Random Boolean trees.

Inria's Junior Seminar
Inria Rocquencourt.

Cécile Mailler

Laboratoire de Mathématiques de Versailles
Université de Versailles-Saint-Quentin

September 18, 2012

Headlines

What's a Boolean function?

Definition

$$f: \begin{array}{ccc} \{0,1\}^k & \to & \{0,1\} \\ (x_1,\dots,x_k) & \mapsto & f(x_1,\dots,x_k) \end{array}$$

We denote by $\mathcal{F}_k$ the set of all Boolean functions on k variables.

$$\begin{array}{ccc} 1 & = & \textit{True} \\ 0 & = & \textit{False} \end{array}$$

What's a Boolean function?

Definition

$$\begin{array}{rccc} f: & \{0,1\}^k & \to & \{0,1\} \\ & (x_1,\dots,x_k) & \mapsto & f(x_1,\dots,x_k) \end{array}$$

We denote by $\mathcal{F}_k$ the set of all Boolean functions on k variables.

$$\begin{array}{ccc} 1 & = & \textit{True} \\ 0 & = & \textit{False} \end{array}$$

The cardinality of $\mathcal{F}_k$ is

$$2^{2^k}$$

What's a Boolean function?

Definition

$$f : \begin{array}{ccc} \{0,1\}^k & \to & \{0,1\} \\ (x_1,\dots,x_k) & \mapsto & f(x_1,\dots,x_k) \end{array}$$

We denote by $\mathcal{F}_k$ the set of all Boolean functions on k variables.

1	=	*True*
0	=	*False*

The cardinality of $\mathcal{F}_k$ is

$$2^{2^k}$$

A Boolean function can be represented by:

- a truth table
- Boolean expressions

$x_1 \wedge x_2$	0	1
0	0	0
1	0	1

Some Boolean functions

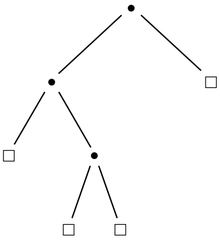
- **Constants:** $((x_1, \dots, x_k) \mapsto 1)$ and $((x_1, \dots, x_k) \mapsto 0)$ denoted by *True* and *False*.
- **Projections:** $((x_1, \dots, x_k) \mapsto x_i)$
- **Negations:** $((x_1, \dots, x_k) \mapsto \overline{x_i})$

$$\boxed{\overline{x_i} = 1 - x_i}$$

- **And:** $((x_1, \dots, x_k) \mapsto x_i \wedge x_j)$
- **Or:** $((x_1, \dots, x_k) \mapsto x_i \vee x_j)$
- **Xor:** $((x_1, \dots, x_k) \mapsto x_i \text{ XOR } x_j)$

$$\boxed{x_i \text{ XOR } x_j = (x_i \wedge \overline{x_j}) \vee (\overline{x_i} \wedge x_j)}$$

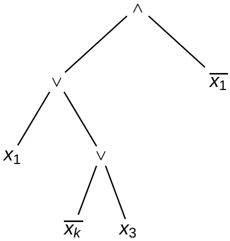
What's a Boolean AND/OR tree?



Definition

The **size** of a tree is the number of its internal nodes.

What's a Boolean AND/OR tree?



$\rightsquigarrow$

Boolean expression

$$\left[x_1 \vee \left(\overline{x_k} \vee x_3\right)\right] \wedge \overline{x_1}$$

$\rightsquigarrow$

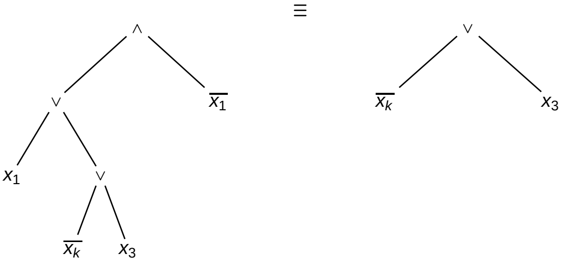
Boolean function

$$(x_1, \ldots, x_k) \mapsto \overline{x_k} \vee x_3$$

Definition

The **size** of a tree is the number of its internal nodes.

Different trees can compute the same function!



$$\Phi : \text{ Boolean trees } \to \text{ Boolean functions}$$

The function Φ is surjective but **not injective**.

Objective of the talk:
define a probability distribution $\mathbb{P}$ on trees and study the distribution $\mathbb{p}$ induced by Φ on Boolean functions.

$$\Phi : \text{ Boolean trees } \to \text{ Boolean functions}$$

The function Φ is surjective but **not injective**.

