(W_n > r) + \mathbb{P}(W_n = r) \end{aligned}$$

We have for any  $\lambda > 0$ ,

$$\mathbb{P}\left(\frac{\max_{k \in \llbracket 0, n \rrbracket} W_k}{\sqrt{n}} \geq \lambda\right) - 2\mathbb{P}\left(\frac{W_n}{\sqrt{n}} \geq \lambda\right) \longrightarrow 0$$

point-wise, for all  $\lambda > 0$ . To conclude, we use  $\frac{\max_{k \in \llbracket 0, n \rrbracket} W_k}{\sqrt{n}} \xrightarrow{\frac{d}{n}} M_1$  (corollary of Donsker's theorem), and  $\frac{W_n}{\sqrt{n}} \xrightarrow{\frac{d}{n}} \mathcal{N}(0, 1)$  (Central limit theorem).

**Corollary 3** *Let  $[a, b]$  be an interval, included in  $[0, 1]$ , we have*

$$\max_{t \in [a, b]} B_t - B_a \stackrel{loi}{=} \sqrt{b-a} M_1.$$

**Proof .**  $(B_t - B_a)_{t \in [a, b]}$  is a Brownian motion, according to Proposition 15. Therefore

$$\max_{t \in [a, b]} B_t - B_a \stackrel{d}{=} \max_{t \in [0, b-a]} B_t.$$

According to Proposition 15 (iv),  $\sqrt{b-a} \tilde{B}_{t/(b-a)} = B_t$  is a Brownian motion. Thus

$$\begin{aligned} \max_{t \in [0, b-a]} B_t &\stackrel{d}{=} \max_{t \in [0, b-a]} \sqrt{b-a} \tilde{B}_{t/(b-a)} \\ &= \max_{t \in [0, 1]} \sqrt{b-a} \tilde{B}_t \stackrel{d}{=} \sqrt{b-a} M_1. \end{aligned}$$

As a by-product, we obtained that if  $M_t = \max\{B_s, s \in [0, t]\}$  (for  $t > 0$ ) then

$$M_t \stackrel{d}{=} \sqrt{t} M_1. \quad (42)$$

**Corollary 4** *Let  $a > 0$  and  $T_a = \inf\{t, B_t = a\}$  the hitting time of  $a$  by the Brownian motion. We have*

$$T_a \stackrel{d}{=} \frac{a^2}{(N(0, 1))^2}$$

**Proof .** Let  $(B_t)_{t \geq 0}$  be a Brownian motion. For  $a > 0$  and  $t > 0$ ,

$$\{T_a \leq t\} = \{M_t \geq a\}.$$

Hence

$$\mathbb{P}(T_a \leq t) = \mathbb{P}(M_t \geq a) = \mathbb{P}(\sqrt{t}|N(0, 1)| \geq a) = \mathbb{P}\left(\frac{a^2}{|N(0, 1)|^2} \geq t\right).$$

Therefore, clearly  $T_a$  has the same distribution as  $\frac{a^2}{|N(0, 1)|^2}$ .  $\square$

**Corollary 5** *A.s., all local maxima of  $B$  are different.*

Of course, the same property holds for the minima of Brownian motion (this will appear to have an application in terms of limiting tree).

**Proof .** Let

$$A = \left\{ x \mid x \in C[0, 1], \exists (a, b), (c, d), b < c \text{ s.t. } \max_{t \in (a, b)} x_t = \max_{t \in (c, d)} x_t \right\},$$

be the set of continuous functions having (at least) two equal local maxima. First, the set  $A$  is the same, if instead of taking all  $a, b, c$  and  $d$ , these points are taken in  $\mathbb{Q}$ . The result of this restriction is that only a countable pairs of intervals are to be tested. IF we can prove that for  $a, b, c, d, b \neq c$  rational numbers, we have

$$\mathbb{P} \left( \max_{t \in (a, b)} B_t = \max_{t \in (c, d)} B_t \right) = 0$$

one will be able to prove  $\mathbb{P}(B \in A) = 0$  since the countable union of negligible sets is still negligible. Since the evolution of the Brownian motion on  $[c, d]$  is independent on the part on  $[0, c)$ , and since the maximum on  $[c, d]$  has a density (conditionally on  $B_c$ ), then  $\mathbb{P}(\max_{t \in (a, b)} B_t = \max_{t \in (c, d)} B_t) = 0$  (one may condition first by the value of  $B_c$  if the argument seems not sufficient).  $\square$

Notice that the same argument works exactly the same on  $\mathbb{R}^+$  instead, since the only used argument is the countability of the number of intervals with rational extremities on  $\mathbb{R}^+$ .

