Une nouvelle bijection pour la génération aléatoire de chemins de m-Dyck

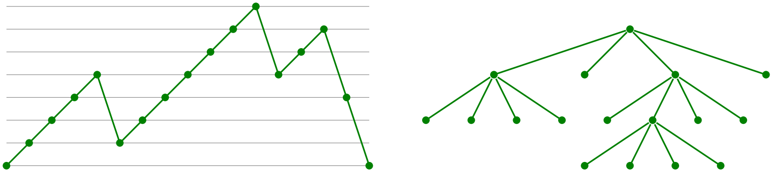
Axel Bacher

LIPN, Université Paris 13

8 mars 2016

Sommaire

Chemins de m-Dyck et arbres $(m+1)$-aires

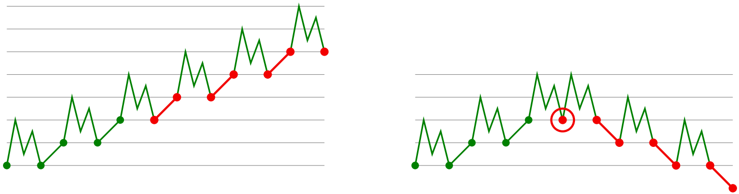


- Chemin de m-Dyck : chemin dans $\mathbb{N}$ de 0 à 0 à pas dans $\{+1, -m\}$.
- Le nombre de chemins de longueur $n = (m+1)d$ est :

$$\frac{1}{n+1}\binom{n+1}{d}$$

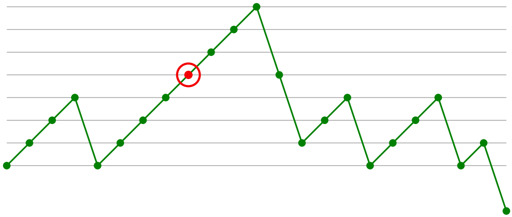
  (nombres de Fuss-Catalan).
- Génération aléatoire basée sur le lemme cyclique [Devroye 2012] : nécessité de tirer une permutation aléatoire, d'où coût $n \log n$.

Préfixes de Dyck et chemins de Łukasiewicz ($m = 1$)



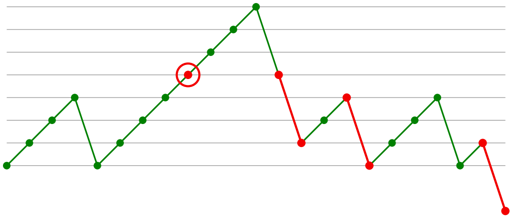
- Il existe une bijection (repliement) des préfixes de Dyck vers les chemins de Łukasiewicz pointés.
- Génération aléatoire des préfixes de Dyck en $\mathcal{O}(n)$.
  [Barcucci, Pinzani & Sprugnoli 1992 ; B., Bodini & Jacquot 2015]

Chemins de m-Łukasiewicz pointés



- Chemin de m-Łukasiewicz : chemin positif sauf à son extrémité.

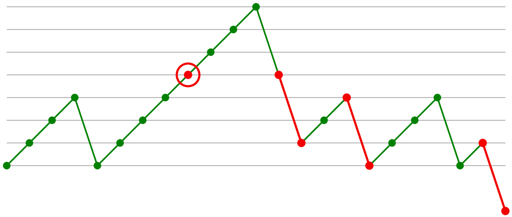
Chemins de m-Łukasiewicz pointés



- Chemin de m-Łukasiewicz : chemin positif sauf à son extrémité.
- Factorisation associée à un point distingué :

$$p\, q_0\mathbf{d} \cdots q_k\mathbf{d}.$$

Chemins de m-Łukasiewicz pointés



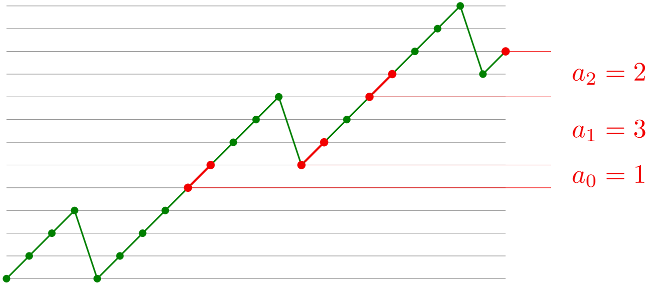
- Chemin de m-Łukasiewicz : chemin positif sauf à son extrémité.
- Factorisation associée à un point distingué :

$$p\, q_0 \mathbf{d} \cdots q_k \mathbf{d}.$$

- Contraintes si hauteur finale $\ell - m - 1$:

$$\begin{cases} 0 \leq h(q_i) \leq m-1, \quad 0 \leq i \leq k-1 \\ 0 \leq h(q_k) \leq \ell - 1. \end{cases}$$

