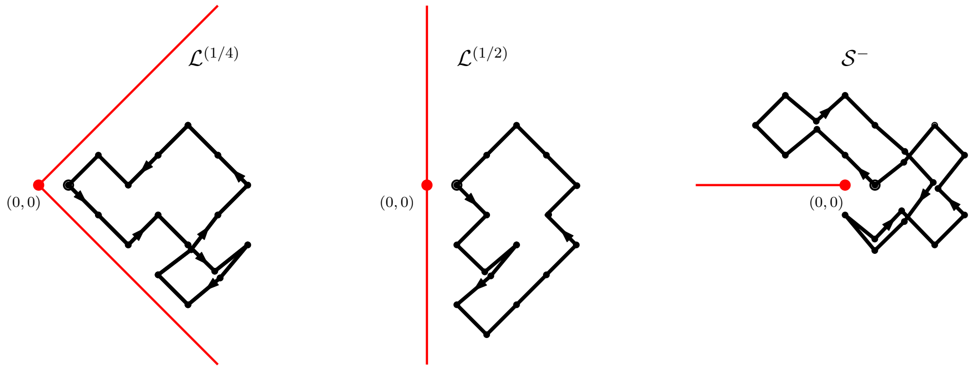
Aspects algorithmiques de la combinatoire (deuxième période)
Exercices du 21/2/25
(Correction)

Exercice 1 : Génération de chemins à pas diagonaux (7 points)

On considère des boucles dans $\mathbb{Z}^2$ à pas $\{(\pm 1, \pm 1)\}$, issus du point $(1, 0)$, et soumis à différentes restrictions illustrées par la figure ci-dessous :



On considère tout d'abord les chemins de $\mathcal{L}_{2n}^{(1/4)}$ formés de $2n$ pas, partant et revenant au point $(1, 0)$ et qui ne visitent que les points (x, y) avec $-x < y < x$ (quart de plan).

1. Montrer que

$$|\mathcal{L}_{2n}^{(1/4)}| = \sum_{k=0}^{n} \binom{2n}{2k} \cdot \frac{1}{k+1}\binom{2k}{k} \cdot \frac{1}{n-k+1}\binom{2n-2k}{n-k}$$

à l'aide d'un argument de mélange de deux chemins de longueur respective $2k$ et $2(n-k)$ pris dans une famille de chemins classiques. En déduire un algorithme de tirage aléatoire uniforme des chemins de $\mathcal{L}_{2n}^{(1/4)}$ et discuter la complexité en temps et en espace des différentes étapes.

Correction : On mélange deux chemins de Dyck de longueur $2i$ et $2j$ avec $i + j = n$ pour obtenir un chemin de $\mathcal{L}_{2n}^{(1/4)}$: Pour le tirage à n fixé on peut tabuler les valeurs en fonction de k et effectuer un premier tirage pour choisir k, puis ensuite engendrer en temps linéaire les deux mots de Dyck et le mot de mélange nécessaires. Cependant le coût de l'étape de choix de k est quadratique. On pourrait aussi tirer directement un mot binaire de longueur $4n$ et le rejeter s'il n'est pas de la forme cherchée (les $2n$ premiers pas déterminent k puis on vérifie si on a deux Dyck de la bonne longueur) mais comme le montre la question suivante le nombre de rejet serait cubique et il n'est pas clair qu'on puisse appliquer une approche florentine).

2. Montrer que le nombre de chemins de la question précédente peut se réécrire plus simplement à l'aide de l'identité de Chu-Vandermonde :

$$\sum_{k=0}^{n} \binom{n+1}{k}\binom{n+1}{k+1} = \binom{2n+2}{n+2}.$$

et en déduire l'asymptotique du nombre de ces chemins quand n tend vers l'infini.