The following theorem gives a very nice property of the Brownian motion around 0, and then by Markovianity, around each position.

**Proposition 17** (*Law of the Iterated Logarithm for the Brownian motion*) Let  $\phi(t) = \sqrt{2t \log \log(1/t)}$ . We have

$$\limsup_{t \downarrow 0} B_t / \phi(t) = 1 \text{ a.s.}$$

**Proof .** this is a proof in two steps.

a) For all  $\delta > 0$ ,

$$\mathbb{P} \left( \limsup_{t \downarrow 0} B_t / \phi(t) > 1 + \delta \right) = 0.$$

b) For all  $\delta > 0$ ,

$$\mathbb{P} \left( \limsup_{t \downarrow 0} B_t / \phi(t) > 1 - \delta \right) = 1.$$

**Proof of (a)** Let  $q \in ]0, 1[$ ; set  $t_n = q^n$ . Let us show that

$$C_n = \{x \mid x(t) > (1 + \delta)\phi(t) \text{ for at least one } t \in [t_{n+1}, t_n]\},$$

then  $\mathbb{P}(B \in C_n \text{ infinitely often}) = 0$ , that is, the probability that  $B \in \liminf C_n$ , that is that  $B$  satisfies infinitely often  $B_t > (1 + \delta)\phi(t)$  for at least one  $t \in [t_{n+1}, t_n]$  is 0. But, we have :

$$C_n \subset \{M_{t_n} > (1 + \delta)\phi(t_{n+1})\},$$

since  $\phi(t)$  increases with  $t$ . Thanks to the bound

$$\begin{aligned} \mathbb{P}\left(\frac{M_{t_n}}{\sqrt{t_n}} > x\right) &= \frac{2}{\sqrt{2\pi}} \int_x^{+\infty} e^{-z^2/2} dz \\ &\leq \frac{\sqrt{2}}{x\sqrt{\pi}} \int_x^{+\infty} z e^{-z^2/2} dz \\ &= \sqrt{\frac{2}{\pi}} \frac{e^{-x^2/2}}{x} \end{aligned}$$

one sees that if  $q \in ]0, 1[$  is chosen in such a way that  $(1 + \delta)^2 q > 1$  then,

$$\mathbb{P}(M_{t_n} \geq (1 + \delta)\phi(t_{n+1}))$$

is the term of a converging sum. By the Borel-Cantelli theorem, we deduce that the probability for  $C_n$  to be satisfied infinitely often is 0.

**proof of (b)** : Let  $q \in ]0, 1[$ ; let  $Z_n = B_{t_n} - B_{t_{n+1}}$ . The  $Z_i$  are independent. Assume for a moment that

$$\mathbb{P}(Z_n > (1 - \varepsilon)\phi(t_n) \text{ infinitely often}) = 1. \quad (43)$$

From (a), since  $-B$  is also a Brownian motion, with probability 1, there exists  $n_0$  such that for any  $n \geq n_0$ ,

$$B_{t_{n+1}} \geq -(1 + \varepsilon)\phi(t_{n+1});$$

indeed, this formula is false for a finite number of  $n$  only.

