# Random walks on $\mathbf{R}$ and ordered trees : First applications.

Frédéric Menous*

March 5, 2002

**Abstract**

We study some sets of probabilities associated to a random walk on $\mathbf{R}$. These sets are also called averages induced by diffusion in J. Ecalle's resummation theory. They are strongly related to ladder epochs for a random walk on the real axis : an average is a set of "weights" indexed by words of plus or minus signs and, for a given word $(\varepsilon_1, \ldots, \varepsilon_n)$ $(\varepsilon_i = \pm)$, the weight of the average is simply the probability to be on $\mathbf{R}^{\varepsilon_1}$ at time 1, on $\mathbf{R}^{\varepsilon_2}$ at time 2, ..., on $\mathbf{R}^{\varepsilon_n}$ at time $n$.

Such a coefficient can be decomposed in a sum of elementary coefficients which are indexed by *ordered trees and forests.*

The proof for the tree-decomposition is based on combinatorial methods, as well as a result of E. Sparre Andersen on a formal series associated to first ladder epoch probabilities. This result can also be recovered with the help of the tree-decomposition.

We also prove that this decomposition, which is valid for any random walk on $\mathbf{R}$, gives back, for the simple random walk, the well-known bijection between Dyck paths and Catalan trees.

## 1 Introduction.

There exists a vast literature about random walks on $\mathbf{R}$ (see, for example, [6]) and, among numerous results, we could cite one, by E. Sparre Andersen, which is based on combinatorial methods.

*UMR 8628, Laboratoire d'analyse harmonique, Département de Mathématiques, Batiment 425, Université Paris-Sud, Centre d'Orsay, 91405 Orsay Cedex. E-mail : Frederic.Menous@math.u-psud.fr

Let $(X_n)_{n\geq 1}$ be a sequence of independent, identically distributed, real random variables, with a common probability density $f \in L^1(\mathbf{R})$. We can define the associated random walk, that is the collection of random variables $(S_n)_{n\geq 1}$ :

$$\forall n \geq 1\,; \quad S_n = X_1 + \cdots + X_n$$

Considering the probabilities associated to the *first ladder epoch* :

$$\forall n \geq 1\,; \quad \tau_n = P(S_1 \leq 0, \ldots, S_{n-1} \leq 0, S_n > 0)$$

E. Sparre Andersen found that :

$$\log \frac{1}{1-\tau(s)} = \sum_{n=1}^{+\infty} \frac{s^n}{n} P(S_n > 0) \quad \text{with} \quad \tau(s) = \sum_{n=1}^{+\infty} \tau_n s^n \tag{1.1}$$

The reader could refer to [6] for a proof. On the one hand, the theorem remains valid if the variables $X_n$ share the same arbitrary distribution, but on the other hand, in the present conditions ($f \in L^1(\mathbf{R})$), the result holds whether the inequalities are strict or not.

As a corollary, if $f$ is even,

$$\tau(s) = 1 - \sqrt{1-s} = \sum_{n=1}^{+\infty} \frac{2}{4^n} ca_{n-1} s^n \qquad \left( \forall n \geq 0\ ;\ ca_n = \frac{(2n)!}{n!\,(n+1)!} \right) \tag{1.2}$$

These results illustrate how combinatorial methods lead to an "explicit" computation of probabilities.

Let us consider generalizations of the "first ladder epoch". We define the collection $\mathbf{m}_f$ of probabilities (or *weights*) :

$$\forall n \geq 1\,; \forall (\varepsilon_1, \ldots, \varepsilon_n) \in \{+, -\}\,; \quad \mathbf{m}_f^{\varepsilon_1, \ldots, \varepsilon_n} = \mathbf{P}(\varepsilon_1 S_1 > 0, \ldots, \varepsilon_n S_n > 0) \tag{1.3}$$

thus

$$\forall n \geq 1 \quad ; \quad \tau_n = \mathbf{m}_f^{\overbrace{-, \ldots, -}^{(n-1)\ \text{times}},+}$$

and $\mathbf{m}_f$ is the *average induced by the diffusion* $f$. A short explanation on this terminology is given in appendix A. The exponential law, that is, the "diffusion" $x \mapsto \frac{1}{2} e^{-|x|}$, yields an induced average **man** than can be explicitly computed :

$$\mathbf{man}^{\varepsilon_1, \ldots, \varepsilon_n} \equiv 4^{-n} ca_{n_1} ca_{n_2} \ldots ca_{n_s} (1 + n_s) \tag{1.4}$$

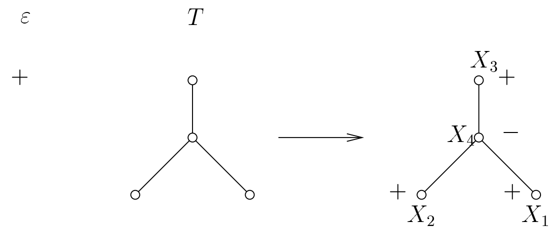
where the integers $n_1, n_2, \ldots, n_s$ denote the numbers of identical consecutive signs within the sequence $(\varepsilon_1, \ldots, \varepsilon_n)$ :

$$(\varepsilon_1, \ldots, \varepsilon_n) = (\pm)^{n_1} (\mp)^{n_2} \ldots (\varepsilon_n)^{n_s} \quad \text{(of course } n_1 + \cdots + n_s = n\text{)} \tag{1.5}$$

This average is the *Catalan* average, as the classical Catalan numbers $ca_n$ appear. This result is of course compatible with Andersen's theorem and points out that the combinatorial properties of the Catalan number could be helpful in the study of such random walks.

The Catalan numbers enumerate the ordered trees and we prove, in the first sections of this paper, that :

**1.** we can associate to each tree, combined with a given sign $\varepsilon = \pm$, a specific probability linked to the random walk.

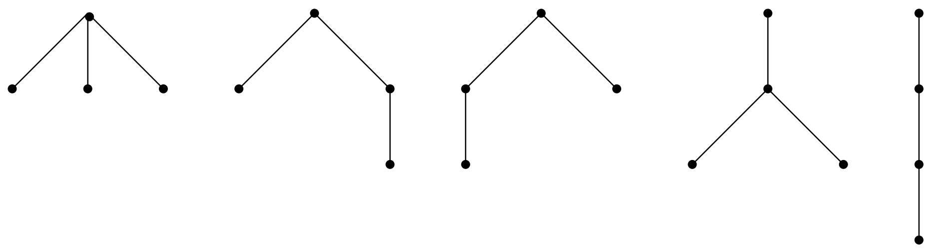



$$\mathbf{m}^{T,+} = P(+X_1 > 0, +X_2 > 0, -(X_4 + X_2 + X_1) > 0, +(X_3 + X_4 + X_2 + X_1) > 0)$$

**2.** For a given sequence $(\varepsilon_1, \ldots, \varepsilon_n)$, we can associate a set of ordered trees $\mathcal{F}$ such that $\mathbf{m}^{\varepsilon_1,\ldots,\varepsilon_n} = \sum_{T \in \mathcal{F}} \mathbf{m}^{T,\varepsilon_n}$. For example,

$$\mathbf{m}^{+++-} = \sum_{T \in Ca_4} \mathbf{m}^{T,-}$$

where $Ca_4$ is the set of rooted-ordered trees with four vertices.



The set $Ca_4$ ($\mathrm{Card}(Ca_4) = ca_3$).

The definitions, as well as the proofs, are enclosed in sections 2 to 7.

This tree decomposition extends the one developed in [12]. It had been discovered by considering diffusionss belonging to the vector space $\mathrm{Vect}_{\mathbf{C}}\{x \mapsto e^{-\lambda|x|}\ ;\ \lambda > 0\}$. Note that such vectors, called *linear exponential diffusions*, are not necessarily probability densities. Nonetheless, every weight of an average $\mathbf{m}_f^{\varepsilon_1,\ldots,\varepsilon_n}$ is an integral on a given measurable domain of $\mathbf{R}^n$ : their definition can be extended to the case $f \in L^1(\mathbf{R})$ (see section 2). One of the most interesting properties of such diffusions is that the tree coefficients $\mathbf{m}^{T,\varepsilon}$ can be explicitly computed.

Sections 8 and 9 are devoted to applications in probability theory. We first prove Andersen's result with the help of tree decomposition. We also prove that this tree decomposition, when applied to the simple random walk, yields back the bijection between

Dyck paths and ordered trees, which underlies the well-known isomorphism between the positive excursions of the simple random walk and the geometric Galton-Watson process (see [1, 8, 13]). Further results shall be given in a forthcoming paper.

# Contents



# 2 Averages induced by a diffusion.

We remind here some definitions for mathematical objects associated to a diffusion. Starting with an integrable function, we build a family of coefficients that are indexed by sequences of + or − signs. A complete expository on the need for such objects can be found in [3, 10, 11].

## 2.1 Definition.

Let us consider an integrable function $f$ on $\mathbf{R}$, then,

**Definition 2.1** *The* ***generalized average m*** *induced by $f$ is the collection of weights :*

$$\boldsymbol{m} = \{\boldsymbol{m}^{\varepsilon_1,\dots,\varepsilon_n} \ , \ n \geq 1 \ , \ \varepsilon_i = \pm\} \tag{2.6}$$

*With :*

$$\forall n \geq 1 \ , \ \forall(\varepsilon_1, \ldots, \varepsilon_n) \in \{-,+\}^n \quad \boldsymbol{m}^{\varepsilon_1,\ldots,\varepsilon_n} = \int_{\boldsymbol{R}^n} \prod_{i=1}^{n} f(x_i)\sigma_{\varepsilon_i}(\check{x}_i)\, dx_i \tag{2.7}$$

*and*

$$\forall 1 \leq i \leq n \ , \ \check{x}_i = x_1 + \cdots + x_i \quad , \quad \sigma_{\varepsilon_i} = \boldsymbol{1}_{\boldsymbol{R}^{\varepsilon_i}}$$

**Remarks**

- If $f$ is non-negative and its integral is 1, then we recover the probability sets introduced above.
- If $f$ is even and its integral is 1 then we obtain a well-behaved average, see [3, 10, 11], which is useful in real resummation theory.
- When the integral of $f$ is non zero, the average is very close (and trivially related) to the average induced by $f$ *normalized* by its integral.
- If the integral of $f$ is zero, then $\mathbf{m}$ defines an *alien operator*, see [3, 10].
- We will use the identities $\sigma_+ + \sigma_- = Id_{\mathbf{R}}$ and $\sigma_+\sigma_- = 0$. This is "almost sure" and the coming identities should be understood this way. That's why we consider random walks with a probability density in $L^1(\mathbf{R})$. For an arbitrary random walk (and for the simple random walk), our results remain valid with one of the "non-symmetric" definitions :

  $$\sigma_+ = \mathbf{1}_{\mathbf{R}^+} \text{ and } \sigma_- = \mathbf{1}_{\mathbf{R}^{-*}}$$

  or

  $$\sigma_+ = \mathbf{1}_{\mathbf{R}^{+*}} \text{ and } \sigma_- = \mathbf{1}_{\mathbf{R}^-}$$

- We can, in every case, note $\mathbf{m}^{\emptyset} = \int_{\mathbf{R}} f(x)\, dx$ and then,

  $$\forall(\varepsilon_1, \ldots, \varepsilon_{n-1}) \in \{-,+\}^{n-1} \quad \sum_{\varepsilon_n = \pm} \mathbf{m}^{\varepsilon_1,\ldots,\varepsilon_n} = \mathbf{m}^{\varepsilon_1,\ldots,\varepsilon_{n-1}}.\mathbf{m}^{\emptyset}$$

There is also a family of "weighted functions" induced by $f$ :

**Definition 2.2** *For $n \geq 1$, and any sequence $(\varepsilon_1, \ldots, \varepsilon_n)$ of plus or minus signs, one can define the weighted function $f^{\varepsilon_1,\ldots,\varepsilon_n}$ by the following induction :*

$$\begin{cases} f^{\varepsilon_1} & = & f \,.\, \sigma_{\varepsilon_1} \\ f^{\varepsilon_1,\ldots,\varepsilon_n} & = & \left(f * f^{\varepsilon_1,\ldots,\varepsilon_{n-1}}\right) .\, \sigma_{\varepsilon_n} \quad \text{if } n \geq 2 \end{cases} \tag{2.8}$$

*and $*$ is the usual convolution in $L^1(\boldsymbol{R})$.*

Note that for any sequence $(\varepsilon_1, \ldots, \varepsilon_n)$ :

$$\mathbf{m}^{\varepsilon_1,\ldots,\varepsilon_n} = \int_{\mathbf{R}} f^{\varepsilon_1,\ldots,\varepsilon_n}(x)\, dx$$

In order to study the general averages and their weighted functions, we give some compact notations and introduce a new family of coefficients that plays a central role in the coming results.