Correction :

$$
\begin{aligned}
|\mathcal{L}_{2n}^{(1/4)}| &= \sum_{k=0}^{n} \binom{2n}{2k} \cdot \frac{1}{k+1}\binom{2k}{k} \cdot \frac{1}{n-k+1}\binom{2n-2k}{n-k} = \\
&= \sum_{k=0}^{n} \frac{(2n)!}{(2k)!(2n-2k)!} \cdot \frac{(2k)!}{(k+1)!k!} \cdot \frac{(2n-2k)!}{(n-k+1)!(n-k)!} \\
&= \frac{1}{(n+1)^2}\binom{2n}{n} \sum_{k=0}^{n} \frac{(n+1)!}{(k+1)!(n-k)!} \cdot \frac{(n+1)!}{k!(n-k+1)!} \\
&= \frac{1}{(n+1)}\binom{2n}{n} \cdot \frac{1}{n+2}\binom{2n+2}{n+1} \\
&\sim cte \cdot n^{-3} \cdot 4^n.
\end{aligned}
$$

3. Donner une interprétation combinatoire de l'identité de Chu-Vandermonde et en déduire une façon de choisir en temps linéaire la valeur de k nécessaire au tirage des chemins de $\mathcal{L}_{2n}^{(1/4)}$ par la méthode de mélange.

(On ne demande pas de donner une preuve combinatoire de la formule simplifiée pour $\mathcal{L}_{2n}^{(1/4)}$: c'est possible mais assez difficile, cf Cori-Dulucq-Viennot 1986, Bernardi 2007.)

Correction : La formule de Chu-Vandermonde revient à raffiner le comptage des mots binaires de longueur $2n+2$ avec n lettres 0 en fonction du nombre k de lettres 0 dans leur préfixe de longueur $n+1$.

Le tirage d'un mot binaire quasi-équilibré de longueur $2n+2$ peut se faire en temps linéaire en adaptant la méthode florentine. La distribution du paramètre k sur ces mots est alors la même que la distribution du paramètre k pour le mélange nécessaire au tirage des chemins de $\mathcal{L}_n^{(1/4)}$, ce qui permet d'éviter de tabuler les valeurs.

On considère maintenant les chemins de $\mathcal{L}_{2n}^{(1/2)}$, constitué de $2n$ pas, issus de $(1,0)$ et y revenant, et qui ne visitent que des points de coordonnée (x,y) avec $x>0$ (demi-plan).

4. Donner un algorithme pour le tirage aléatoire uniforme des chemins de $\mathcal{L}_{2n}^{(1/2)}$ à l'aide d'un argument de bicoloriage de pas appliqué à une famille de chemins classique. Quelle est la complexité de l'algorithme obtenu ?

Correction : La projection d'un chemin de $\mathcal{L}_{2n}^{(1/2)}$ sur l'axe horizontal est un chemin de Dyck de longueur $2n$. Si on bicolorie les pas du chemin projeté suivant que leur antécédant est orienté vers le haut ou vers le bas on obtient des mots de Dyck bicolores tels que la distribution des couleurs de pas soit équilibrée. Il suffit donc d'engendrer un mot de Dyck de longueur $2n$ et un mot binaire équilibré de longueur $2n$ et de les superposer pour obtenir le résultat voulu.

Finalement on considère les chemins de $\mathcal{S}_{2n}^{-}$, constitué de $2n$ pas, issus de $(1,0)$ et y revenant, et qui ne visitent pas les points de la forme $(x,0)$ avec $x \leqslant 0$ (quatre quart de plan). Remarquer que ces chemins se décomposent en séquences de boucles *premières*, qui ne visitent pas le point $(1,0)$ en dehors du premier et dernier pas. On peut montrer (ce n'est pas demandé ici) plus généralement que l'étude des chemins de $\mathcal{S}_{2n}^{-}$ se ramène à celle des chemins de $\mathcal{S}'_{2n}$ issus du point $(0,0)$ et finissant au point $(2,0)$ sans jamais revisiter les points de coordonnée $(x,0)$ avec $x \leqslant 0$ après leur premier pas.

5. En combinant un principe de réflexion avec une conjugaison au niveau du $i$ème pas de w, donner une bijection entre les paires (w,i) où w est un chemin de $\mathcal{S}'_{2n}$ et $i \in [\![0, 2n-1]\!]$, et les chemins dans $\mathbb{Z}^2$ constitués de $2n$ pas de $\{(\pm1,\pm1)\}$, issus de l'origine et terminant sur la droite verticale d'équation $x=2$.

