A propos des grands arbres simples
et
une notion d'arbre continu

Jean-François Marckert

— ○ — ○ — ○ — ○ — ○ — ○ — ○ — ○ —

Arbres simples et arbres de Galton-Watson

- Définitions et lien entre les modèles
- Codages des arbres
- Liens entre les différents codages
- Propriétés asymptotiques des codages des arbres de GW

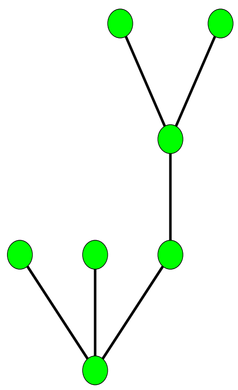
Une notion d'arbre continu

- Notion de convergence d'une suite d'arbres
- Les arbres simples de taille n, renormalisés, convergent
- À quoi ça sert?

Arbres de Galton-Watson

Modèle : processus de GW de loi de reproduction $\mathbf{p} = (p_i)_{i \geq 0}$

$\longrightarrow$ Tous les individus ont un nombre aléatoire d'enfants de loi $\mathbf{p}$ indépendamment les uns des autres.
$\longrightarrow$ Au début on a un individu (la racine).



- L'arbre de GW est l'arbre généalogique (aléatoire) de cette population.
- C'est un arbre planaire (totalement ordonné).

La probabilité qu'un arbre aléatoire $\mathbf{T}$ soit égal à un arbre donné t est

$$\mathbb{P}(\mathbf{T} = t) = \prod_{u \in t} p_{d(u)}$$

où $d(u)$= nombre d'enfants du noeud u

Arbres de Galton-Watson conditionnés par la taille

Proposition 1 *Soit ξ une variable de loi $\mathbf{p}$. Si $\mathbb{E}(\xi) \leq 1$, l'arbre de GW est p.s. fini.*

On suppose maintenant $\mathbb{E}(\xi) = 1$ et $0 < \sigma^2 = \mathrm{var}(\xi) < +\infty$.

Que peut-on dire de t sachant $|t| = n$?

Rappel :

$$\mathbb{P}_n(\mathbf{T} = t) := \mathbb{P}\left(\mathbf{T} = t \,|\, |\mathbf{T}| = n\right) = \frac{\mathbb{P}(\mathbf{T} = t)\mathbb{1}_{|t|=n}}{\mathbb{P}(|\mathbf{T}| = n)}$$

Exemples de familles d'arbres de GW sous $\mathbb{P}_n$:

- Si $\mathbf{p} = (p(1-p)^i)_{i\geq 0}$ est la loi géométrique, $\mathbb{P}_n$= loi uniforme sur les arbres plans à n noeuds.
- Si $p_2 = 1 - p_0$, $\mathbb{P}_{2n+1}$ est la loi uniforme sur les arbres binaires à $2n+1$ noeuds.
- Arbres k-aires uniformes
- Arbres non uniformes aussi

Lien avec les arbres simples

ARBRE SIMPLE :On se donne une famille de poids $(\varphi_k)_{k\geq 0}$.
Le poids de l'arbre t est:

$$\omega(t) = \prod_{u\in t} \varphi_{d(u)}.$$

Soit $a_n = \sum_{t,\ |t|=n} \omega(t)$. On a

$$A(z) = z\Phi(A(z)).$$

Si on note $\mathbb{Q}_n$ la "loi" des arbres de taille n :

$$\mathbb{Q}_n(t) = \frac{\omega(t)}{a_n}\mathbb{1}_{|t|=n}.$$

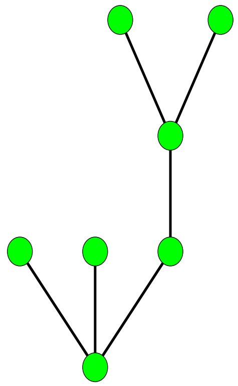
ARBRE DE GW :Pour tout $x \in]0, R_\varphi[$, $p_k = \dfrac{\varphi_k x^k}{\Phi(x)}$ est une proba.

Sous cette loi de reproduction, pour tout n,

$$\mathbb{P}_n = \mathbb{Q}_n$$

De plus pour x solution de $x\Phi'(x) = \Phi(x)$, est un arbre de Galton-Watson critique ($\mathbb{E}(\xi) = 1$).

Quelques éléments sur les arbres sous $\mathbb{P}_n$



Si $\mathbb{P}$ est critique et $0 < \sigma^2 < +\infty$,

$$\mathbb{P}(|\mathbf{T}| = n) \sim n^{-3/2}/(\sqrt{2\pi}\sigma)$$

Hauteur :

$$\frac{H_n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} \frac{2}{\sigma}H$$ Flajolet & Odlyzko (1982)