## 2.2 The family $M$.

Let us consider first pairs in $\{+,-\} \times \mathbf{R}$, that is $\binom{\varepsilon}{x}$ $(\varepsilon = \pm,\ x \in \mathbf{R})$. For $n \geq 1$ and for a given sequence $\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}} = \binom{\varepsilon_1,\ldots,\varepsilon_n}{x_1,\ldots,x_n}$ in $(\{+,-\} \times \mathbf{R})^n$, the previous section suggests to define the coefficient :

$$M^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}}} = M^{\binom{\varepsilon_1,\ldots,\varepsilon_n}{x_1,\ldots,x_n}} = \sigma_{\varepsilon_1}(x_1)\sigma_{\varepsilon_2}(x_1 + x_2)\ldots\sigma_{\varepsilon_n}(x_1 + \cdots + x_n) \tag{2.9}$$

On the same way, if $\boldsymbol{x} = (x_1, \ldots, x_n)$ is a sequence in $\mathbf{R}^n$, then :

$$\left\{ \begin{array}{lllllll} F^{\boldsymbol{x}} & = & f(x_1)\ldots f(x_n) & & \|\boldsymbol{x}\| & = & x_1 + \cdots + x_n \\ & & & & & & \\ d\boldsymbol{x} & = & dx_1 \ldots dx_n & & l(\boldsymbol{x}) & = & n \end{array} \right. \tag{2.10}$$

We can thus write that, for a given finite sequence $\boldsymbol{\varepsilon}$ of plus or minus signs,

$$\mathbf{m}^{\boldsymbol{\varepsilon}} = \int_{\boldsymbol{x} \in \mathbf{R}^{l(\boldsymbol{\varepsilon})}} F^{\boldsymbol{x}} M^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}}} d\boldsymbol{x} \tag{2.11}$$

and, introducing the Dirac distribution $\delta$, see definition 2.2, for a sequence $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$ $(n \geq 1)$ :

$$\forall x \in \mathbf{R}\ , \quad f^{\boldsymbol{\varepsilon}}(x) = \int_{\boldsymbol{y} \in \mathbf{R}^{l(\boldsymbol{\varepsilon})}} F^{\boldsymbol{y}} M^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{y}}} \delta(x - \|\boldsymbol{y}\|) d\boldsymbol{y} \tag{2.12}$$

These representations for generalized averages and weighted functions suggest that many of their generic properties will stem from the study of the family $M$. This is the case for their *Tree-decomposition* formulas.

# 3 Ordered trees and forests.

We follow the terminology given in [15].

## 3.1 Ordered trees.

Ordered trees may be defined recursively as follows : An *ordered* (or plane) *tree* is a finite set of *vertices* such that :

**(a)** One specially designated vertex is called the root and

**(b)** the remaining vertices (excluding the root) are put in an ordered partition $(T_1, \ldots, T_m)$ of $m \geq 0$ disjoint non-empty sets $T_1, \ldots, T_m$, each of which is an ordered tree. The trees $T_1, \ldots, T_m$ are called *subtrees*.

Of course, a tree is also a graph, considering that, in the recursive definition, we can put edges between the root and the roots of the subtrees. The *size* of a tree is the number of its vertices. Note that a tree of size $n$ has exactly $n-1$ edges. For $n \geq 1$, $Ca_n$ is the set of ordered trees of size $n$, and we note $ca_{n-1}$ its cardinal. This notation is natural as the numbers $ca_n$ are the Catalan numbers :

$$ca_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!} \tag{3.13}$$

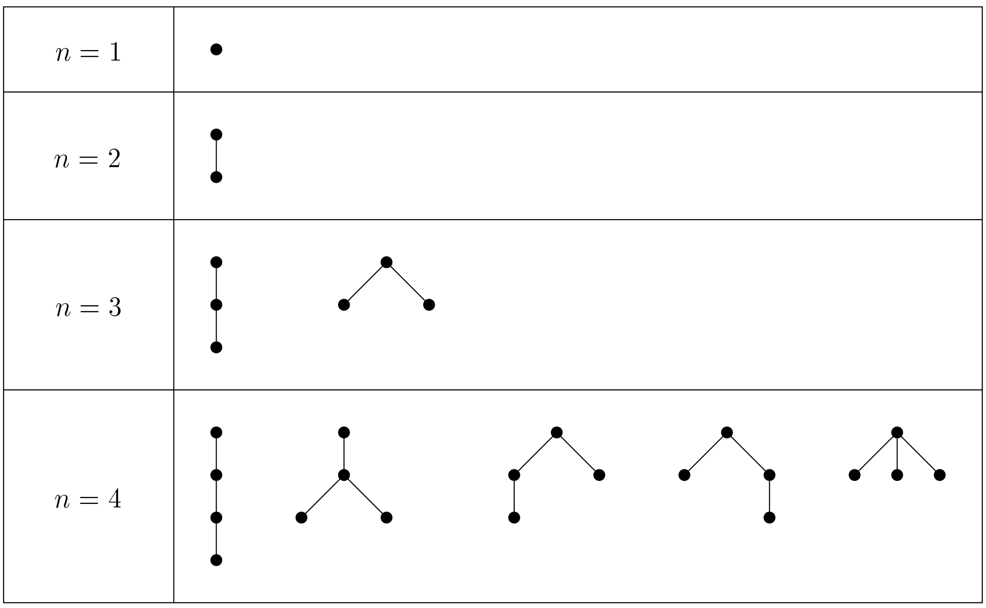
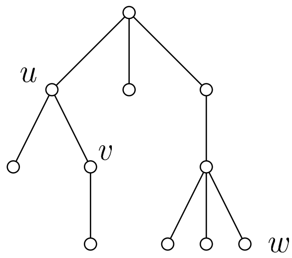



Figure 1: First sets $Ca_n$ of ordered trees

We assume that the definitions of the son (or successor) of a vertex, the father (or predecessor) of a vertex, as well as the notion of leafs (or endpoints), are obviously illustrated by the following figure.




$u$ is the father of $v$
$v$ is the son of $u$
$w$ is a leaf.

### 3.2 Forests.

A forest $F$ is a sequence $T_1, \ldots, T_m$ $(m \geq 1)$ of ordered trees. Its size is the sum of the sizes of its ordered trees. For $n \geq 1$, $Fa_n$ is the set of forests of size $n$ and its cardinal is $ca_n$. In fact, there is an obvious bijection between $Fa_n$ and $Ca_{n+1}$. One can transform a forest into an ordered tree by adding a common root to the sequence of trees composing the forest. Reciprocally, cutting the root of an ordered tree gives rise to a forest. Let us give the first sets of forests and suggest by dotted lines the corresponding ordered trees :

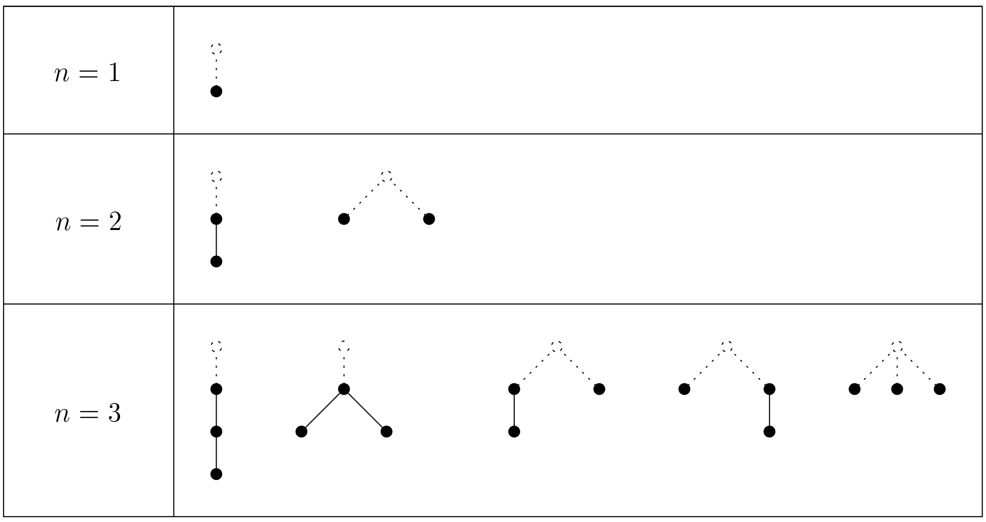



Figure 2: First sets $Fa_n$ of forests

Let us make some final remarks on the terminology for a forest $F$ $(= (T_1, \ldots, T_m))$ :

- For any vertex $v$ in $F$, we note $\rho(v)$ the root of the ordered tree containing this vertex. Thus, $\rho(F)$ is the root set of the forest $F$.
- The *height* $h(v)$ of a vertex $v$ is the number of edges between $v$ and its root $\rho(v)$.
- The forest $F$ naturally defines a partial order $\underset{F}{\succ}$ on the set of its vertices.For two vertices $u$ and $v$, $u \underset{F}{\succ} v$ means that $v$ belongs to the subtree having $u$ as root. For

example, the set of vertices has a maximal element iff $F$ contains only one tree. In this case, the maximal element is the unique root of $F$.

# 4 The tree-indexed family.

We define a family of coefficients similar to the family $M$ (see section 2.2) but indexed by a tree structure. To do so, we first give a way to label a tree or a forest by a sequence of variables.

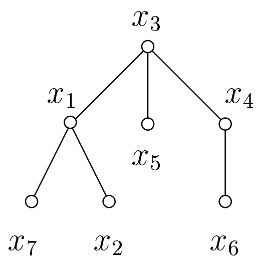
## 4.1 Labeled trees and forests.