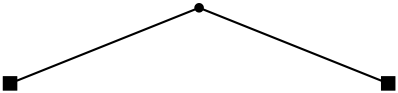
Objective of the talk:
define a probability distribution $\mathbb{P}$ on trees and study the distribution $\mathbb{p}$ induced by Φ on Boolean functions.

Litterature:

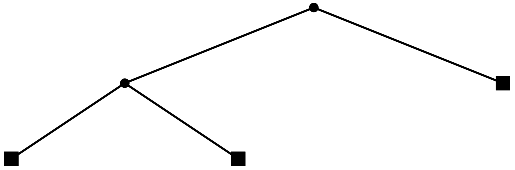
- uniform distribution on $\mathcal{F}_k$ [Shannon 1963]
- uniform distribution on the set of trees of size n, on k variables [Lefmann & Savický 1997] [Chauvin et al. 2006]
- Galton-Watson distribution on the set of trees on k variables [Chauvin et al. 2006]

Random Binary Search Tree

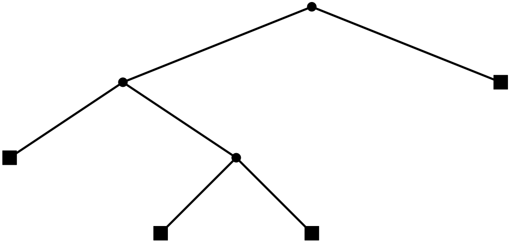
Random Binary Search Tree



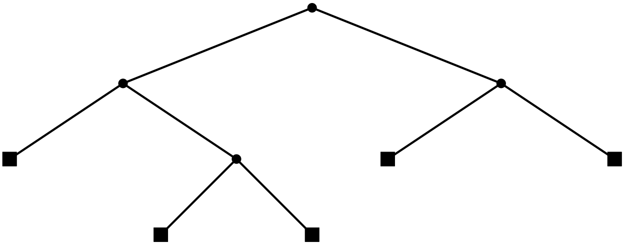
Random Binary Search Tree



Random Binary Search Tree

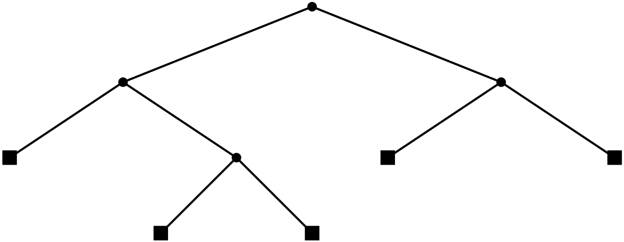


Random Binary Search Tree



We stop after the n^{th} step.

Random Binary Search Tree



We stop after the n^{th} step.

- For each internal node, we flip a coin to choose between $\wedge$ and $\vee$.
- For each external node, we choose uniformly at random a label in $\{x_1, \overline{x_1}, \ldots, x_k, \overline{x_k}\}$.

Main result

We denote by $\mathcal{T}_n$ the growing tree of size n, and by f_n the random Boolean function it computes.

Theorem

When the size n of the growing tree tends to infinity,

$$\mathbb{P}(f_n = True) = \mathbb{P}(f_n = False) \to \frac{1}{2}.$$

Actually, we have just **flipped a coin** to choose between the constant function *True* and the constant function *False*!

Proof of the result

Theorem

When the size n of the growing tree tends to infinity,

$$\mathbb{P}(f_n = True) = \mathbb{P}(f_n = False) \to \frac{1}{2}.$$

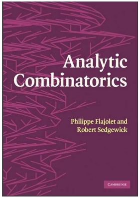
Two approaches:

- Via analytic combinatorics
  - because it is the **classical** approach
  - but it is quite **technical**
- Via continuous time embedding
  - because it is more **efficient** and gives the speed of convergence
  - but it is an approach **specific** to the growing tree model

Analytic combinatorics

- We want to study the sequences $(\mathbb{p}_n(f) := \mathbb{p}(f_n = f))_{n\geq 0}$ for all Boolean function $f \in \mathcal{F}_k$.
- We introduce the **generating functions**

$$\phi_f(z) = \sum_{g\geq 0} \mathbb{p}_n(f) z^n.$$



- If you get information about the **dominant singularity** of $\phi_f(z)$ and the behaviour of $\phi_f(z)$ around this singularity, then you get some information about the **asymptotic behaviour** of $\mathbb{p}_n(f)$.