+ CV moments où la loi de H vérifie

$$\mathbb{P}(H \leq x) = \sum_{k\in\mathbb{Z}} (1 - 2k^2x^2)\exp(-k^2x^2).$$

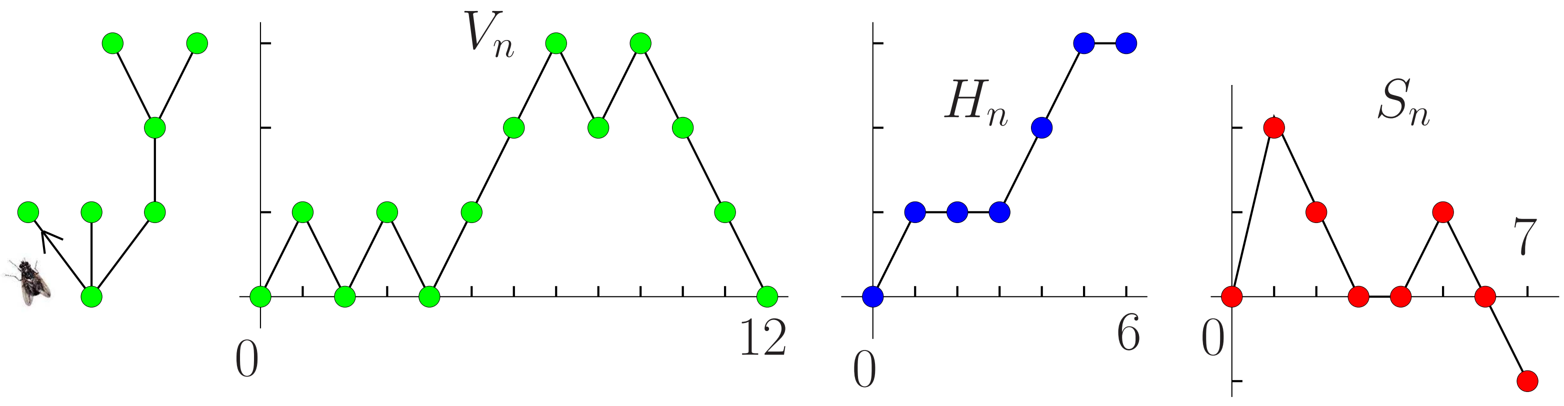
Largeur : Takàcs (1993)

$$\frac{L_n}{\sqrt{n}} \xrightarrow[n\to\infty]{loi} \sigma H$$

+ résultats sur le comportement global (Aldous, Drmota & Gittenberger, Le Gall ...)

Parcours en profondeur et codage des arbres

Tour, Processus des hauteurs et PPP des arbres ayant n nœuds



La mouche marche à la vitesse 1.

- Tour de l'arbre V_n : ordonnée de la mouche en fonction du temps.

Soit v_i le $i$ème nouveau nœud visité par la mouche

- Processus des hauteurs H_n :

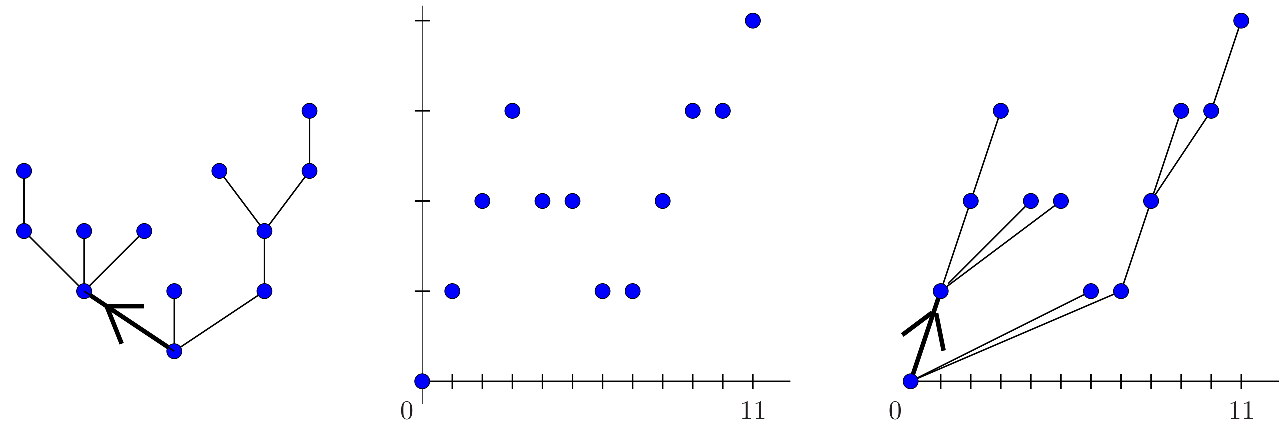
$$H_n(i) = d(racine, v_i) \quad \text{pour} \quad 0 \leq i \leq n-1.$$

- PPP : Pile du parcours en profondeur S_n :

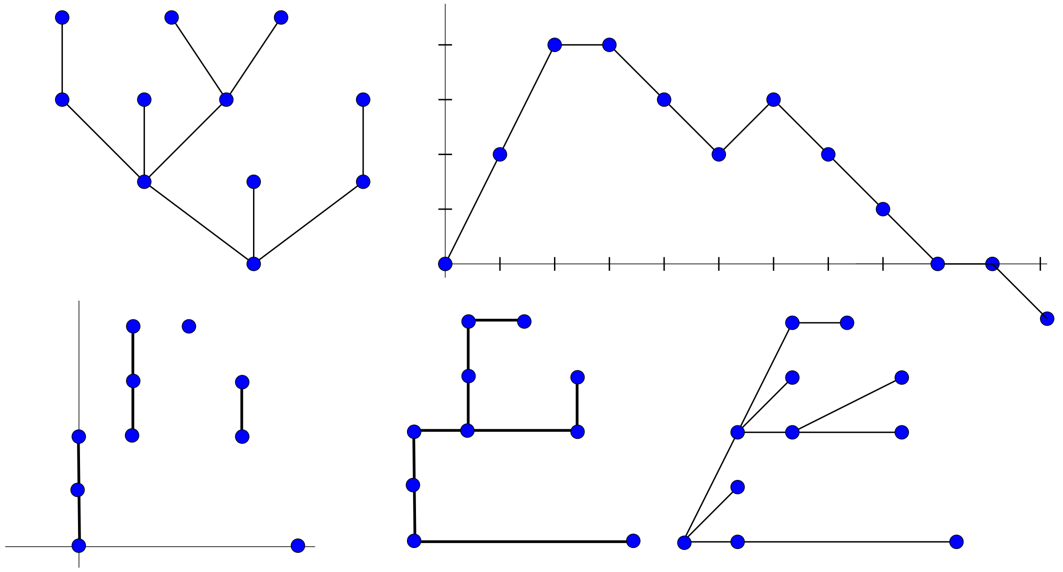
$$S_n(j) = \sum_{i=0}^{j-1} (\xi_i - 1) \text{ pour tout } 1 \leq j \leq n.$$