For $n \geq 1$, consider $F$ a forest of size $n$ and $(x_1, \ldots, x_n)$ a sequence of variables. A *labeling* of $F$ by the sequence $(x_1, \ldots, x_n)$ is a bijection $\phi$ between the set of vertices of $F$ and $\{x_1, \ldots, x_n\}$. For a given labeling $\phi$, one defines the "beginning sums" of variables associated to $F$ and $\phi$ :

$\forall\, 1 \leq j \leq n\,;$

$$\overset{\curlyvee}{x}_j = \sum_{\phi^{-1}(x_j) \underset{F}{\succeq} u} \phi(u) \tag{4.14}$$

For example :



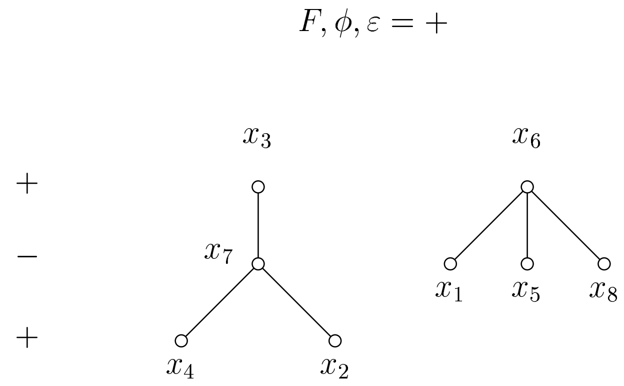

$$\begin{aligned} \overset{\curlyvee}{x}_1 &= x_1 + x_7 + x_2 \\ \overset{\curlyvee}{x}_4 &= x_4 + x_6 \\ \overset{\curlyvee}{x}_5 &= x_5 \end{aligned}$$

Figure 3: Some beginning sums for a given labeling.

## 4.2 Tree coefficients.

**Definition 4.3** *Let $F$ be a forest of size $n$, $\phi$ a labeling of $F$ by the sequence of variables $(x_1, \ldots, x_n) = \boldsymbol{x}$, for a given plus or minus sign $\varepsilon$, we define the coefficient*

$$M^{\varepsilon, \boldsymbol{x}}_{F,\phi} = \prod_{u \in F} \sigma_{\varepsilon(-1)^{h(u)}} \left( \overset{\curlyvee}{\phi(u)} \right) \tag{4.15}$$




Let us give an illustration of this definition :

$$
\begin{aligned}
M_{F,\phi}^{\varepsilon,\boldsymbol{x}} &= \sigma_+(x_3+x_7+x_4+x_2)\sigma_-(x_7+x_4+x_2)\sigma_+(x_2)\sigma_+(x_4) \\
&\quad \times \sigma_+(x_6+x_1+x_5+x_8)\sigma_-(x_1)\sigma_-(x_5)\sigma_-(x_8)
\end{aligned}
$$

A noticeable fact is that, if the forest $F$ is composed of several trees $(T_1,\dots,T_m)$ $(m \geq 2)$, then a labeling $\phi$ of $F$ by the sequence $\boldsymbol{x} = (x_1,\dots,x_n)$ induces for each tree $T_i$ a unique labeling $\phi_i$ by a subsequence $\boldsymbol{x}^i$ of $\boldsymbol{x}$. Thus,

$$
\begin{aligned}
M_{F,\phi}^{\varepsilon,\boldsymbol{x}} &= \prod_{u \in F} \sigma_{\varepsilon(-1)^{h(u)}} \left( \overset{\curlyvee}{\phi}(u) \right) \\
&= \prod_{\rho \in \rho(F)} \prod_{\rho \underset{F}{\succ} u} \sigma_{\varepsilon(-1)^{h(u)}} \left( \overset{\curlyvee}{\phi}(u) \right) \qquad (4.16) \\
&= \prod_{i=1}^{m} M_{T_i,\phi_i}^{\varepsilon,\boldsymbol{x}^i}
\end{aligned}
$$

This results is illustrated by the previous figure and there is no ambiguity on the ordering and the beginning sums : The restriction of $\underset{F}{\succ}$ to the set of vertices of a tree $T_i$ is precisely $\underset{T_i}{\succ}$.

We shall be able now to relate these coefficients to the initial family $M$ (see section 2.2).

# 5 Tree decomposition for the family $M$.

## 5.1 Some specific labeled forests $\mathcal{F}$.

For a *bisequence* $\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}} = \binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}$ in $(\{+,-\} \times \mathbf{R})^n$ $(n \geq 1)$, we give the inductive definition of a set $\mathcal{F}^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}}}$ that corresponds to a set of forests of size $n$, each of them having a given labeling $\phi_F$ by the sequence $(x_1, \dots, x_n)$.

If $\underline{n=1}$, $\mathcal{F}^{\binom{\varepsilon_1}{x_1}}$ is the unique forest of size 1, with its trivial labeling by $x_1$ (There is only one vertex to be labeled by $x_1$).

If $\underline{n \geq 2}$, Consider first the set $\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_{n-1}}{x_1,\dots,x_{n-1}}}$.

- If $\varepsilon_n \neq \varepsilon_{n-1}$, the elements $(F, \phi_F)$ of $\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}}$ are obtained by the following transformation. Pick an element $(F', \phi_{F'})$ of $\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_{n-1}}{x_1,\dots,x_{n-1}}}$ and consider $F'$ as the sequence of subtrees of a rooted tree with its root labeled by $x_n$. This gives birth to a forest of size $n$ (with only one tree), and its labeling by the sequence $(x_1, \dots, x_n)$ is naturally induced.

- If $\varepsilon_n = \varepsilon_{n-1}$ then, once again, consider first an element $(F', \phi_{F'})$ of $\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_{n-1}}{x_1,\dots,x_{n-1}}}$. We shall now give several ways to add a new vertex labeled by $x_n$. We remind that $F'$ is composed of trees $(T_1, \dots, T_m)$ $(m \geq 1)$. We do either of the following transformations to get an element of $\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}}$.

   1. Add the unique tree of size 1, labeled by $x_n$, to the sequence $(T_1, \dots, T_m)$, on its right.
   2. For $1 \leq i \leq m$, add the vertex labeled by $x_n$ as the rightest son of the root of $T_i$ and consider then $(T_{i+1}, \dots, T_m)$ as subtrees of the vertex labeled by $x_n$.

These operations are illustrated on figure 4.

In every case, for $n \geq 2$, starting with an element $(F, \phi)$ of $\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_{n-1}}{x_1,\dots,x_{n-1}}}$ the above construction yields a set of forests labeled by $(x_1, \dots, x_n)$ : $S^{x_n}_{\varepsilon_{n-1},\varepsilon_n}(F, \phi)$ and

$$\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}} = \bigcup_{(F,\phi)\in\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_{n-1}}{x_1,\dots,x_{n-1}}}} S^{x_n}_{\varepsilon_{n-1},\varepsilon_n}(F, \phi) \tag{5.17}$$

Moreover, one easily check that there is no repetition of the same tree (independently of the labeling) in such sets.

This rather obscure construction becomes natural in the results on tree decomposition for the family $M$.

## 5.2 Main theorem.

**Theorem 5.1** *For a bisequence* $\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}} = \binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}$ *in* $(\{+,-\} \times \boldsymbol{R})^n$ $(n \geq 1)$*,*

$$M^{\binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}} = \sum_{(F,\phi_F)\in\mathcal{F}^{\binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}}} M^{\varepsilon_n,\boldsymbol{x}}_{F,\phi_F} \tag{5.18}$$

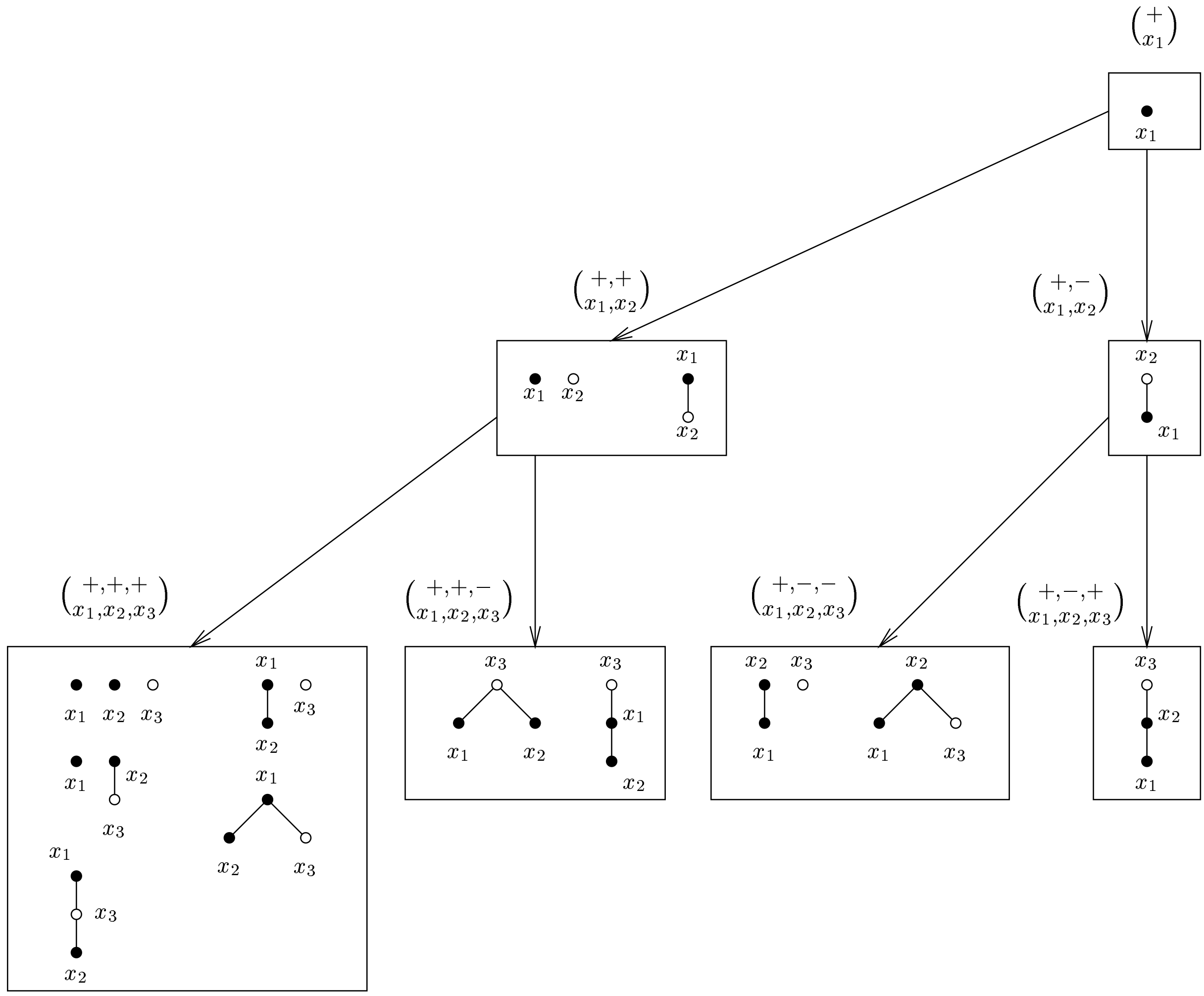



Figure 4: First labeled forests.