The induction property of the model

Proposition

The two subtrees of a growing tree are growing trees, and if ℓ_n is the size of the left subtree of $\mathcal{T}_n$, for all $q \in \{0, n-1\}$,

$$\mathbb{P}(\ell_n = q) = \frac{1}{n}.$$

The induction property of the model

Proposition

The two subtrees of a growing tree are growing trees, and if ℓ_n is the size of the left subtree of $\mathcal{T}_n$, for all $q \in \{0, n-1\}$,

$$\mathbb{P}(\ell_n = q) = \frac{1}{n}.$$

We have $\mathbb{p}_{n+1}(f) = \sum_{q=0}^{n} \frac{1}{n+1} \mathbb{p}(f_{n+1} = f | \ell_{n+1} = q)$. Thus,

$$\mathbb{p}_{n+1}(f) = \sum_{q=0}^{n} \frac{1}{n+1} \left(\frac{1}{2} \sum_{g \wedge h = f} \mathbb{p}_q(g) \mathbb{p}_{n-q}(h) + \frac{1}{2} \sum_{g \vee h = f} \mathbb{p}_q(g) \mathbb{p}_{n-q}(h) \right),$$

The induction property of the model

Proposition

The two subtrees of a growing tree are growing trees, and if ℓ_n is the size of the left subtree of $\mathcal{T}_n$, for all $q \in \{0, n-1\}$,

$$\mathbb{P}(\ell_n = q) = \frac{1}{n}.$$

We have $\mathbb{p}_{n+1}(f) = \sum_{q=0}^{n} \frac{1}{n+1} \mathbb{p}(f_{n+1} = f | \ell_{n+1} = q)$. Thus,

$$(n+1)\mathbb{p}_{n+1}(f) = \sum_{q=0}^{n} \left(\frac{1}{2} \sum_{g \wedge h = f} \mathbb{p}_q(g)\mathbb{p}_{n-q}(h) + \frac{1}{2} \sum_{g \vee h = f} \mathbb{p}_q(g)\mathbb{p}_{n-q}(h) \right),$$

The induction property of the model

Proposition

The two subtrees of a growing tree are growing trees, and if ℓ_n is the size of the left subtree of $\mathcal{T}_n$, for all $q \in \{0, n-1\}$,

$$\mathbb{P}(\ell_n = q) = \frac{1}{n}.$$

We have $\mathbb{p}_{n+1}(f) = \sum_{q=0}^{n} \frac{1}{n+1} \mathbb{p}(f_{n+1} = f | \ell_{n+1} = q)$. Thus,

$$(n+1)\mathbb{p}_{n+1}(f) = \sum_{q=0}^{n} \left(\frac{1}{2} \sum_{g \wedge h = f} \mathbb{p}_q(g)\mathbb{p}_{n-q}(h) + \frac{1}{2} \sum_{g \vee h = f} \mathbb{p}_q(g)\mathbb{p}_{n-q}(h) \right),$$

$$\sum_{n \geq 0} (n+1)\mathbb{p}_{n+1}(f) z^n = \frac{1}{2} \sum_{g \wedge h = f} \left(\sum_{n \geq 0} \sum_{q=0}^{n} \mathbb{p}_q(g)\mathbb{p}_{n-q}(h) z^n \right) + \frac{1}{2} \sum_{g \vee h = f} \left(\sum_{n \geq 0} \sum_{q=0}^{n} \mathbb{p}_q(g)\mathbb{p}_{n-q}(h) z^n \right).$$

Abracadabra!

$$\phi'_f(z) = \frac{1}{2}\sum_{g \wedge h = f} \phi_g(z)\phi_h(z) + \frac{1}{2}\sum_{g \vee h = f} \phi_g(z)\phi_h(z).$$

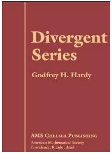
After some long and technical arguments, we get:

$$\phi_{True}(z) = \phi_{False}(z) \quad \sim \frac{1}{2(1-z)} \qquad \text{when } z \to 1^-,$$

$$\phi_f(z) \quad = o\left(\frac{1}{1-z}\right) \qquad \text{when } z \to 1^-, \text{ for all } f \notin \{True, False\}.$$

We apply a ~~transfer lemma~~ Tauberian theorem and get immediately:

$$\mathbb{p}_n(True) = \mathbb{p}_n(False) \to \frac{1}{2} \text{ when } n \to \infty.$$



Embedding in continuous time
The exponential law

- radioactive desintegration
- arrivals of phone calls at a call center
- arrivals of buses at a bus stop