A propos des processus

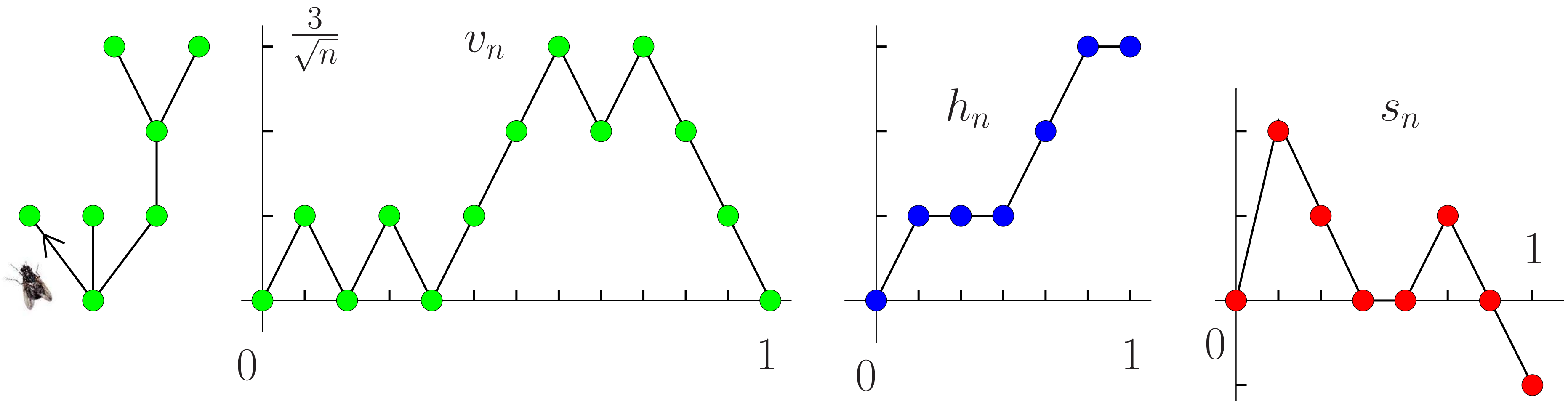
le processus des hauteurs



La PPP $S_n(j) = \sum_{i=0}^{j-1}(\xi_i - 1)$ est une marche aléatoire dont les incréments sont des variables de loi $\xi - 1$ conditionnée par $\left\{S_n(j) \geq 0 \text{ pour } j \leq n-1 \text{ et } S_n(n) = -1\right\}$



Normalisation et plongement dans $(C[0,1], \|.\|_\infty)$



$$v_n(t) = \frac{V_n(2(n-1)t)}{\sqrt{n}}, \quad h_n(t) = \frac{H_n((n-1)t)}{\sqrt{n}}, \quad s_n(t) = \frac{S_n(nt)}{\sqrt{n}}$$

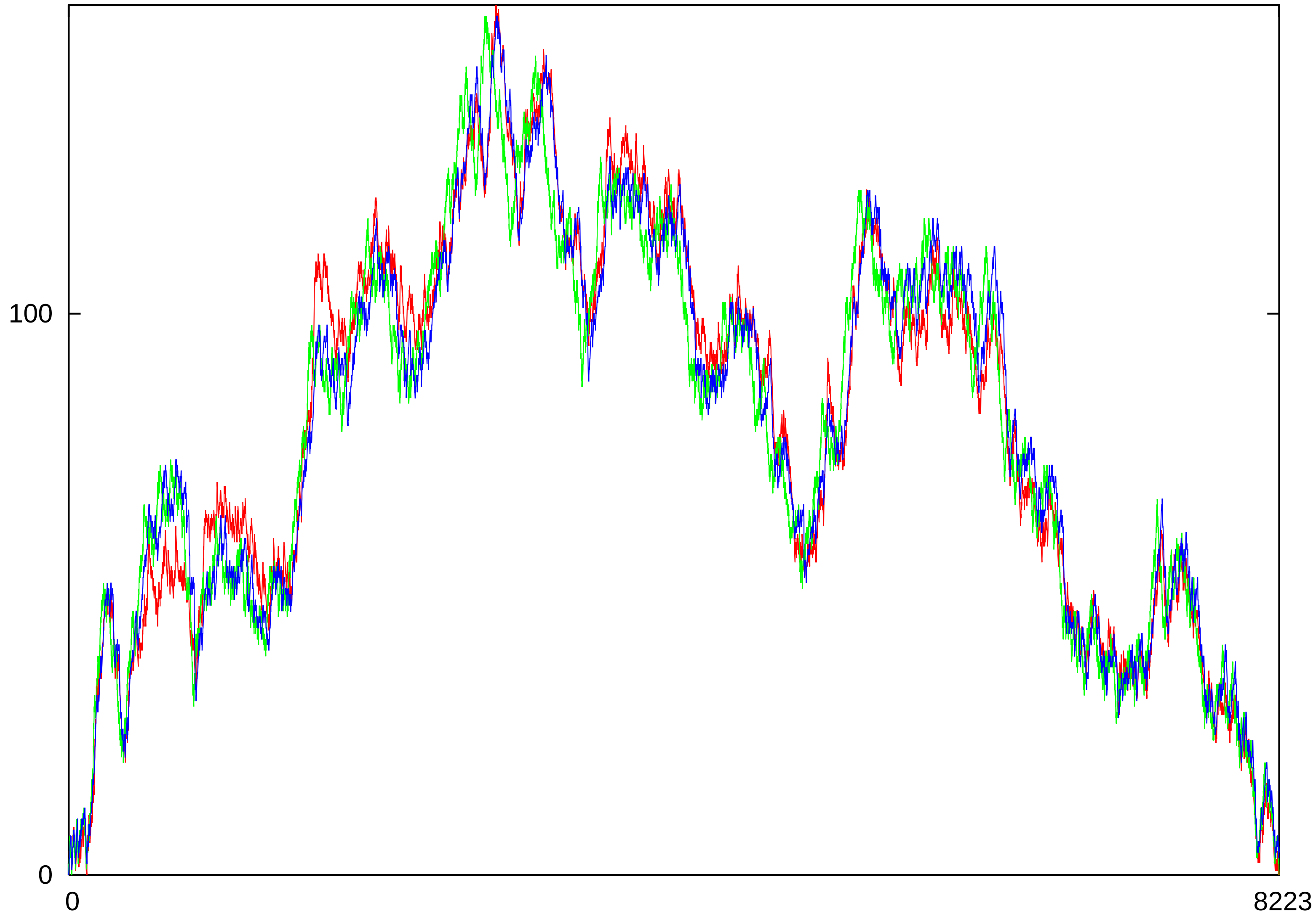
Théorème (Ces trois processus sont asympt. proportionnels. (M. & Mokkadem)).
Pour tout $\nu > 0$, il existe $c > 0$, tel que