**Sketch of the proof :** It can be proved by induction that the sets $\mathcal{F}$ are the right ones. For $n \geq 2$, thanks to the definition of the family $M$ (eq. 2.9) :

$$M^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}}} = M^{\binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}} = M^{\binom{\varepsilon_1,\dots,\varepsilon_{n-1}}{x_1,\dots,x_{n-1}}} \sigma_{\varepsilon_n}(x_1 + \cdots + x_n) \tag{5.19}$$

This identity can then be combined with the following result :

For $\varepsilon = \pm$ and $(y_1, \dots, y_m, z)$ some real variables $(m \geq 1)$ :

$$
\begin{aligned}
&\sigma_\varepsilon(y_1)\sigma_\varepsilon(y_2)\dots\sigma_\varepsilon(y_m)\sigma_\varepsilon(y_1+\cdots+y_m+z) = \\
&\qquad \sigma_\varepsilon(y_1)\sigma_\varepsilon(y_2)\dots\sigma_\varepsilon(y_m)\sigma_\varepsilon(z) \;\; + \\
&\qquad \sum_{i=1}^{r} \begin{pmatrix} \sigma_\varepsilon(y_1)\dots\sigma_\varepsilon(y_{i-1})\sigma_\varepsilon(y_i+\cdots+y_m+z) \\ \times\; \sigma_{\bar\varepsilon}(y_{i+1}+\cdots+y_m+z)\sigma_\varepsilon(y_{i+1})\dots\sigma_\varepsilon(y_m) \end{pmatrix}
\end{aligned}
\tag{5.20}
$$

Using this property and eq. (5.19) : For $n \geq 1$, let $F$ be a forest of size $n$ and $\phi$ a given labeling by $\boldsymbol{x}^n = (x_1,\dots,x_n)$, then, for $\varepsilon = \pm$, $\eta = \pm$ and $\boldsymbol{x}^{n+1} = (x_1,\dots,x_{n+1})$, we get (see also eq. (4.16)) :

$$
M_{F,\phi}^{\varepsilon,\boldsymbol{x}^n}\,\sigma_\eta(x_1+\cdots+x_{n+1}) = \sum_{(G,\gamma)\in S_{\varepsilon,\eta}^{\boldsymbol{x}^{n+1}}(F,\phi)} M_{G,\gamma}^{\eta,\boldsymbol{x}^{n+1}} \tag{5.21}
$$

It becomes obvious that the sets $\mathcal{F}$ were build on purpose. We leave to the reader the complete proof for equations (5.20) and (5.21), but the following illustrations should be convincing.

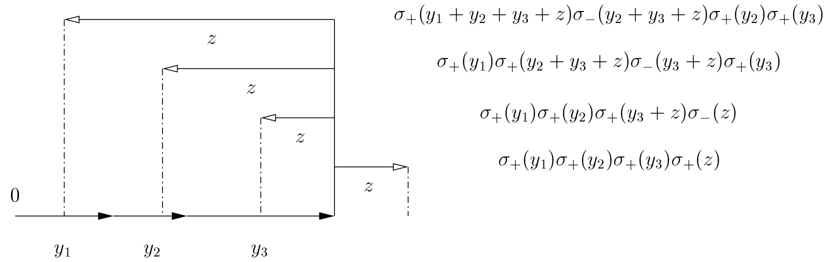



Figure 5: Eq. (5.20) for $\sigma_+(y_1)\sigma_+(y_2)\sigma_+(y_3)\sigma_+(y_1+y_2+y_3+z)$ .

Equation (5.21) is illustrated by the figure 6. We start from the coefficient :

$$
M_{F,\phi}^{+,x_1,x_2,x_3} = \sigma_-(x_1)\sigma_+(x_3+x_1)\sigma_+(x_2)
$$

The arrow (1) is the multiplication by $\sigma_-(x_1+x_2+x_3+x_4)$ and the underlying tree construction is obvious. The second arrow (2) is the multiplication by $\sigma_+(x_1+x_2+x_3+x_4)$ and, thanks to equation (5.20) ($y_1 = x_3+x_1, y_2 = x_2, z = x_4$) :

$$
\begin{aligned}
&\sigma_+(x_3+x_1)\sigma_+(x_2)\sigma_+(x_3+x_1+x_2+x_4) = \\
&\qquad\qquad \sigma_+(x_3+x_1)\sigma_+(x_2)\sigma_+(x_4) \\
&\qquad\qquad +\sigma_+(x_3+x_1)\sigma_+(x_2+x_4)\sigma_-(x_4) \\
&\qquad\qquad +\sigma_+(x_3+x_1+x_2+x_4)\sigma_-(x_2+x_4)\sigma_+(x_2)
\end{aligned}
$$

and the tree construction is now obvious.

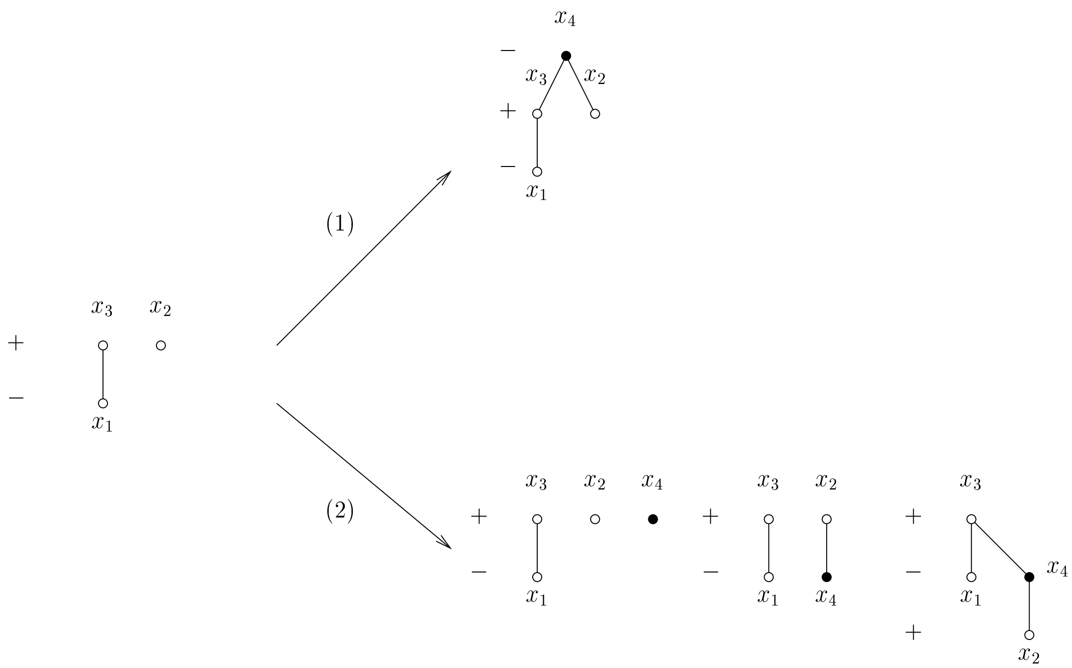



Figure 6: Illustration for Eq. (5.21)

In the following section we give some combinatorial properties of the family of forests.

We will then examine the consequences of theorem 5.1 for generalized averages and weighted functions (see section 2).

# 6 Some combinatorial properties of sets of forests $\mathcal{F}$.

Let us consider, as in section 5.1, a set $\mathcal{F}^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}}}$, with $\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}} = \binom{\varepsilon_1,\dots,\varepsilon_n}{x_1,\dots,x_n}$ in $(\{+,-\}\times\mathbf{R})^n$ $(n \geq 1)$. This is a set of labeled forests, and, once the labeling is omitted, it remains a set of forests of size $n$. This set depends only on the consecutive stacks of identical signs. We note these sets $\mathcal{F}_{n_1,\dots,n_s}$ if $\boldsymbol{\varepsilon} = (\pm)^{n_1}(\mp)^{n_2}\dots(\varepsilon_n)^{n_s}$. Let us give some combinatorial results on these sets of forests.

## 6.1 Enumeration.

For a sequence $n_1,\dots,n_s$ $(s \geq 1,\, n_i \geq 1)$ :

$$Card\,(\mathcal{F}_{n_1,\dots,n_s}) = ca_{n_1}\dots ca_{n_s} \tag{6.22}$$

Moreover, a forest in $\mathcal{F}_{n_1,\dots,n_s}$ has at most $n_s$ trees and if, for $1 \leq k \leq n_s$, $\mathcal{F}^k_{n_1,\dots,n_s}$ is the subset of forests of $\mathcal{F}_{n_1,\dots,n_s}$ having $k$ trees, then :

$$Card\left(\mathcal{F}^k_{n_1,\dots,n_s}\right) = ca_{n_1}\dots ca_{n_{s-1}}\,\frac{k}{2n_s-k}\begin{pmatrix} 2n_s-k \\ n_s \end{pmatrix} \tag{6.23}$$

This also means that, for $n \geq 1$ :

$$\sum_{k=1}^{n} \frac{k}{2n-k} \left( \begin{array}{c} 2n-k \\ n \end{array} \right) = ca_n \tag{6.24}$$

The reader shall convince himself on figure 4.

## 6.2 Construction.

All these sets were defined in section 5.1. Nonetheless, we give an equivalent construction of the sets $\mathcal{F}_{n_1,\ldots,n_s}$ by induction on $s$ ($n_1 \geq 1, \ldots, n_s \geq 1$).

- If $s = 1$ then $\mathcal{F}_{n_1}$ is the complete set of forests of size $n_1$ (see section 3.2) :

$$\mathcal{F}_{n_1} = Fa_{n_1}$$

- If $s \geq 2$ then consider the sets $\mathcal{F}_{n_1,\ldots,n_{s-1}}$ and $Fa_{n_s}$. The elements of $\mathcal{F}_{n_1,\ldots,n_s}$ are the forests obtained by considering now an element of $\mathcal{F}_{n_1,\ldots,n_{s-1}}$ as supplementary (and to its left) subtrees of the leftest root of an element of $Fa_{n_s}$.

Note that this construction yields a part of the proof for the enumeration results.

# 7 Application to generalized averages.

Let us first remind, see section 2, that, once an integrable function $f$ is given, the family $M$ plays a central role in the definition of :

- **Generalized average** (see eq. (2.11)) :

$$\mathbf{m}^{\boldsymbol{\varepsilon}} = \int_{\boldsymbol{x} \in \mathbf{R}^{l(\boldsymbol{\varepsilon})}} f(x_1) \ldots f(x_n) M^{\left(\begin{smallmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{x} \end{smallmatrix}\right)} d\boldsymbol{x}$$

- **Weighted functions** (see eq. (2.12)) :

$$\forall y \in \mathbf{R} \ , \quad f^{\boldsymbol{\varepsilon}}(y) = \int_{\boldsymbol{x} \in \mathbf{R}^{l(\boldsymbol{\varepsilon})}} f(x_1) \ldots f(x_n) M^{\left(\begin{smallmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{x} \end{smallmatrix}\right)} \delta(y - \|\boldsymbol{x}\|) d\boldsymbol{x}$$

Assuming that, as in equation (2.10), $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n) \in \{+, -\}^n$, $l(\boldsymbol{\varepsilon}) = n$ and, in addition, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n$ and $\|\boldsymbol{x}\| = x_1 + \cdots + x_n$.