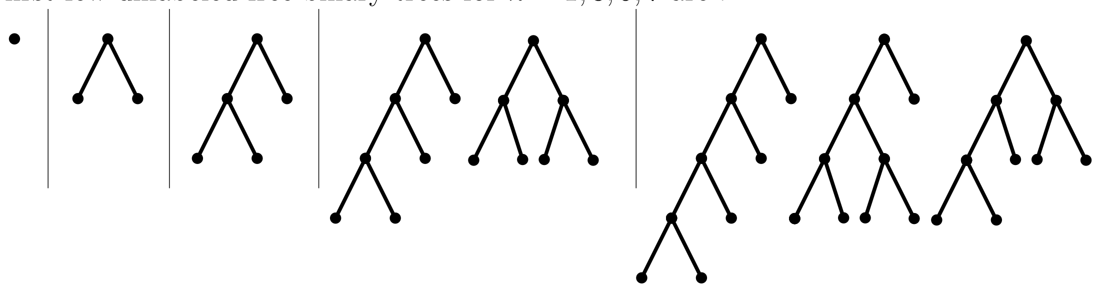
Correction : Supposons que $w = w_1 \cdots w_{2n}$ avec $w_i \in \{(\pm1,\pm1)\}$. On pose $\Psi(w,i) = \bar{w}_i \cdots \bar{w}_{2n} \cdot w_1 \cdots w_i$ où $\bar{u} = (i,-j)$ if $u=(i,j)$ with $u \in \{(\pm1,\pm1)\}$. Par construction cette opération transforme la paire (w,i) en un chemin qui termine sur la droite d'équation $x=2$. Inversement étant donné un tel chemin terminant au point $(2, 2\ell)$, on considère son dernier passage au point visité le plus à gauche sur la droite d'équation $y = \ell$ et on décompose le chemin à ce point. Il reste à prouver que ça marche bien (omis).

6. Déduire de la question précédente un algorithme de génération pour les chemins de $\mathcal{S}'_{2n}$.
Correction : Puisque le facteur n de marquage ne biaise pas la génération, il suffit d'engendrer un chemin qui termine sur la droite $y = 2$. Cela se fait par le même principe que pour $\mathcal{L}_{2n}^{(1/2)}$ sans la contrainte d'équilibrage des couleurs.

Exercice 2 : Random generation of unlabeled free binary trees

The set $\mathcal{U}_n$ of unlabeled free binary trees of size n can be defined as follows :
- For $n = 1$ there is a unique unlabeled free binary tree with one node.
- For $n > 1$ each unlabeled free binary tree of size n consists of a root vertex carrying an unordered pair of unlabeled free binary trees with respective size k, $n-1-k$ for some $1 \leqslant k \leqslant n-2-k$.

The first few unlabeled free binary trees for $n = 1, 3, 5, 7$ are :



Observe that all (unlabeled free) binary trees have odd size and let $\mathcal{U} = \cup_{n\geqslant 0}\mathcal{U}_{2n+1}$. Finally let $U(z) = \sum_{a\in\mathcal{U}} z^{|a|}$ be the ordinary generation funtion of unlabelled free trees.

1. Show that $U(z) = z + \frac{z}{2}U(z)^2 + \frac{z}{2}U(z^2)$.
Correction : $U(z) = \sum_{a\in\mathcal{U}} z^{|b|} = z + \sum_{a=\{b,c\}} z^{|a|} = z + \frac{1}{2}\sum_{a=(b,c)} z^{1+|b|+|c|} + \frac{1}{2}\sum_{a=(b,b)} z^{1+2|b|} = z + \frac{z}{2}\sum_{b\in\mathcal{U}} z^{|b|} \cdot \sum_{c\in\mathcal{U}} z^{|c|} + \frac{z}{2}\sum_{b\in\mathcal{U}} z^{2|b|} = z + \frac{z}{2}U(z)^2 + \frac{z}{2}U(z^2)$.

2. Give the pseudocode of a Boltzmann sampler for the class $\mathcal{U}$.
Correction : Assuming we have an evaluation function Eval :
$\Gamma\mathcal{U}(z)$:
  if $\mathrm{Bern}(z/U(z)) = 1$ return $\bullet$
  if $\mathrm{Bern}(\frac{z}{2}U(z)/(U(z)-z)) = 1$ return $\bullet[\Gamma\mathcal{U}(z), \Gamma\mathcal{U}(z)]$
  let $u = \Gamma\mathcal{U}(z^2)$; return $\bullet[u, u]$.

3. Give the pseudocode of a recursive uniform sampler for the class $\mathcal{U}_{2n+1}$.