$$\mathbb{P}\left(\|v_n - h_n\|_\infty \geq n^{\nu - 1/4}\right) \leq e^{-cn^\nu}, \qquad \mathbb{P}\left(\left\|h_n - \frac{\sigma^2}{2} s_n\right\|_\infty \geq n^{\nu-1/4}\right) \leq e^{-cn^\nu}$$

Une simulation

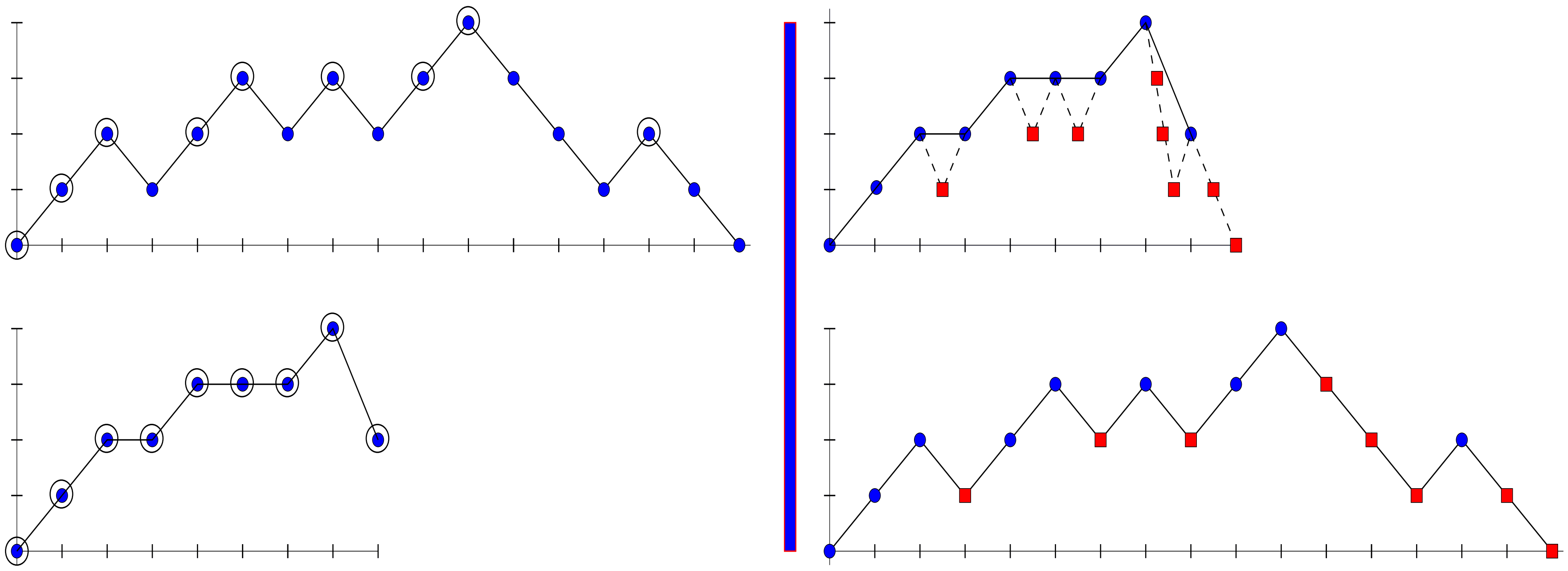
Distribution : $n = 8223,\ p_0 = 7/10, p_2 = p_3 = p_5 = 1/10.$

$\mathbb{E}(\xi) = 1$, $\sigma^2 = 14/5$, $\sigma^2/2 = 7/5.$ $V_n(2.)$, $H_n(.)$, $\dfrac{S_n(.)}{7/5}$



Explications...