From  $B_{t_n} = Z_n + B_{t_{n+1}}$ , we get

$$\begin{aligned} B_{t_n} &\geq (1 - \varepsilon)\phi(t_n) - (1 + \varepsilon)\phi(t_{n+1}) \\ &= \phi(t_n) \left(1 - \varepsilon - (1 + \varepsilon) \frac{\phi(t_{n+1})}{\phi(t_n)}\right) \end{aligned}$$

infinitely often. But  $\phi(t_{n+1})/\phi(t_n) \rightarrow \sqrt{q}$ . Hence, for  $q$  and  $\varepsilon$  sufficiently small (we can choose them as wanted), we have

$$1 - \varepsilon - 2(1 + \varepsilon)\sqrt{q} \geq 1 - \delta,$$

and then (b) is established. It remains to establish (43). We have

$$\mathbb{P}(Z_n > (1 - \varepsilon)\phi(t_n)) = \mathbb{P}\left(N > \frac{(1 - \varepsilon)\phi(t_n)}{\sqrt{t_n - t_{n+1}}}\right),$$

for  $N \sim \mathcal{N}(0, 1)$ , by using the formula of scaling of the Brownian motion. We have

$$\frac{(1 - \varepsilon)\phi(t_n)}{\sqrt{t_n - t_{n+1}}} = \frac{1 - \varepsilon}{1 - q} \sqrt{2} \sqrt{\log(n \log 1/q)}. \quad (44)$$

It remains to find a good bound of the tail of the distribution of  $N$ . Write

$$\int_x^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \underset{+\infty}{\sim} \frac{e^{-x^2/2}}{x\sqrt{2\pi}};$$

which is obtained by an integration by part. After a simple computation, we get if  $(1 - \varepsilon)^2/(1 - q) < 1$  then  $\mathbb{P}(Z_n > (1 - \varepsilon)\phi(t_n))$  is the term of a diverging sum (it suffices for this to take  $q$  small enough). According to Borel-Cantelli, (43) holds true.  $\square$



**Corollary 6** *A.s., in any neighborhood of 0, the Brownian motion crosses the x-axis an infinite number of times.*

**Proof .** By symmetry

$$\liminf_{t \downarrow 0} B_t / \phi(t) = -1 \text{ a.s..}$$

Using the intermediate value theorem, the Brownian motion crosses 0, an infinite number of times in  $[0, t]$  for any  $t$ . Notice again, that since the Brownian motion is constant on no interval (if it were, its derivative would exist and be 0 on this interval), this result is not trivial.  $\square$

**Proposition 18** *Let  $(B_t)_{t \geq 0}$  be a Brownian motion on  $\mathbb{R}^+$ ; then the process  $\tilde{B}$ , defined by  $\tilde{B}_0 = 0, \tilde{B}_t = tB_{1/t}$  for  $t > 0$  is a Brownian motion.*

**Proof .** First it is a centered Gaussian process. We have for  $0 < s < t$ ,

$$\begin{aligned} \text{cov}(\tilde{B}_s, \tilde{B}_t) &= \text{cov}(sB_{1/s}, tB_{1/t}) \\ &= st \text{cov}(B_{1/s}, B_{1/t}) \\ &= st \left( \frac{1}{s} \wedge \frac{1}{t} \right) = st(1/t) = s = s \wedge t. \end{aligned}$$

We just show that on  $(0, +\infty)$ , the FDD were those of the Brownian motion; the continuity on  $(0, +\infty)$  is clear (continuous transform of continuous function). It remains to show the continuity at 0. We have to show that  $\lim_{t \rightarrow 0} tB_{1/t} = 0$  (that is, the limit exists, and is 0). We will show that  $\lim_{x \rightarrow +\infty} x^{-1}B_x = 0$ . First, according to the law of large numbers,

$$\lim_{\substack{n \rightarrow +\infty \\ n \in \mathbb{N}}} \frac{B_n}{n} = 0.$$

Set  $Z_k = \max_{t \in [0,1]} |B_{k+t} - B_k|$ . We have for  $t$  in  $[k, k+1]$ ,

$$\begin{aligned} \left| \frac{B_t}{t} - \frac{B_k}{k} \right| &= \left| \frac{B_t - B_k}{t} + \left( \frac{1}{t} - \frac{1}{k} \right) B_k \right| \\ &\leq \frac{Z_k}{k} + \frac{|B_k|}{k^2} \end{aligned}$$