Préfixes de m-Dyck décorés

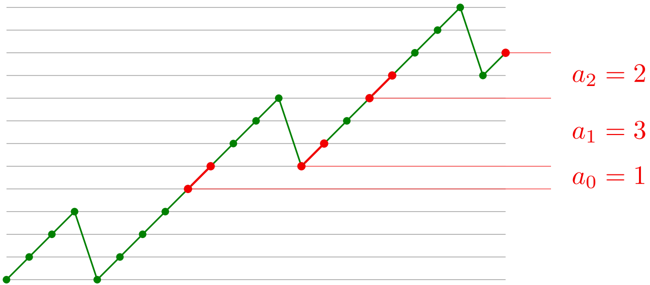


- Soit w un préfixe de m-Dyck de hauteur $(m+1)k+\ell$.
  Une décoration de w est une suite $(a_0, \ldots, a_k)$ avec :

$$\begin{cases} 1 \leq a_i \leq m, \quad i = 0, \ldots, k-1 \\ 1 \leq a_k \leq \ell \end{cases}$$

  (il y a $m^k\ell$ décorations possibles).

Préfixes de m-Dyck décorés



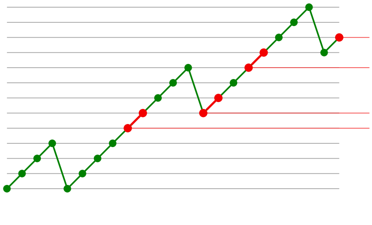
- Soit w un préfixe de m-Dyck de hauteur $(m+1)k + \ell$. Une décoration de w est une suite $(a_0, \ldots, a_k)$ avec :

$$\begin{cases} 1 \leq a_i \leq m, & i = 0, \ldots, k-1 \\ 1 \leq a_k \leq \ell \end{cases}$$

  (il y a $m^k \ell$ décorations possibles).
- Factorisation associée à la décoration :

$$p\, \mathbf{u}q_0 \cdots \mathbf{u}q_k, \qquad h(\mathbf{u}q_i) = a_i.$$

Dépliement et repliement

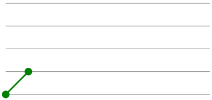


$p\,\mathbf{u}q_0 \cdots \mathbf{u}q_k$ $\qquad$ $p\,q_0\mathbf{d} \cdots q_k\mathbf{d}$

Théorème

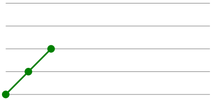
Le repliement est une bijection des préfixes de m-Dyck décorés vers les chemins de m-Łukasiewicz pointés. Déplier ou replier ne demande de lire que la partie du chemin après le point.

Préfixe de m-Dyck aléatoire



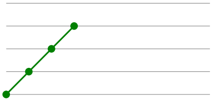
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.

Préfixe de m-Dyck aléatoire



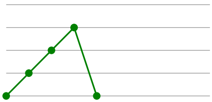
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.

Préfixe de m-Dyck aléatoire



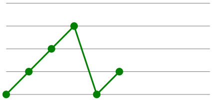
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.

Préfixe de m-Dyck aléatoire



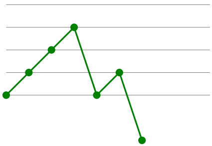
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.

Préfixe de m-Dyck aléatoire



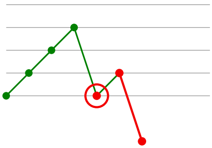
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.

Préfixe de m-Dyck aléatoire



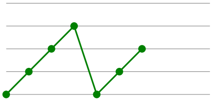
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.

Préfixe de m-Dyck aléatoire



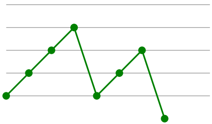
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.