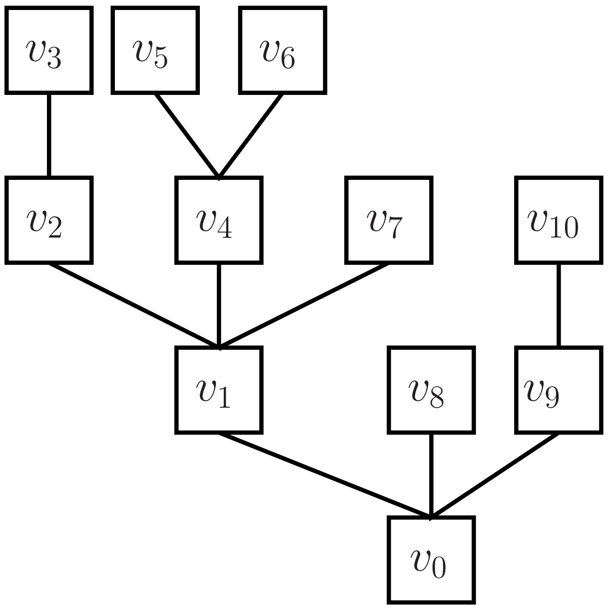
Proximité du tour et du processus des hauteurs :

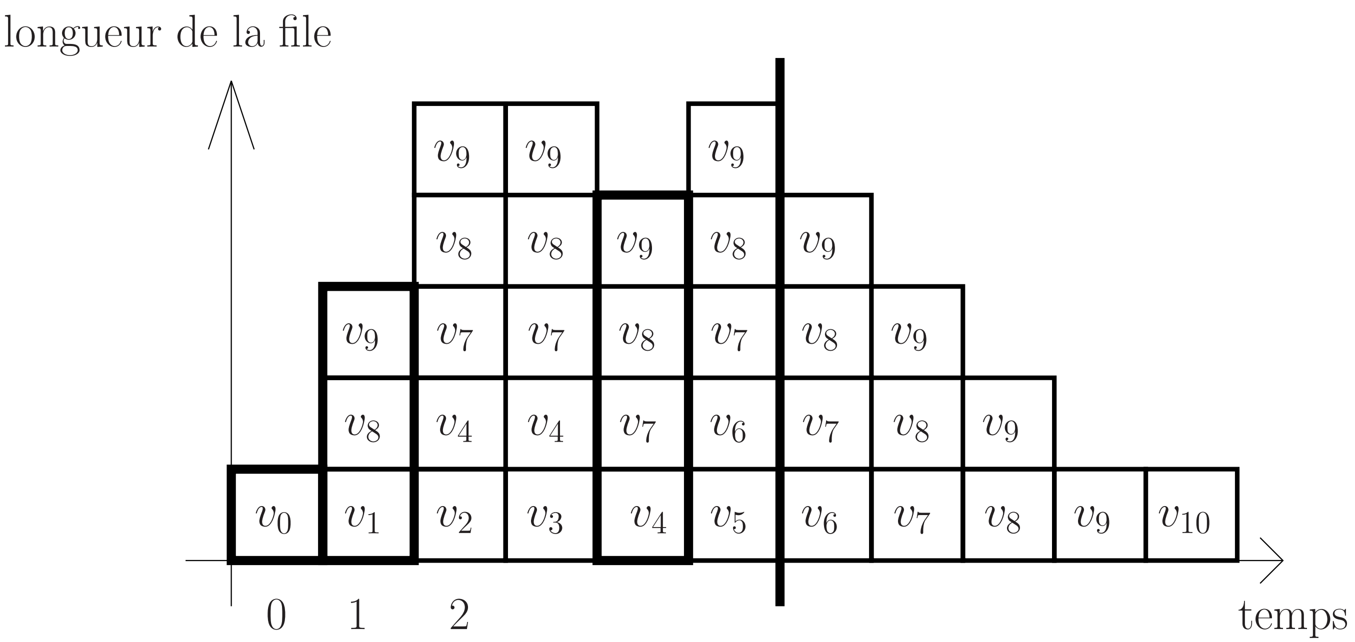


Raison: un point sur 2 (ou presque) suit une montée dans un chemin de Dyck ...

Explications...

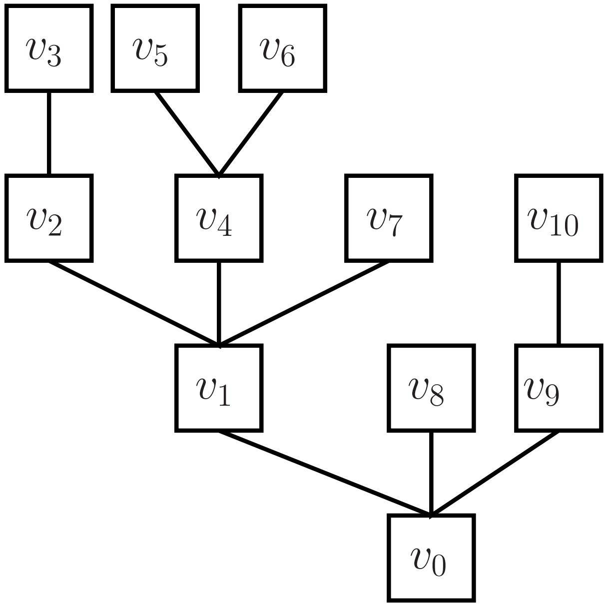
Proximité du processus des hauteurs et de la PPP