## 7.1 More Tree coefficients and Tree functions.

**Definition 7.4** *For a forest $F$ of size $n$, a labeling $\phi$ of $F$ by $\boldsymbol{x} = (x_1, \dots, x_n)$, and a given sign $\varepsilon = \pm$, let us define :*

$$\boldsymbol{m}_{\phi}^{F,\varepsilon} = \int_{\boldsymbol{x}\in\boldsymbol{R}^n} f(x_1)\dots f(x_n) M_{F,\phi}^{\varepsilon,\boldsymbol{x}} d\boldsymbol{x} \tag{7.25}$$

*and*

$$\forall y \in \boldsymbol{R}\ , \quad f_{\phi}^{F,\varepsilon}(y) = \int_{\boldsymbol{x}\in\boldsymbol{R}^n} f(x_1)\dots f(x_n) M_{F,\phi}^{\varepsilon,\boldsymbol{x}} \delta(y - \|\boldsymbol{x}\|) d\boldsymbol{x} \tag{7.26}$$

Here are two noticeable facts on these coefficients :
**1.** Except for the tree coefficient $M_{F,\phi}^{\varepsilon,\boldsymbol{x}}$, the integrated function is symmetric : the previous coefficients *do not* depend on the labeling :

$$\begin{aligned} \mathbf{m}_{\phi}^{F,\varepsilon} &= \mathbf{m}^{F,\varepsilon} \\ f_{\phi}^{F,\varepsilon} &= f^{F,\varepsilon} \end{aligned}$$

**2.** If $F$ is a forest composed of $k$ trees $F = (T_1, \dots, T_k)$, then :

$$\begin{aligned} \mathbf{m}^{F,\varepsilon} &= \mathbf{m}^{T_1,\varepsilon}.\mathbf{m}^{T_2,\varepsilon}\dots\mathbf{m}^{T_k,\varepsilon} \\ f^{F,\varepsilon} &= f^{T_1,\varepsilon} * f^{T_2,\varepsilon} * \dots * f^{T_k,\varepsilon} \end{aligned}$$

These results point out that the coefficients corresponding to single trees should be the most important. Let us give now the conclusion of the previous work.

## 7.2 Tree decomposition for generalized averages and weighted functions.

Let us remind that, for a sequence $\boldsymbol{\varepsilon} \in \{+,-\}^n$ also written $\boldsymbol{\varepsilon} = (\pm)^{n_1}(\mp)^{n_2}\dots(\varepsilon_n)^{n_s}$, we defined in the previous sections a set of of forests $\mathcal{F}^{\boldsymbol{\varepsilon}}$ that corresponds either to the forests involved in $\mathcal{F}^{\binom{\boldsymbol{\varepsilon}}{\boldsymbol{x}}}$ (section 5) or to $\mathcal{F}_{n_1,\dots,n_s}$ (section 6). Combining now the previous results :

**Theorem 7.2** *For $n \geq 1$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n) \in \{+,-\}^n$, the set of forests $\mathcal{F}^{\boldsymbol{\varepsilon}}$ ($\subset Fa_n$) is such that :*

$$\boldsymbol{m}^{\boldsymbol{\varepsilon}} = \sum_{F\in\mathcal{F}^{\boldsymbol{\varepsilon}}} \boldsymbol{m}^{F,\varepsilon_n} \tag{7.27}$$

*and*

$$\forall y \in \boldsymbol{R}\ , \quad f^{\boldsymbol{\varepsilon}}(y) = \sum_{F\in\mathcal{F}^{\boldsymbol{\varepsilon}}} f^{F,\varepsilon_n}(y) \tag{7.28}$$

Here comes a first application of the tree decomposition : a proof for Andersen's formula.

# 8 Back to Andersen's formula.

## 8.1 The free associative algebra $\mathbf{Z}\langle\mathcal{T}\rangle$.

The set of ordered trees $\mathcal{T} = \{T \in Ca_n \ ; \ n \geq 1\}$ can be considered as a set of noncommuting variables, the multiplication being the concatenation. It generates the free associative $\mathbf{Z}$-algebra $\mathbf{Z}\langle\mathcal{T}\rangle$ (see [9, 14]) with a natural graduation given by the number of vertices in a tree. As any forest of $\mathcal{F} = \{F \in Fa_n \ ; \ n \geq 1\}$ is easily identified to a concatenation of trees, $\mathbf{Z}\langle\mathcal{T}\rangle = \mathbf{Z}\langle\mathcal{F}\rangle$. We can also identify some sets to the sum of their elements :

For $n \geq 1$,

$$Ca_n = \sum_{T \in Ca_n} T \tag{8.29}$$

$$Fa_n = \sum_{F \in Fa_n} F = \sum_{\substack{n_1+\cdots+n_s=n \\ s\geq 1 \ ; \ n_i \geq 1}} \sum_{\substack{T_i \in Ca_{n_i} \\ 1 \leq i \leq s}} T_1 \dots T_s = \sum_{\substack{n_1+\cdots+n_s=n \\ s\geq 1 \ ; \ n_i \geq 1}} Ca_{n_1}.Ca_{n_2} \dots Ca_{n_s} \tag{8.30}$$

## 8.2 An operator on $\mathbf{Z}\langle\mathcal{T}\rangle$.

Section 6 suggests to introduce the operator $LA$ that is called the "Left Attachment". It is a bilinear morphism from $\mathbf{Z}\langle\mathcal{T}\rangle \times \mathbf{Z}\langle\mathcal{T}\rangle$ to $\mathbf{Z}\langle\mathcal{T}\rangle$ and it can be defined on a pair of forests : Let $F_1, F_2$ be two forests, we can "add" $F_1$ to the forest $F_2$ by considering $F_1$ as supplementary (and to its left) subtrees of the leftest root of $F_2$. The resulting forest is the Left Attachment of $F_1$ to $F_2$ : $LA(F_1, F_2)$.

We have the following identities, if $F_1 = T_1^1 \dots T_t^1$ and $F_2 = T_1^2, \dots T_s^2$ then :

$$LA(F_1, T_1^2 \dots T_s^2) = LA(F_1, T_1^2)T_2^2 \dots T_s^2 \tag{8.31}$$

$$LA(T_1^1 \dots T_t^1, F_2) = LA(T_1^1, LA(T_2^1, LA(\dots, LA(T_{t-1}^1, LA(T_t^1, F_2)) \dots))) \tag{8.32}$$

Let us end this section with a combinatorial result :

**Lemma 8.1** *For $n \geq 2$,*

$$\sum_{\substack{n_1+\cdots+n_s=n \\ s\geq 2}} n_{s-1} LA(Ca_{n_1} \dots Ca_{n_{s-1}}, Ca_{n_s}) = (n-1)Ca_n \tag{8.33}$$

The proof is straightforward. Consider a tree $T$ in $Ca_n$, it can be decomposed : on one side there is the root, on the other side, there is the forest of subtrees $T' = F$ which belongs to $Fa_{n-1}$. There exists a unique composition $(m_1, \dots, m_t)$ of $n-1$ $(m_1 + \cdots + m_t = n-1)$ such that $F \in Ca_{m_1} \dots Ca_{m_t}$, thus $F = T_1 \dots T_t$. Note that we just remind here that there is a natural bijection between $Ca_n$ and $Fa_{n-1}$. For $1 \leq i \leq t$, consider $T^i$ the tree of $Ca_{n-(m_1+\cdots+m_i)}$ obtained by omitting $T_1 \dots T_i$ in $T$ :

$$T = LA(T_1 \dots T_i, T^i)$$

and

$$(n-1)T = (m_1 + \cdots + m_t)T = \sum_{i=1}^{t} m_i LA(T_1 \dots T_i, T^i)$$

But if we note $r(F)$ the tree such that $r(F)' = F$ ($r(\emptyset) = \bullet$), then :

$$\begin{aligned}
(n-1)Ca_n &= \sum_{T \in Ca_n} (n-1)T \\
&= \sum_{n_1+\cdots+n_s=n-1} \sum_{F \in Ca_{n_1}\dots Ca_{n_s}} (n-1)r(F) \\
&= \sum_{n_1+\cdots+n_s=n-1} \sum_{\substack{T_k \in Ca_{n_k} \\ 1\le k\le s}} (n_1+\cdots+n_s)r(T_1\dots T_s) \\
&= \sum_{n_1+\cdots+n_s=n-1} \sum_{\substack{T_k \in Ca_{n_k} \\ 1\le k\le s}} \sum_{i=1}^{s} n_i LA(T_1\dots T_i, r(T_{i+1}\dots T_s)) \\
&= \sum_{i=1}^{n-1} \sum_{n_1+\cdots+n_i+m=n} n_i LA(Ca_{n_1}\dots Ca_{n_i}, Ca_m)
\end{aligned}$$

This is the attempted result.

Thanks to section 6, Andersen's formula is a corollary of this lemma.

## 8.3 A proof of Andersen's formula.

We remind that, to any sequence $(n_1, \dots, n_s)$ of positive integers, a set of forest is associated by :

$$\mathcal{F}_{n_1,\dots,n_s} = \begin{cases} Fa_{n_1} & \text{if} \quad s=1 \\ \\ LA(\mathcal{F}_{n_1,\dots,n_{s-1}}, Fa_{n_s}) & \text{if} \quad s \ge 2 \end{cases}$$

We also define a similar family by

$$\widetilde{\mathcal{F}}_{n_1,\dots,n_s} = \begin{cases} Fa_{n_1} & \text{if} \quad s=1 \\ \\ LA(\sigma(\widetilde{\mathcal{F}}_{n_1,\dots,n_{s-1}}), Fa_{n_s}) & \text{if} \quad s \ge 2 \end{cases}$$

where $\sigma(F) = \sigma(T_1 \dots T_k) = T_k \dots T_1$. The first family is useful as, for a given sequence $(n_1, \dots, n_s)$ :

$$\mathbf{m}_f^{(\pm)^{n_1}(\mp)^{n_2}\dots(\varepsilon)^{n_s}} = \sum_{F \in \mathcal{F}_{n_1,\dots,n_s}} \mathbf{m}_f^{F,\varepsilon}$$

Because of the definition of the tree coefficients, one also gets :

$$\mathbf{m}_f^{(\pm)^{n_1}(\mp)^{n_2}\dots(\varepsilon)^{n_s}} = \sum_{F \in \widetilde{\mathcal{F}}_{n_1,\dots,n_s}} \mathbf{m}_f^{F,\varepsilon}$$

and these sets inherit some combinatorial properties from the previous lemma :

For $n \geq 1$

$$\sum_{n_1+\cdots+n_s=n} \widetilde{\mathcal{F}}_{n_1,\ldots,n_s} = \sum_{l_1+\cdots+l_t=n} l_1 Ca_{l_1}\ldots Ca_{l_t} \tag{8.34}$$

Let us give the proof by induction. This result is obviously true for $n=1$. Let $n \geq 2$ and suppose that the equation holds for $1 \leq k < n$.