Préfixe de m-Dyck aléatoire



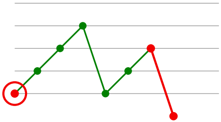
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.

Préfixe de m-Dyck aléatoire



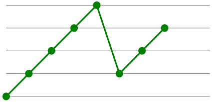
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.

Préfixe de m-Dyck aléatoire



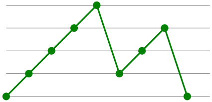
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.

Préfixe de m-Dyck aléatoire



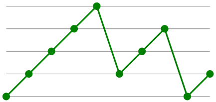
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.

Préfixe de m-Dyck aléatoire



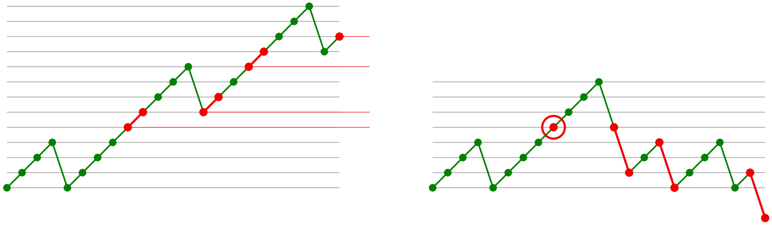
- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.

Préfixe de m-Dyck aléatoire



- On tire des pas $\mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ avec probabilité $\frac{1}{m+1}$.
- Si on crève le plancher, on pointe aléatoirement et on déplie.
- À tout moment, les chemins de hauteur $(m+1)k+\ell$ ont pour probabilité proportionnelle à m^k.

Chemin de m-Łukasiewicz aléatoire



- On tire un préfixe de m-Dyck aléatoire avec la procédure précédente de longueur $n = (m+1)d + \ell$ et hauteur $h = (m+1)k + \ell$.
- On décore aléatoirement ce préfixe et on replie.
- Le résultat est un chemin de m-Łukasiewicz pointé uniforme.

Complexité

Chemin de m-Łukasiewicz aléatoire

$w \leftarrow \varepsilon$
for $i = 1, \ldots, n$ **do**
  $s \leftarrow \mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ sinon
  $w \leftarrow ws$
  if $h(w) < 0$ **then**
    pointer aléatoirement w
    déplier w (oublier la décoration)
  end if
end for
décorer aléatoirement w
replier w (oublier le point)
return w

Complexité

Chemin de m-Łukasiewicz aléatoire

$w \leftarrow \varepsilon$
for $i = 1, \ldots, n$ **do**
  $s \leftarrow \mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ sinon
  $w \leftarrow ws$
  if $h(w) < 0$ **then**
    pointer aléatoirement w
    déplier w (oublier la décoration)
  end if
end for
décorer aléatoirement w
replier w (oublier le point)
return w

- On s'intéresse à la complexité en bits aléatoires et en accès mémoire.

Complexité

Chemin de m-Łukasiewicz aléatoire

$w \leftarrow \varepsilon$	
for $i = 1, \ldots, n$ **do**	
$s \leftarrow \mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ sinon	β
$w \leftarrow ws$	1
if $h(w) < 0$ **then**	
pointer aléatoirement w	$\mathcal{O}(\log i)$
déplier w (oublier la décoration)	$\mathrm{Unif}\{1, \ldots, i\}$
end if	
end for	
décorer aléatoirement w	$\mathcal{O}(\sqrt{n})$
replier w (oublier le point)	$\mathrm{Unif}\{1, \ldots, n\}$
return w	

- On s'intéresse à la complexité en bits aléatoires et en accès mémoire.

Complexité

Chemin de m-Łukasiewicz aléatoire

$w \leftarrow \varepsilon$	
for $i = 1, \ldots, n$ **do**	
$s \leftarrow \mathbf{u}$ avec probabilité $\frac{m}{m+1}$, $\mathbf{d}$ sinon	β
$w \leftarrow ws$	1
if $h(w) < 0$ **then**	
pointer aléatoirement w	$\mathcal{O}(\log i)$
déplier w (oublier la décoration)	$\mathrm{Unif}\{1, \ldots, i\}$
end if	
end for	
décorer aléatoirement w	$\mathcal{O}(\sqrt{n})$
replier w (oublier le point)	$\mathrm{Unif}\{1, \ldots, n\}$
return w	

- On s'intéresse à la complexité en bits aléatoires et en accès mémoire.
- Les branches **if** sont indépendantes de probabilité $\approx \frac{1}{2i}$.