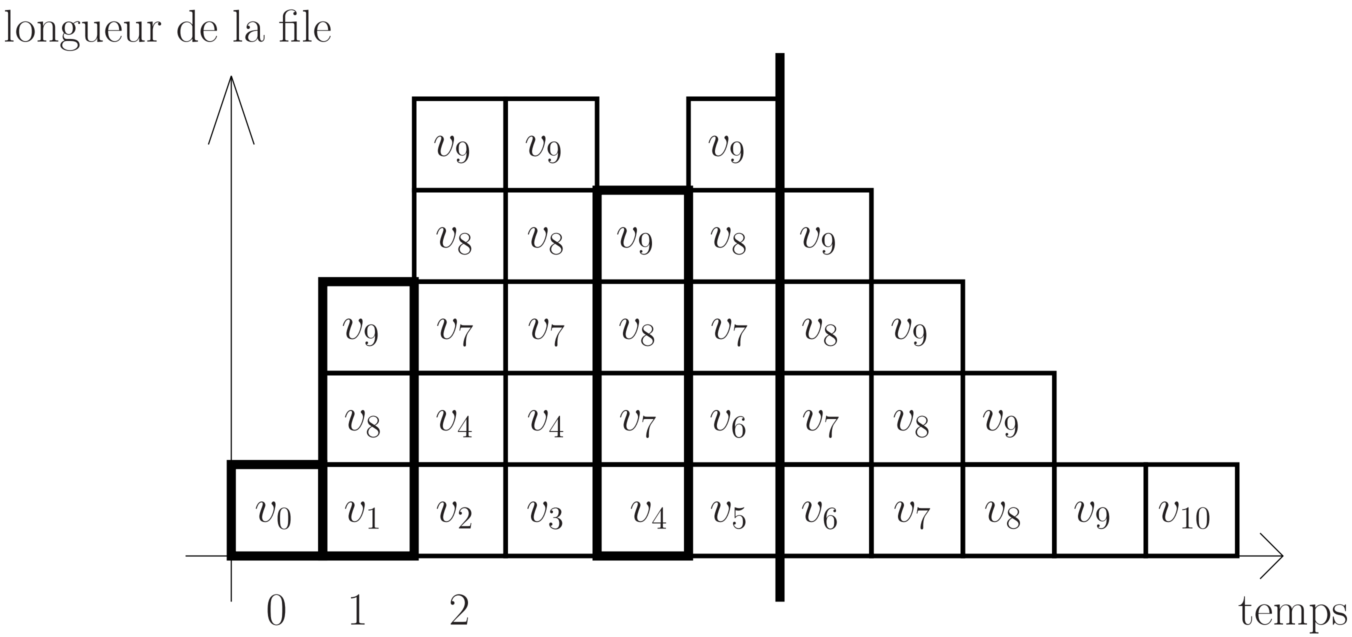




Explications...

Proximité du processus des hauteurs et de la PPP

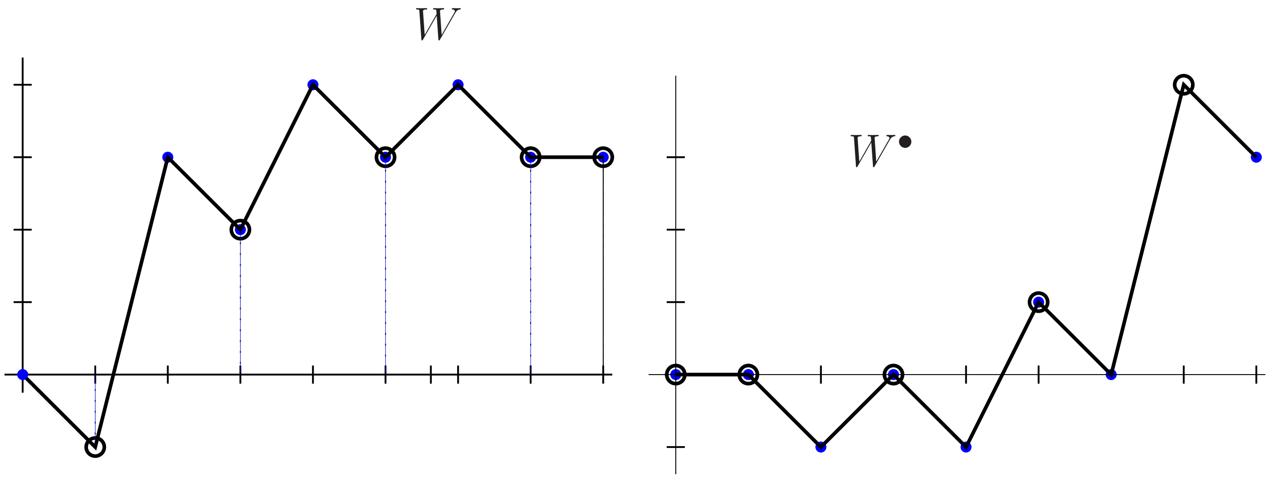




Lemme 1 *(Le Gall & Le Jan) Pour tout $l \in [\![0, n-1]\!]$ et $j \in [\![0, l-1]\!]$, $S_n(j)$ est un minimum à droite pour S_n sur $[\![0, l-1]\!]$ si et seulement si v_j est un ancêtre de v_l.*