$$\begin{aligned}
\sum_{n_1+\cdots+n_s=n} \widetilde{\mathcal{F}}_{n_1,\ldots,n_s} &= \widetilde{\mathcal{F}}_n + \sum_{\substack{n_1+\cdots+n_s=n \\ s\geq 2}} \widetilde{\mathcal{F}}_{n_1,\ldots,n_s} \\
&= Fa_n + \sum_{\substack{n_1+\cdots+n_s=n \\ s\geq 2}} LA(\sigma(\widetilde{\mathcal{F}}_{n_1,\ldots,n_{s-1}}), Fa_{n_s}) \\
&= Fa_n + \sum_{k=1}^{n-1} LA\left(\sigma\left(\sum_{\substack{m_1+\cdots+m_{s'}=n-k \\ s'\geq 1}} \widetilde{\mathcal{F}}_{n_1,\ldots,n_{s'}}\right), Fa_k\right) \\
&= Fa_n + \sum_{k=1}^{n-1} \sum_{\substack{l_1+\cdots+l_t=n-k \\ k_1+\cdots+k_u=k}} LA(l_t Ca_{l_1}\ldots Ca_{l_t}, Ca_{k_1}\ldots Ca_{k_u}) \\
&= Fa_n + \sum_{k=1}^{n-1} \sum_{\substack{l_1+\cdots+l_t=n-k \\ k_1+\cdots+k_u=k}} LA(l_t Ca_{l_1}\ldots Ca_{l_t}, Ca_{k_1}) Ca_{k_2}\ldots Ca_{k_u} \\
&= Fa_n + \sum_{k'=0}^{n-1} \sum_{k'_1+\cdots+k'_u=k'} (n-k'-1) Ca_{n-k'} Ca_{k'_1}\ldots Ca_{k'_u} \\
&= \sum_{l_1+\cdots+l_t=n} Ca_{l_1}\ldots Ca_{l_t} + \sum_{l_1+\cdots+l_t=n} (l_1-1) Ca_{l_1}\ldots Ca_{l_t} \\
&= \sum_{l_1+\cdots+l_t=n} l_1 Ca_{l_1}\ldots Ca_{l_t}
\end{aligned}$$

and this ends the proof.

Let us remember that

$$\forall n \geq 1 \quad ; \quad \tau_n = \mathbf{m}_f^{\overbrace{-,\ldots,-}^{(n-1)\text{ times}},+} = \sum_{T\in Ca_n} \mathbf{m}_f^{T,+}$$

As $F \mapsto \mathbf{m}_f^{F,+}$ defines a morphism from $\mathbf{Z}\langle\mathcal{T}\rangle$ to $\mathbf{R}$, the equation (8.34) reads :

$$\begin{aligned}
\sum_{l_1+\cdots+l_t=n} l_1 \tau_{l_1}\ldots\tau_{l_t} &= \sum_{n_1+\cdots+n_s=n} \sum_{F\in\widetilde{\mathcal{F}}_{n_1,\ldots,n_s}} \mathbf{m}_f^{F,+} \\
&= \sum_{\varepsilon_i=\pm\ ;\ 1\leq i\leq n-1} \mathbf{m}_f^{\varepsilon_1\ldots\varepsilon_{n-1}+}
\end{aligned}$$

If $f$ is a probability density, then the last term is $P(S_n > 0)$ for the associated random walk and from this relation, we get, if $\tau(s) = \sum_{n=1}^{+\infty} \tau_n s^n$,

$$\frac{\tau'(s)}{1-\tau(s)} = \sum_{n=1}^{+\infty} s^{n-1} P(S_n > 0)$$

We have proved Andersen's formula :

$$\log \frac{1}{1-\tau(s)} = \sum_{n=1}^{+\infty} \frac{s^n}{n} P(S_n > 0) \tag{8.35}$$

Note that, if we don't care about probability, we have a similar equation for generalized averages, as soon as $f$ is an integrable function.

# 9 Trees and excursion for the simple random walk.

## 9.1 Introduction

"Trees" and "random walk" are words that are already associated in the probability literature. We can cite the bijection between the Galton-Watson process and the excursions of the simple random walk, see [1, 8, 13]. This result is based on a combinatorial bijection between Dyck paths and ordered trees.

We shall here focus on this combinatorial approach and prove that our tree-decomposition gives back this results. The main point is that our decomposition is a generalization of this result : it establishes a bijection between labeled trees and "sets of paths". Note that this relation seems to differ from the one that can be found in [1].

Let us first remind the bijection between Dyck paths and ordered trees.

## 9.2 Paths and trees.

**Definition 9.5** *A Dyck path of length $2n$ ($n \in \boldsymbol{N}^*$) is a positive path with jumps $+1$ or $-1$ starting at $0$ and finishing at $0$ for the $2n^{th}$ jump. We note $D_n$ the set of Dyck paths of length $2n$.*

**Proposition 9.1** *For $n \geq 1$, the set of ordered trees of size $n$ and the set of Dyck paths of length $2n$ have the same cardinal.*

This is a very classical result and from a given bijection between these two kind of sets, one can deduce the usual isomorphism between excursions of the simple random walk and the geometric Galton-Watson process.

For $n \geq 1$, the natural bijection $\Psi_n$ between $D_n$ and $Ca_n$ can be defined by induction. For $n = 1$, the unique Dyck path of size 2 is associated to the unique tree of size 1. For $n \geq 2$, the bijection is defined as follows :

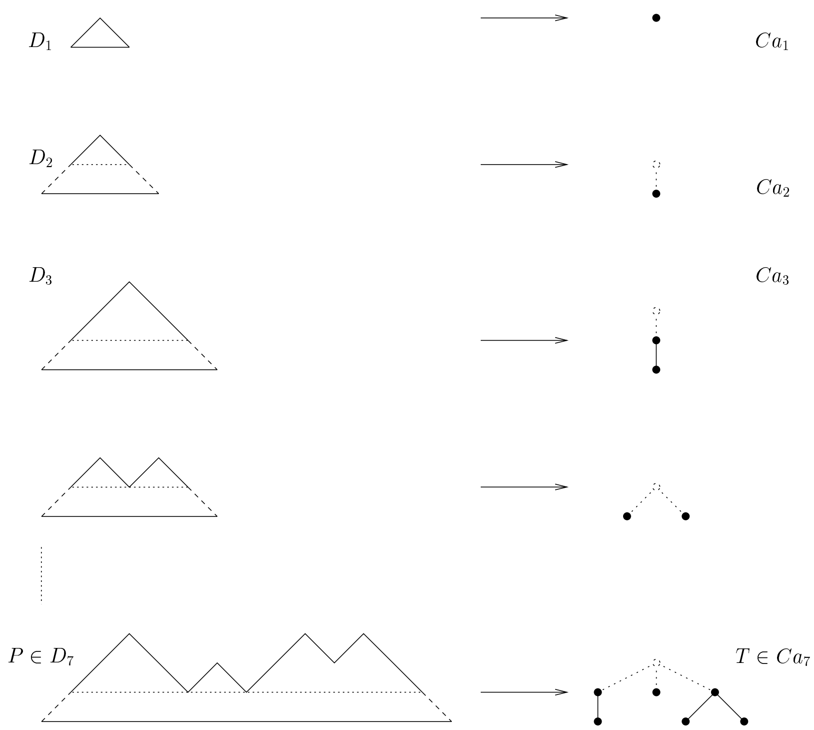



Figure 7: Inductive definition of the bijection.

Consider a Dyck path $P$ of size $2n$, omit the first and the last jump and consider that this new path starts at the origin. This new path $P'$ is of length $2n-2$, is finishing at 0 and is non negative : it is a concatenation of $s \geq 1$ Dyck paths $P_1, \ldots, P_s$ of size $2n_1, \ldots, 2n_s$ $(n_1 + \cdots + n_s = n-1)$. Using the induction, each of the paths $P_i$ corresponds to a unique tree $T_i$ of $Ca_{n_i}$. The tree of $Ca_n$ that is associated to $P$ is the tree obtained by adding a common root to the ordered forest $(T_1, \ldots, T_s)$. The construction of the inverse map is obvious. The figure 7 illustrates this construction.

We prove in the following section that this bijection can be seen as a particular case of our tree-decomposition. To do so, we first define a generalization of Dyck paths.

## 9.3 Generalized Dyck paths.

A path of length $n$ is simply a sequence of coordinates $((0,0),(1,y_1),\ldots,(n,y_n))$ or, equivalently, a sequence of "jumps" $(x_1, x_2, \ldots, x_n)$ (with $y_i = \check{x}_i = x_1 + \cdots + x_i$, for $1 \leq i \leq n$). For example, Dyck paths are paths with only $+1$ or $-1$ jumps.

**Definition 9.6** *A generalized Dyck path (GDP) of length $n$ ($n \geq 1$) is a path, with real jumps, such that :*

$$y_n = \check{x}_n = x_1 + \cdots + x_n \leq 0$$

*and, if $n \geq 2$,*

$$\forall\ 1 \leq i \leq n-1\ ,\ y_i = \check{x}_i = x_1 + \cdots + x_i > 0$$

Dyck paths of length $2n$ are GDP of length $2n$, and, if $GD_n$ stands for GDP of length $n$, then considering paths as sequences of jumps, the set $GD_n$ can easily be expressed in term of the family $M$ (see section 2.2) :

$$GD_n = \left\{ (x_1, \ldots, x_n) \in \mathbf{R}^n\ ;\ M^{\binom{\overbrace{+, \ \cdots\ ,+}^{n-1}\ ,\ -}{x_1, \ldots, x_{n-1}, x_n}} = 1 \right\} \tag{9.36}$$

with

$$M^{\binom{\varepsilon_1, \ldots, \varepsilon_n}{x_1, \ldots, x_n}} = \sigma_{\varepsilon_1}(x_1)\sigma_{\varepsilon_2}(x_1 + x_2) \ldots \sigma_{\varepsilon_n}(x_1 + \cdots + x_n)$$

and

$$\sigma_+ = \mathbf{1}_{\mathbf{R}^{+*}} \quad and \quad \sigma_- = \mathbf{1}_{\mathbf{R}^-}$$

Note that

$$\forall n \geq 1\ ;\ GD_n \cap \{+1, -1\}^n = \begin{cases} D_{n/2} & \text{if } n \text{ is even} \\ \\ \emptyset & \text{if } n \text{ is odd} \end{cases}$$

We described in the first section the well-known bijection between Dyck paths and trees. Let us now investigate how the definition of generalized Dyck paths interact with the tree decomposition of the family $M$.

## 9.4 GDP and tree-decomposition : back to the labeling.

For $n \geq 1$ the sets are $GD_n$ are strongly linked to some elements of the family $M$ (see above), especially, the coefficients $M^{\binom{+,\ \cdots\ ,+\ ,\ -}{x_1, \ldots, x_{n-1}, x_n}}$. Because of theorem 5.1,

$$M^{\binom{+,\ \cdots\ ,+\ ,\ -}{x_1, \ldots, x_{n-1}, x_n}} = \sum_{(F, \phi_F) \in \mathcal{F}^{\binom{+,\ \cdots\ ,+\ ,\ -}{x_1, \ldots, x_{n-1}, x_n}}} M_{F, \phi_F}^{-, \boldsymbol{x}} \tag{9.37}$$

The forests in the set $\mathcal{F}^{\binom{+,\ \cdots\ ,+\ ,\ -}{x_1, \ldots, x_{n-1}, x_n}}$ are exactly the trees of $Ca_n$ (see section 6). It means that the tree decomposition induces a partition of the set $GD_n$ into $ca_{n-1}$ subsets :

$$GD_n = \bigcup_{T \in Ca_n} GD^T \quad \text{where} \quad GD^T \overset{def}{=} \{(x_1, \ldots, x_n) \in \mathbf{R}^n\ ;\ M_{T, \phi_T}^{-, (x_1, \ldots, x_n)} = 1\} \tag{9.38}$$

The labeling were not explicit in section 5 but, for a better understanding of the above partition, we need to give a precise statement on the labeling of these trees.