It remains to control  $\frac{Z_k}{k}$ . The distribution of  $Z_k$  is equal to that of  $\max\{|B_t|, t \in [0, 1]\}$ . This r.v. admits a finite mean since  $\max |B| \leq \max B + \max -B$  and also of order 2 since  $\mathbb{E}((\max |B|)^2) \leq \mathbb{E}((\max B + \max -B)^2) \leq \mathbb{E}((\max B)^2) + \mathbb{E}((\max -B)^2) + 2\mathbb{E}((\max B)(\max -B))$ ; this last mean is bounded (according to Cauchy-Schwarz) by  $(\mathbb{E}((\max B)^2)\mathbb{E}((\max -B)^2))^{1/2}$  (recall that the distribution of  $\max B$  is known : it's the law of the absolute value of a normal r.v.. Thus

$$\mathbb{P}\left(\frac{Z_k}{k} \geq \varepsilon\right) = \mathbb{P}\left(\frac{Z_k^2}{k^2} \geq \varepsilon^2\right) \leq \frac{\mathbb{E}(Z_k^2)}{k^2 \varepsilon^2} = O(1/k^2).$$

Hence, by the Borel-Cantelli lemma,  $\frac{Z_k^2}{k^2}$  can be larger than  $\varepsilon$  only a finite number of times (almost surely). Hence,  $\lim_{t \rightarrow +\infty} \frac{B_t}{t} = 0$  a.s..  $\square$

**Proposition 19** (Law of the iterated logarithm in  $+\infty$ )

$$\limsup_{t \rightarrow +\infty} \frac{B_t}{\sqrt{2t \log \log t}} \stackrel{a.s.}{=} 1$$

$$\liminf_{t \rightarrow +\infty} \frac{B_t}{\sqrt{2t \log \log t}} \stackrel{a.s.}{=} -1$$

**Proof .**

$$\begin{aligned} \limsup_{t \rightarrow +\infty} \frac{B_t}{\sqrt{2t \log \log t}} &= \limsup_{t \rightarrow +\infty} \frac{t\tilde{B}_{1/t}}{\sqrt{2t \log \log t}} \\ &= \limsup_{x \rightarrow 0} \frac{1/x\tilde{B}_x}{\sqrt{(2/x) \log \log 1/x}} = 1 \end{aligned}$$

according to law of the iterated logarithm at 0.  $\square$

**Proposition 20** (Normalized Brownian excursion and Brownian motion) Let  $(B_t)_{t \geq 0}$  be a Brownian motion. Let  $g = \max\{t \mid t < 1, B_t = 0\}$  and  $d = \inf\{t \mid t > 1, B_t = 0\}$ . Then :

- a.s.  $g < 1$  and  $d > 1$ .
- The process a.s. defined by  $\left| \frac{B_{g+(d-g)t}}{\sqrt{d-g}} \right|$  is a normalized Brownian excursion.

*Sketch of proof (due to Csaki & Mohanty) :* Take a random walk  $W$  with i.i.d. increments  $\pm 1$  with proba.  $1/2$ , having length  $Kn$  with  $K > 1$ . On this random walk we are interested in the part of the trajectory which straddles position  $n$ . Let  $g_n := \max\{k \leq n \mid W_k = 0\}$ , and  $d_n = \min\{k \geq n \mid W_k = 0\}$ . This last moment exists with proba.  $p_K^n$ , depending on  $K$ , for  $p_K^n \rightarrow 1$  when  $K \rightarrow +\infty$  (a computation is needed, here). Hence, for all  $\varepsilon > 0$ , there exists  $K$  such that  $d_n \leq Kn$  with probability  $1 - \varepsilon$  at least. Consider the set of trajectories for which  $d_n \leq Kn$ . Consider now the part on  $[g_n, d_n]$ . By symmetry, conditionally on  $g_n - d_n = l$ , the absolute value of the trajectory on the interval  $[g_n, d_n]$  is uniform among the paths of size  $l$ , positive strictly between  $[g_n + 1, d_n - 1]$ . Moreover, it is clear also, that for any  $\varepsilon > 0$ , there exists  $K$  such that for  $n$  large enough  $\mathbb{P}(K/n \leq d_n - g_n \leq Kn) \geq 1 - \varepsilon$  (that is, in the discrete case, the portion of the path which straddles  $n$  as length which goes to  $\infty$ , but linearly, in proba..