A propos des processus

$$W^{\bullet}(j) = W(l) - W(l-j) \text{ pour tout } j \in [\![0, l]\!].$$



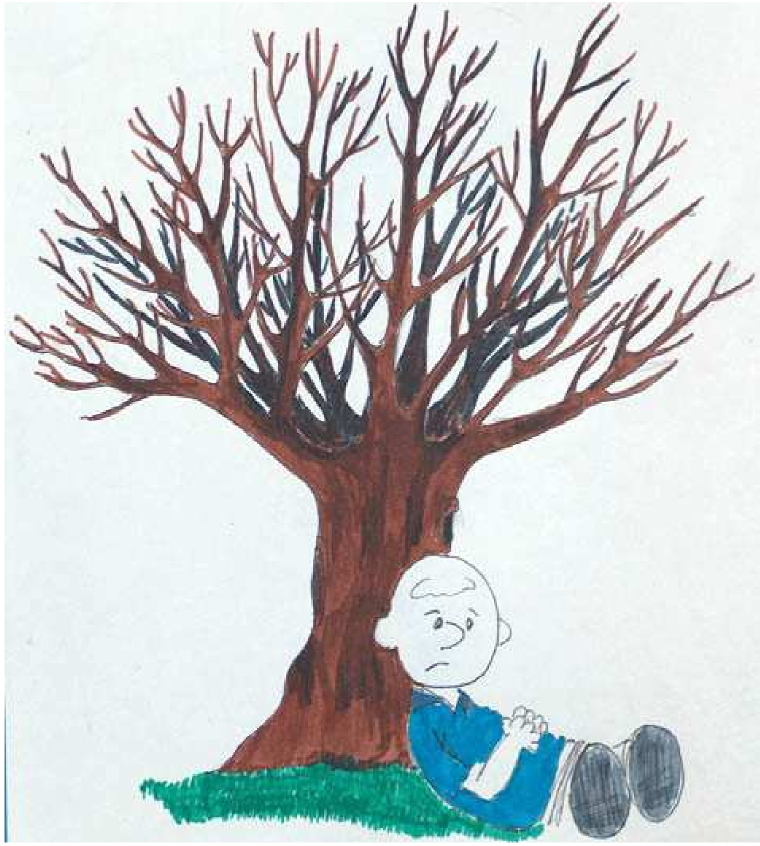
Retournement de la marche:

les minima à droite deviennent des records:

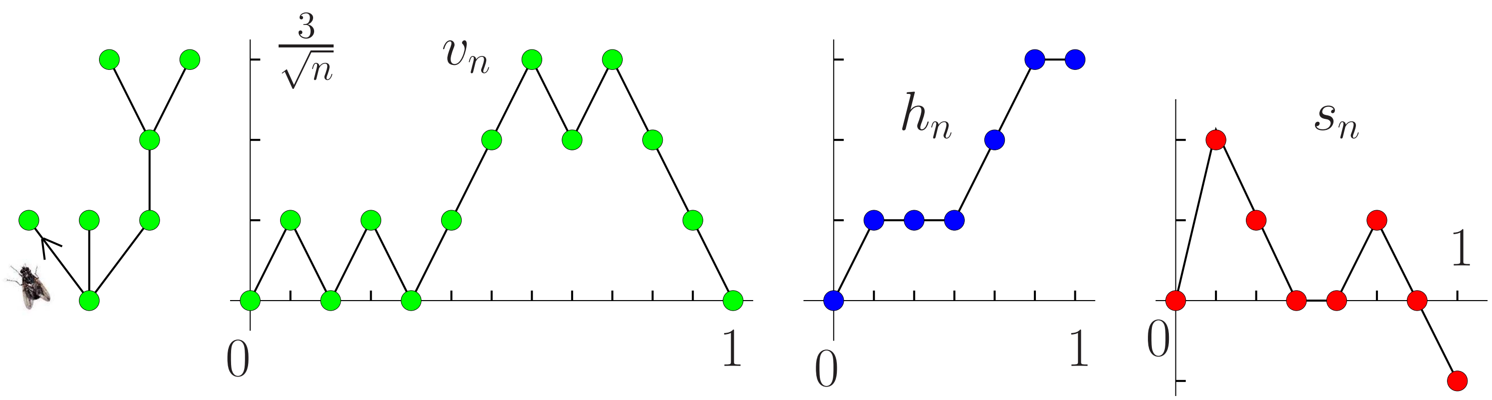
$\sigma^2/2 =$ moyenne d'un accroissement de records.

$$\longrightarrow \qquad \frac{\text{maximum}}{\#\text{accroissement de record}} \sim \frac{\sigma^2}{2}$$

Des grands arbres



Normalisation et plongement dans $(C[0,1], \|.\|_\infty)$

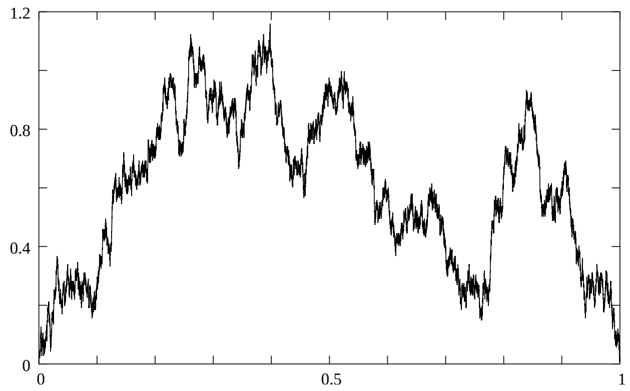


Théorème (M-Mokkadem)

$$(v_n, h_n, s_n) \xrightarrow[n]{faible} \left(\frac{2}{\sigma}e, \frac{2}{\sigma}e, \sigma e\right) \text{ dans } C([0,1]^3)$$

où e est l'excursion brownienne normalisée.