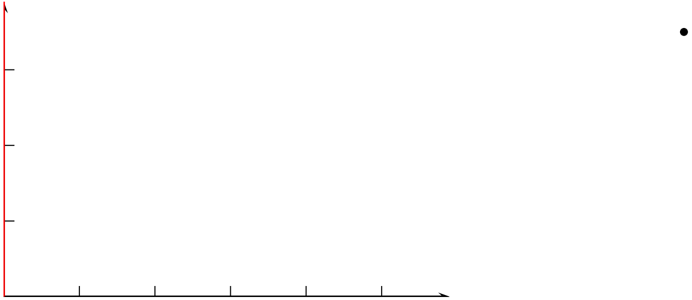
The random variable $\tau \in \mathbb{R}^+$ follows the exponential law of parameter λ if

$$\mathbb{P}(\tau > t) = e^{-\lambda t}.$$

Properties

- no memory: for all $t < s \in \mathbb{R}^+$, $\mathbb{P}(\tau > s \,|\, \tau > t) = \mathbb{P}(\tau > s - t)$.
- minimum: if $\tau_1, \ldots, \tau_m \sim \mathcal{E}(1)$ are independent, $\mathbb{P}(\min\{\tau_1, \ldots, \tau_m\} = \tau_i) = \frac{1}{m}$ for all $i \in \{1, \ldots, m\}$.

Embedding in continuous time
What's a Yule tree?



Labelled Yule tree

We label the process, uniformly at random.

Labelled Yule tree

We label the process, uniformly at random.

→ We thus have $\mathcal{Y}_t$, a continuous time process of binary AND/OR trees. We denote by g_t the random Boolean function computed by $\mathcal{Y}_t$.

→ $g_t = f_{n(t)}$ almost surely, $n(t)$ being the size of $\mathcal{Y}_t$

In a Yule tree, the right and left subtrees of each node are **independent**.

Strategy of proof

- We have to show that, asymptotically, only constant functions have a nonzero probability.
- Let us consider $\mathbb{p}_t^{10} := P(g_t(a) = 1 \text{ and } g_t(b) = 0)$ for fixed a and $b \in \{0,1\}^k$ two assignments of the k variables.
- Let us show that $\mathbb{p}_t^{10}$ tends to zero when $t \longrightarrow +\infty$, independently from the choice of a and b.

After **two lines of easy calculations**, you get, if $\pi(t) := \mathbb{p}_t^{10}$:

$$\pi' + \pi^2 = 0 \rightarrow \pi(t) = \frac{1}{t+cst}$$

After **two lines of easy calculations**, you get, if $\pi(t) := \mathbb{p}_t^{10}$:

$$\pi' + \pi^2 = 0 \rightarrow \pi(t) = \frac{1}{t+cst}$$

If $n(t)$ is the number of leaves of $\mathcal{Y}_t$, then $n(t) \sim e^t$ a.s., thus:

$$t \sim \ln n(t).$$

Theorem:

When n tends to inifnity, $\mathbb{p}_n(True) = \mathbb{p}_n(False) \rightarrow \frac{1}{2}$ and the speed of convergence is of order $\mathcal{O}\left(\frac{1}{\ln n}\right)$.

Extensions of the result

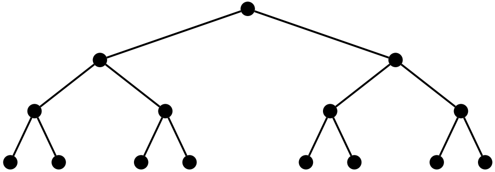
- bias the law on the literals $\mathbb{P}(x_i) = \mathbb{P}(\bar{x}_i) > 0$: no change!

- bias the choice of the connectives: $\mathbb{P}(\wedge) = q$ and $\mathbb{P}(\vee) = 1 - q$.
  - if $q > \frac{1}{2}$ then $\mathbb{p}_n(\textit{False}) \to 1$
  - if $q < \frac{1}{2}$ then $\mathbb{p}_n(\textit{True}) \to 1$

- bias the choice of the connective and allow only positive literals
  - if $q > \frac{1}{2}$ then $\mathbb{p}_n(x_1 \wedge \ldots \wedge x_k) \to 1$
  - if $q < \frac{1}{2}$ then $\mathbb{p}_n(x_1 \vee \ldots \vee x_k) \to 1$
  - if $q = \frac{1}{2}$ then ... (solved but complicated)

- the implication model: connective $\to$ and positive literals only: $\mathbb{p}_n(\textit{True}) \to 1$

Balanced trees
[Genitrini 2009]



The balanced tree induces a distribution on $\mathcal{F}_k$ which behaves exactly as the growing tree distribution!

> “In expectation, a large growing tree contains a large balanced tree”

Conjecture

Every random Boolean tree which “contains a large balanced tree” induces a degenerate distribution on Boolean functions.

Thank you for your attention!

And thanks to my two supervisors Danièle Gardy et Brigitte Chauvin!