Take  $\lambda > 0$ . Conditioned on  $g_n - d_n = \lfloor \lambda n \rfloor$ , and after normalization, the (absolute value of the) path between  $g_n$  and  $d_n$  converges to the Brownian excursion for any  $\lambda > 0$ . Since this “conditioned” limit law (on  $d_n - g_n$  large, but not too much, and, on next, on  $d_n - g_n = \lambda n$ ) is the same for all  $\lambda > 0$ , we are very close to the results: it remains to let  $K$  goes to  $+\infty$ , and to show that the limit of the part between  $g_n$  and  $d_n$  indeed converges to the part between  $d$  and  $g$  on the Brownian motion.  $\square$

This proposition allows one to show that a lot of local properties of the Brownian motion are still true for the normalized Brownian excursion. Some of these properties can be translated in terms of the CRT (for example, since the local minima of the Brownian excursion are pairwise different, the CRT is an unary binary tree).

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Elements on probability theory in <math>\mathbb{R}</math></b>	<b>10</b>
2.1	Distribution on $\mathbb{R}$ : characterizations and convergences . . . . .	10
2.1.1	Characterizations of measures on $\mathbb{R}$ . . . . .	12
2.1.2	Characterization of the convergence of measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . . . . .	13
2.2	Exercise: analysis of the mean profile of binary search trees . . . . .	16
<b>3</b>	<b>Paths and trees : discrete considerations</b>	<b>18</b>
3.1	Planar trees . . . . .	18
3.2	Encoding of finite trees . . . . .	19
3.3	Encoding of trees by paths . . . . .	20
3.3.1	Queue representation . . . . .	20
3.4	Relation between $H_T$ , $V_T$ and $S_T$ . . . . .	24
3.4.1	Link between $H_T$ and $V_T$ . . . . .	24
3.4.2	Link between the DFQ and the height process . . . . .	25
3.5	Random walks . . . . .	27
3.6	A probability distribution on the set of trees . . . . .	28
3.7	Galton-Watson trees conditioned by the size . . . . .	29
3.8	From random walk to GW trees . . . . .	30
3.9	Example of Galton-Watson tree conditioned by the size . . . . .	33
3.10	An exponential family and a critical parameter . . . . .	35
3.11	Problem: Analysis of hashing with linear probing . . . . .	36
<b>4</b>	<b>Asymptotic comparison between the height process and DFQ</b>	<b>38</b>
4.1	Concentration principle : from random walks to Lukaciewish paths . . . . .	38
4.2	The theorem of Petrov . . . . .	39
4.3	Application : How many nodes with degree $k$ in a GW tree with $n$ nodes? . . . . .	40
4.4	Asymptotic comparison between the height process and the DFQ . . . . .	40
4.5	Study of the number of records of $W^\bullet$ : ladder variables . . . . .	42
4.5.1	Concentration of the number of records . . . . .	43
4.5.2	Back to right minima considerations on $W$ . . . . .	45
4.6	Relation between the contour process and the DFQ . . . . .	46
<b>5</b>	<b>Continuous processes and weak convergence in <math>C[0, 1]</math></b>	<b>49</b>
5.1	Probability measure on a Polish set . . . . .	49
5.2	Measure on Polish space. Characterization . . . . .	49
5.3	Random variables with a values in a Polish space . . . . .	52
5.4	Convergence in distribution in a Polish space . . . . .	52
5.5	Probability measures on $(C[0, 1], \mathcal{B}(C[0, 1]))$ . . . . .	53
5.6	Weak convergence in $C[0, 1]$ . . . . .	56

---

5.6.1	Definition and existence of the Brownian motion . . . . .	63
5.7	Weak convergence of Lukaciewish walks . . . . .	65
5.8	Limit of the processes $\mathbf{v}_n, \mathbf{h}_n, \mathbf{s}_n$ . . . . .	68
5.8.1	Consequences in terms of trees . . . . .	69
5.9	Convergence of trees . . . . .	71
<b>6</b>	<b>Brownian motion</b> . . . . .	<b>75</b>
6.1	Brownian motion on $\mathbb{R}$ . . . . .	75
6.2	Irregularities of Brownian trajectories . . . . .	76