**Theorem 9.3** *For $n \geq 1$, the forests of $\mathcal{F}^{\binom{+, \cdots, +, -}{x_1, \ldots, x_{n-1}, x_n}}$ are the trees of $Ca_n$. Moreover, if $T$ be a tree of $Ca_n$, the labeling of this tree, as an element of $\mathcal{F}^{\binom{+, \cdots, +, -}{x_1, \ldots, x_{n-1}, x_n}}$, is given by one of the following procedures :*

**Procedure 1 :**

- *The root of $T$ is labeled by $x_n$.*
- *Once this root is omitted, we get a forest of $Fa_{n-1}$ having $s$ trees $T_1, \ldots, T_s$ of size $\alpha_1, \ldots, \alpha_s$ $(s \geq 1)$. Note that each $\alpha_i$ is positive and $\alpha_1 + \cdots + \alpha_s = n - 1$. We split the sequence $(x_1, \ldots, x_{n-1})$ into $s$ successive subsequences $X^1 \ldots X^s$. The labels of the tree $T_i$ are in the sequence $X^i$, and as the roots of these trees are at odd distance of the original root, we label them by the first element of their respective sequence $X^i$.*
- *We apply then the same procedure to each tree* **but**, *if the new root to label is at even distance of the original root, its label is not the first of the associated subsequence, but the last.*

**Procedure 2 :**

- *The root of $T$ is labeled by $x_n$.*
- *Once this root is omitted, we get a forest of $Fa_{n-1}$ having $s$ trees $T_1, \ldots, T_s$ of size $\alpha_1, \ldots, \alpha_s$ $(s \geq 1)$. We split the sequence $(x_1, \ldots, x_{n-1})$ into $s$ successive subsequences $X^1 \ldots X^s$. The labels of the tree $T_i$ are those of $X^i = (x_{\breve{\alpha}_{i-1}+1}, \ldots, x_{\breve{\alpha}_i})$ and the labeling of $T_i$ is the same as the one encountered for this tree in the set $\mathcal{F}^{\binom{+ \quad , \ldots, \quad + \quad\quad -}{x_{\breve{\alpha}_i} \; , \ldots, \; x_{\breve{\alpha}_{i-1}} \; x_{\breve{\alpha}_{i-1}+1}}}$ (the sequence is reversed).*

The two procedures are obviously equivalent.This result is illustrated in figure 8, for a tree of $\mathcal{F}^{\binom{+ \;, \ldots, \; + \;\; -}{x_1 \;, \ldots, \; x_7 \; x_8}}$

**Proof :** From section 5, we deduce the following results :

*Fact 1 :* Because of the construction in section 5, the set of labeled forests $\mathcal{F}^{\binom{+, \cdots, +, -}{x_1, \ldots, x_{n-1}, x_n}}$ is easily deduced from the set $\mathcal{F}^{\binom{+, \cdots, +}{x_1, \ldots, x_{n-1}}}$ by adding a root labeled by $x_n$ to the forest belonging to this set. It means that the sets to study are

$$\mathcal{F}^{\binom{+, \ldots, +}{x_1, \ldots, x_n}} \stackrel{def}{=} F^{x_1, \ldots, x_n}$$

Note also that, omitting the labeling, the forests in the set $\mathcal{F}^{\binom{+, \ldots, +}{x_1, \ldots, x_n}}$ are exactly those of $Fa_n$ and the forests in $\mathcal{F}^{\binom{+, \cdots, +, -}{x_1, \ldots, x_{n-1}, x_n}}$ are exactly the *trees* of $Ca_n$

*Fact 2 :* If

$$A^{x_1, \ldots, x_n} \stackrel{def}{=} \{F \in F^{x_1, \ldots, x_n} \text{ such that } F \text{ is a tree}\}$$

then $A^{x_1} = F^{x_1}$ has one element, the unique tree of size 1 labeled by $x_1$, and for $n \geq 2$, we have the following properties :

- A labeled tree $T$ of size $n$ belongs to $A^{x_1, \ldots, x_n}$ iff : "the rightest son of its root is labeled by $x_n$, and there exists a unique $k$ such that $1 \leq k \leq n - 1$, the forest *under*

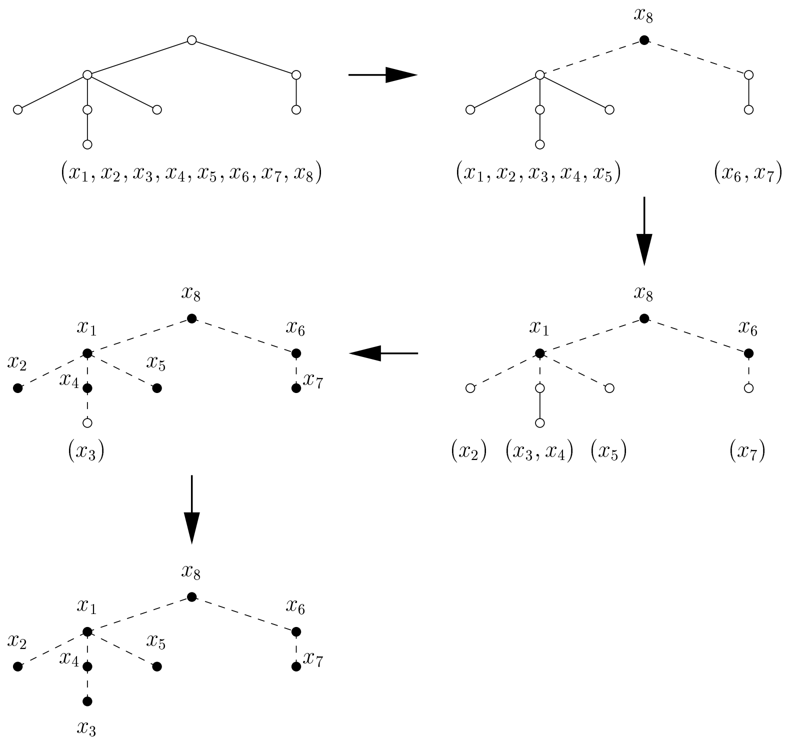



Figure 8: Labeling a tree.

$x_n$ belongs to $F^{x_{k+1},\dots,x_{n-1}}$ and the tree $T$ amputated of the subtree starting at $x_n$ belongs to $A^{x_1,\dots,x_k}$".

- A labeled forest $F$ belongs to $F^{x_1,\dots,x_n}$ iff there exist a decomposition of $(x_1,\dots,x_n)$ in *successive* subsequences $X^1,\dots,X^t$ $(t \geq 1)$ such that $F$ is a sequence of trees $T^1,\dots,T^t$ with
$$\forall 1 \leq i \leq t \ , \ T^i \in A^{X^i}$$

Theorem 9.3 is clearly a consequence of these facts.

## 9.5 From $GD_{2n}$ to $D_n$.

The set $D_n$ is equipotent to the set $Ca_n$, and is a finite subset of $GD_{2n}$. The tree-decomposition of the family $M$ induces a partition of the set $GD_{2n}$ into $ca_{2n-1}$ subsets that are indexed by the trees of $Ca_{2n}$. We will prove that this partition induces a bijection between $D_n$ and a subset of $Ca_{2n}$ that can be naturally identified to $Ca_n$ and that, after identification, the bijection is the one described in section 9.2.

For $n \geq 1$, we remind that

$$GD_{2n} \cap \{-1,+1\}^{2n} = D_n \quad \xrightarrow{\Psi_n} \quad Ca_n$$

and

$$GD_{2n} = \bigcup_{T \in Ca_{2n}} GD^T$$

thus, it seems natural to study, for $T \in Ca_{2n}$, the set

$$\widetilde{GD}^T = GD^T \cap \{-1,+1\}^{2n} = \left\{ (x_1,\dots,x_{2n}) \in \{-1,+1\}^{2n} \ ; \ M_{T,\phi_T}^{-,(x_1,\dots,x_{2n})} = 1 \right\} \qquad (9.39)$$

Let us first define the levels of a tree : the root is at level 1, its sons are at level 2 and so on ... .

**Theorem 9.4** *Let $T \in Ca_{2n}$ $(n \geq 1)$, the set $\widetilde{GD}^T$ is non-empty iff each vertex of $T$ of odd level has exactly one son. In this case the set $\widetilde{GD}^T$ contains one element (a Dyck path of length $2n$).*

**Proof :** Let us consider a tree $T$ of $Ca_{2n}$ and suppose that $(x_1,\dots,x_{2n}) \in \widetilde{GD}^T$. Each $x_i$ is attached to a vertex and we can consider the partial sum $s_i$, with respect to the order induced by the tree : $s_i$ is $x_i$ plus its primogeniture. We can also associate a sign $\varepsilon_i$ to each $x_i$, as in the tree-coefficient : $\varepsilon_i$ is $+$ if $x_i$ is the label of a vertex at even level and is $-$ otherwise.

In these conditions (see section 4),

$$M_{T,\phi_T}^{-,(x_1,\dots,x_{2n})} = \prod_{i=1}^{2n} \sigma_{\varepsilon_i}(s_i) = 1$$

thus,

$$\forall 1 \leq i \leq 2n \ ; \quad \sigma_{\varepsilon_i}(s_i) = 1$$

Suppose that $x_i$ is the label of a leaf at an odd level, then $s_i = x_i = -1$. When we go "up", its father is labeled by $x_{i^-}$. $s_{i^-}$ is the sum of $x_{i^-}$, $s_i = -1$, and other non positive partial sums associated to the other sons of $x_{i^-}$. The sum $s_{i^-}$ is obtained by adding at most $x_{i^-} = +1$ and $s_{i^-} - x_{i^-} \leq -1$ thus $s_{i^-}$ *cannot* be positive. It means that the leafs of $T$ must be at even levels.

Let us now consider a vertex of $T$ that is not a leaf, suppose that it is labeled by $x_i$ and has $k$ sons ($k \geq 1$) labeled by $x_{i_1}, \ldots, x_{i_k}$. If this is vertex is at an even level $s_i = x_i + s_{i_1} + \cdots + s_{i_k} > 0$ whereas, for $1 \leq j \leq k$, $s_{i_j} \leq 0$ : it implies that $x_i = +1$ and for $1 \leq j \leq k$, $s_{i_j} = 0$. If this vertex is at an odd level $s_i = x_i + s_{i_1} + \cdots + s_{i_k} \leq 0$ whereas, for $1 \leq j \leq k$, $s_{i_j} > 0$ : it implies that $x_i = -1$ and $k = 1$ and $s_{i_1} = 1$.

This proves that if $\widetilde{GD}^T$ is non-empty then it satisfies the condition of the theorem. Conversely, starting with the leafs and going "up", the above considerations shows that for, such a tree, the label $x_i$ at odd (resp. even) level are equal to $+1$ (resp. $-1$) with partial sums $s_i = +1$ (resp. $0$) : there is a unique Dyck path in the set $\widetilde{GD}^T$.