Marginales : [Kaigh (1976), Aldous (1991)]

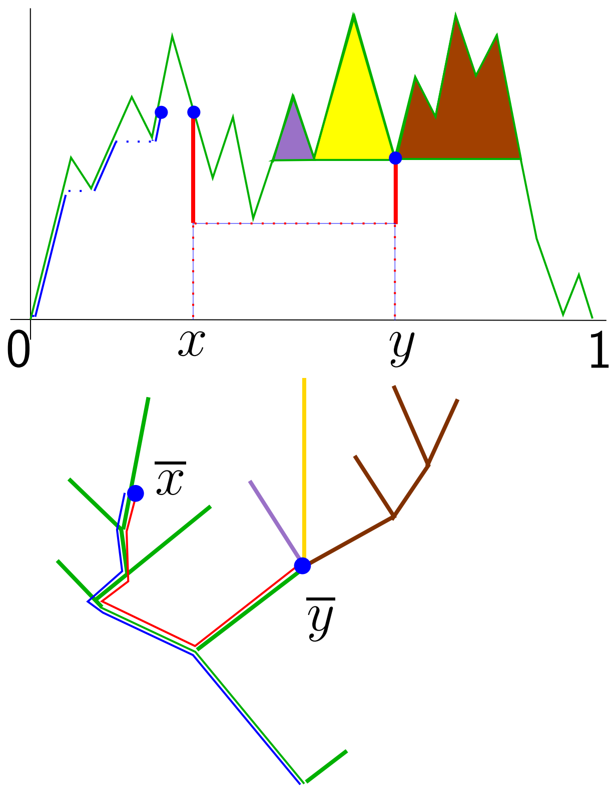


Notion d'arbre continu

À tout $f \in C^+[0,1]$, on associe un arbre $A(f)$:

$$A(f) \;:= [0,1]/\underset{f}{\sim} \text{ où}$$

$$x \underset{f}{\sim} y \iff f(x) = f(y) = \check{f}(x,y) \;:= \min_{u \in [x \wedge y, x \vee y]} f(u)$$



$\star$ $A(f)$ muni de la distance

$$d_f(\overline{x}, \overline{y}) = f(x) + f(y) - 2\check{f}(x,y)$$

est un espace métrique compact, connexe par arcs, sans cycle, etc...

L'espace des arbres $\mathcal{A}$ est muni de la topologie induite par la norme $\|.\|_\infty$ sur $C^+[0,1]$. Ceci en fait un espace Polonais (métrique, complet, séparable)

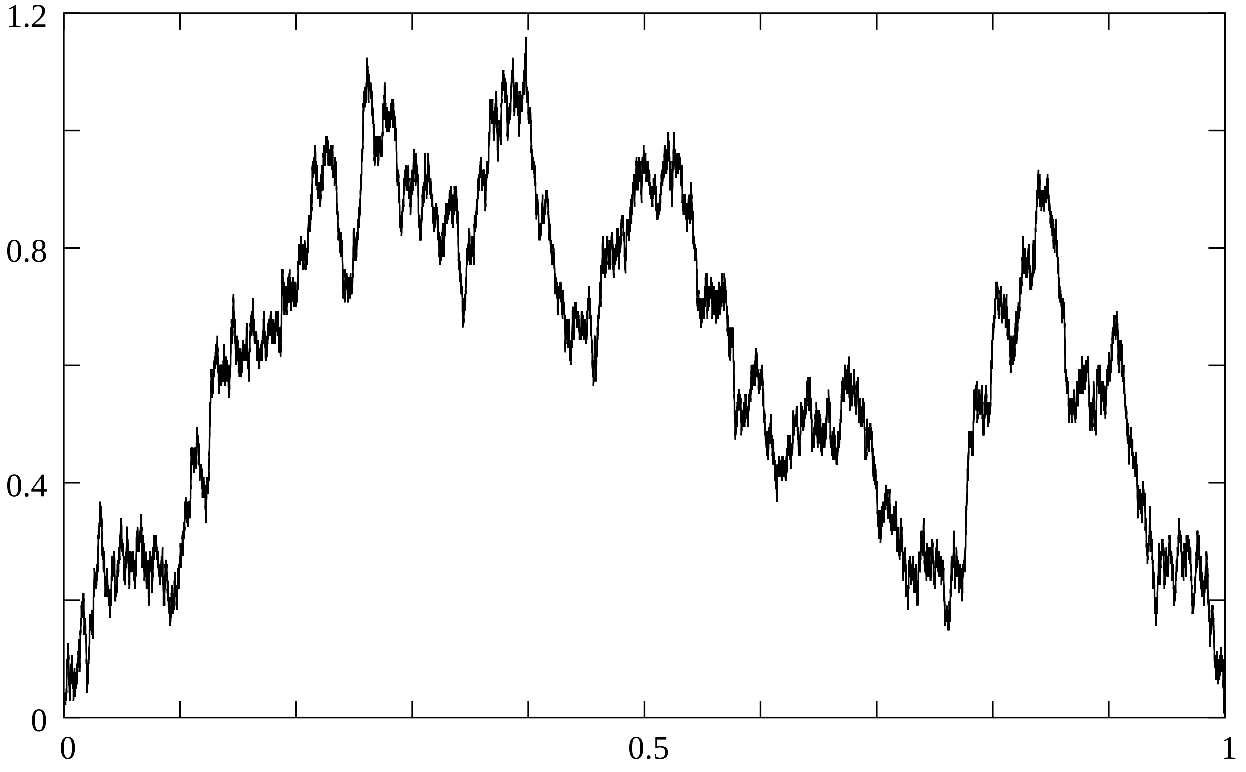
Notion d'arbre continu

Convergence des arbres de GW sous $\mathbb{P}_n$ vers une limite universelle :

l'arbre continu (continuum random tree) .

$$A(v_n) \xrightarrow[n]{faible.} A(\frac{2}{\sigma}e) \text{ dans } (\mathcal{A}, \|.\|_\infty)$$

Un arbre continu (faut encore le quotienter!) :



Et pourquoi les animaux ressemblent à des arbres?

Un animal dirigé ayant 150 000 cellules, choisi uniformément