Complexité (suite)

Théorème

Le coût en bits aléatoires et accès mémoire vérifie :

$$\frac{B_n}{n} \xrightarrow{\mathrm{d}} \beta\,; \qquad\qquad \frac{M_n}{n} \xrightarrow{\mathrm{d}} 1 + X + \mathrm{Unif}[0,1].$$

- Le nombre β est le coût en bits aléatoires de $\mathrm{Bernouilli}\big(\frac{1}{m+1}\big)$.
  D'après [Knuth & Yao 1976], on peut prendre :

$$\beta \sim -\tfrac{1}{m+1}\log_2\big(\tfrac{1}{m+1}\big) - \tfrac{m}{m+1}\log_2\big(\tfrac{m}{m+1}\big).$$

Complexité (suite)

Théorème

Le coût en bits aléatoires et accès mémoire vérifie :

$$\frac{B_n}{n} \xrightarrow{\mathrm{d}} \beta\,; \qquad\qquad \frac{M_n}{n} \xrightarrow{\mathrm{d}} 1 + X + \mathrm{Unif}[0,1].$$

- Le nombre β est le coût en bits aléatoires de $\mathrm{Bernouilli}\left(\frac{1}{m+1}\right)$. D'après [Knuth & Yao 1976], on peut prendre :

$$\beta \sim -\tfrac{1}{m+1}\log_2\left(\tfrac{1}{m+1}\right) - \tfrac{m}{m+1}\log_2\left(\tfrac{m}{m+1}\right).$$

- La loi X est définie par :

$$X = \sum_{x \in S} \mathrm{Unif}[0,x],$$

où S est un processus de Poisson de densité $\lambda(x) = \frac{1}{2x}$ sur $]0,1]$.

Complexité (suite)

Théorème

Le coût en bits aléatoires et accès mémoire vérifie :

$$\frac{B_n}{n} \xrightarrow{\mathrm{d}} \beta\,; \qquad\qquad \frac{M_n}{n} \xrightarrow{\mathrm{d}} 1 + X + \mathrm{Unif}[0,1].$$

- Le nombre β est le coût en bits aléatoires de $\mathrm{Bernouilli}\left(\frac{1}{m+1}\right)$. D'après [Knuth & Yao 1976], on peut prendre :

$$\beta \sim -\tfrac{1}{m+1}\log_2\left(\tfrac{1}{m+1}\right) - \tfrac{m}{m+1}\log_2\left(\tfrac{m}{m+1}\right).$$

- La loi X est définie par :

$$X = \sum_{x \in S} \mathrm{Unif}[0,x],$$

où S est un processus de Poisson de densité $\lambda(x) = \frac{1}{2x}$ sur $]0,1]$.

- En particulier :

$$\mathbb{E}(M_n) \sim \frac{7n}{4}, \qquad\qquad \mathbb{V}(M_n) \sim \frac{n^2}{6}.$$

La loi limite X

Résultats sur la loi X

Les cumulants valent :

$$\kappa_n(X) = \frac{1}{2n(n+1)}.$$

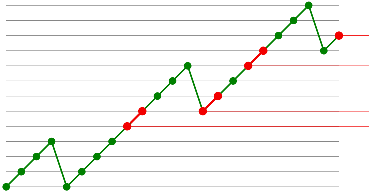
La fonction de répartition $F(x) = \mathbb{P}(X \leq x)$ *est définie par :*

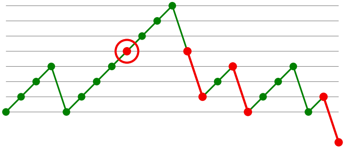
$$\begin{aligned} F(x) + F'(x) + 2xF''(x) &= F(x-1), && x > 0\,; \\ F(x) &= \sqrt{\frac{2e^{1-\gamma}}{\pi}} \sin\sqrt{2x}, && 0 \leq x \leq 1. \end{aligned}$$

La queue de distribution vérifie quand $x \to \infty$ *:*

$$1 - F(x) \sim x^{-x}(\log x)^{-2x}(e/2)^{x+o(x)}.$$

Perspectives





- Peut-on appliquer une méthode semblable pour d'autres objets (clubs de Duchon...) ?