Let us define for $n \geq 1$ :

$$C_n \stackrel{def}{=} \{T \in Ca_{2n} \ ; \ \widetilde{GD}^T \neq \emptyset\} \tag{9.40}$$

It is now obvious that this subset of $Ca_{2n}$ is equipotent to $D_n$ and to $Ca_n$. For $n \geq 1$, there was a first "natural" bijection $\Psi_n$ between $D_n$ and $Ca_n$ ; the above theorem yields a second bijection $\Phi_n$ between $C_n$ and $D_n$ that associate to a tree $T$ of $C_n$ the unique Dyck path in $\widetilde{GD}^T$. We can consider the bijection $\Psi_n \circ \Phi_n$ that happens to be very simple :

**Theorem 9.5** *For $n \geq 1$, the bijection $\Theta_n = \Psi_n \circ \Phi_n$ between $C_n$ and $Ca_n$ can be described as follows :*

*Consider a tree $T \in C_n$, it is a tree of $Ca_{2n}$ such that each vertex of $T$ of odd level has exactly one son. If these vertices of odd level are glued with their respective unique son, such that the in between edges disappear, then a tree $T'$ of $Ca_n$ is obtained and $T' = \Theta_n(T)$.*

**Proof :** First consider that, starting from $C_n$, $\Theta_n$ is defined by the gluing procedure, then it remains to prove by induction that $\Theta_n = \Psi_n \circ \Phi_n$. The reader who wants to skip the proof can have a look to figure 9.

The result is obvious for $n = 1$. For a tree $T$ in $C_n$ ($n \geq 2$), consider its labeling, as an element of $\mathcal{F}^{\binom{+,\ \ldots\ ,+\ ,\ -}{x_1,\ldots,x_{2n-1},x_{2n}}}$ (it is described by theorem 9.3). As an element of $C_n$, its root is labeled by $x_{2n}$ and its unique son is labeled by $x_1$. Now, it becomes obvious, that omitting this edge ($x_{2n} - x_1$) in the tree $T$ corresponds, for the path $\Phi_n(T)$ to omit the first (resp. last) step $x_1 = +1$ (resp. $x_{2n} = -1$) and because of the inductive definition of $\Psi_n$, it corresponds to omit the root in $\Psi_n \circ \Phi_n(T)$, as in $\Theta_n(T)$. Once this omission is done, we get $k$ subtrees of $T$ :

$$(T_1, \ldots, T_k) \in Ca_{n_1} \times \cdots \times Ca_{n_k} \quad (n_1 + \cdots + n_k = 2n - 2)$$

As $T$ is in $C_n$, the integers $n_1, \ldots, n_k$ are even and :

$$\forall 1 \leq i \leq k \ ; \ T_i \in C_{m_i} \ (m_i = n_i/2)$$

Looking at the induced labeling of these trees, the paths $\Phi_{m_1}(T_1), \ldots, \Phi_{m_k}(T_k)$ are exactly those obtained once $x_1$ and $x_{2n}$ are omitted and the trees $\Psi_{m_1} \circ \Phi_{m_1}(T_1), \ldots, \Psi_{m_k} \circ \Phi_{m_k}(T_k)$ are those obtained from $\Psi_n \circ \Phi_n(T)$ once its root is omitted. By induction,

$$\forall 1 \leq i \leq k \ ; \ \Psi_{m_i} \circ \Phi_{m_i}(T_i) = \Theta_{m_i}(T_i)$$

and then $\Psi_n \circ \Phi_n(T) = \Theta_n(T)$. It ends the proof of this theorem.

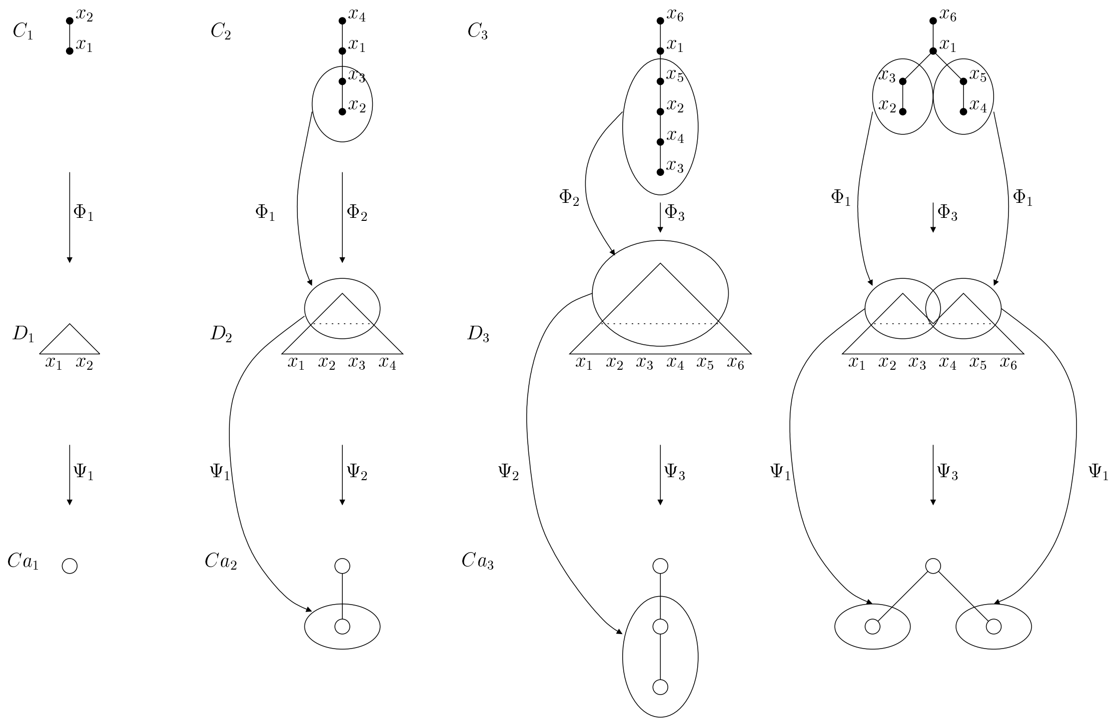



Figure 9: Illustration of the proof.

# 10 Conclusion.

The proof of the Andersen's formula , as well as the results for the simple random walk, are the first applications of our tree decomposition. There is now doubt that this tool will provide other known, and hopefully unknown, formulas or properties for random walks.

Another application will be given in a coming paper. A complete answer will be given to the following question : when do two densities in $\mathrm{Vect}_{\mathbf{R}}\{x \mapsto e^{-\lambda|x|} \ ; \ \lambda > 0\}$ induce the same average ?

# A A note on the terminology.

We end this paper by a short explanation for the terminology of "averages induced by diffusion". This denomination was given in the mathematical theory that originally motivated the author's interest in random walk : the *resummation theory.* We briefly describe here the links between this theory and the random walks.

A specific set of probabilities associated to a random walk on $\mathbf{R}$ has been called an average because this terminology was first adopted in resummation theory. The well-behaved uniformizing averages (WBA) were introduced by J. Ecalle in the framework real resummation, which aims at assigning a real sum to a real divergent series of "natural origin", for example a formal solution of a differential equation. The need for "uniformizing" some analytic ramified functions appears naturally in this kind of problem. For example, let $\varphi$ be an analytic function with singularities over $\mathbf{N}^*$, analytically continuable on the universal covering of $\mathbf{C}/\mathbf{N}^*$. For a given non-negative integer $n$ we can label the $2^n$ determinations of $\varphi$ over the interval $]n, n+1[$ that are obtained by analytic continuation of $\varphi$ along the $2^n$ paths dodging the singularities $\{1, \dots, n\}$ to the left or to the right. If the sign $+$ (resp. $-$) is assigned when dodging to the right (resp. to the left), these $2^n$ determinations of $\varphi$ over $]n, n+1[$ are labeled by the addresses $(\varepsilon_1, \dots, \varepsilon_n)$ ($\varepsilon_i = \pm$ and $1 \leq i \leq n$). Such functions appear naturally in the real-resummation theory and it is almost as natural to try to associate a uniform function to $\varphi$, that is to "uniformize" it. The simplest way is to do, over the interval $]n, n+1[$, an "average" of the $2^n$ determinations of $\varphi$, pondering them by $2^n$ "weights" $\mathbf{m}^{\varepsilon_1,\dots,\varepsilon_n}$, of sum (for a given $n$) equal to 1.

Thus, a *uniformizing average* $\mathbf{m}$ is simply a collection of weights :

$$\mathbf{m} = \{\mathbf{m}^{\varepsilon_1,\dots,\varepsilon_n}\,;\, n \geq 0\,;\, \varepsilon_i = \pm\,;\, 1 \leq i \leq n\} \tag{A.41}$$

Some additionnal conditions, both analytic and algebraic, are imposed to such averages, so that they become a very powerful tool in real resummation. In these conditions, the average $\mathbf{m}$ is called a *well-behaved average* (WBA). For details see [3, 4, 10, 11]. The study of particular averages, which appeared to be WBA, proved the existence of such objects, and, for example, J. Ecalle found a great class of WBA : the *averages induced by diffusion.*

# References


[1] J. Bennies and G. Kersting. A random walk approach to Galton-Watson trees. *J. Theor. Prob.*, 13(3):777–803, 2000.

[2] L. Comtet. *Analyse Combinatoire*, volume 1. P.U.F., 1970.

[3] J. Ecalle. Well-behaved convolution averages and their applications to real resummation. To appear.

[4] J. Ecalle and F. Menous. Well-behaved convolution averages and the non-accumulation theorem for limit-cycles. In *The Stokes Phenomenon and Hilbert's 16th Problem*, pages 71–102. World Scientific, 1996.

[5] C. EVEN. Etude d'une fonction remarquable associée aux moyennes de convolution. *Annales de l'Institut Fourier*, 49(2):687–705, 1999.

[6] W. FELLER. *An Introduction to Probability Theory and Its Applications.*, volume 2. Wiley, 1966.

[7] H.M. FINUCAN. Some elementary aspects of the Catalan numbers. In *Combinat. Math. IV*, volume 560 of *LNM*, pages 41–45. Springer, 1975.

[8] J. F. LE GALL and Y. LE JAN. Branching processes in Lévy-processes : the exploration process. *Ann. Prob.*, 26:213–252, 1998.

[9] C. MALVENUTO and C. REUTENAUER. Duality between Quasi-Symmetric Functions and the Solomon Descent Algebra. *Journal of Algebra*, 177:967–982, 1995.

[10] F. MENOUS. *Les bonnes moyennes uniformisantes et leurs applications à la resommation réelle.* Thèse de doctorat, 1996.

[11] F. MENOUS. Les bonnes moyennes uniformisantes et une application à la resommation réelle. *Annales de la Faculté des Sciences de Toulouse*, VIII(4):579–628, 1999.

[12] F. MENOUS. Well-behaved averages, random walk on $\mathbf{R}$ with "linear exponential law" and rooted-oriented trees. *Prépublications d'Orsay*, (16), 2000.

[13] J. NEVEU. Arbres et processus de Galton-Watson. *Ann. Inst. Henri Poincaré*, 22:199–207, 1986.

[14] C. REUTENAUER. *Free Lie Algebras.* Oxford Univ. Press, 1993.

[15] R. P. STANLEY. *Enumerative Combinatorics*, volume 1. Cambridge University Press, 1997.