CM 4 (Combinatoire)

Adrian Tanasă

Licence Informatique

université de Bordeaux

Annonce

- Étude Prisme - voir flyer et vidéo page Moodle du cours

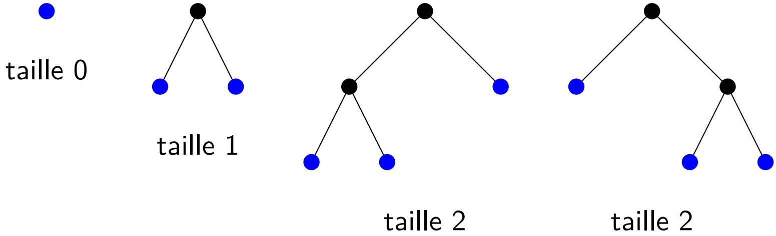
Rappel - ce qu'on a vu la semaine dernière

- Rappel de la définition de la classe combinatoire
- Exemples classe combinatoire
- Pavages et classe combinatoire
- Permutations classe combinatoire
- Coefficients binomiaux ; triangle de Pascal
- Généralisation : coefficients multinomiaux ; application : comptage des anagrammes d'un mot donné

Plan du cours d'aujourd'hui

- Arbres binaires complets (ABC)
- Comptage d'ABC - formule de récurrence (nombres de Catalan)
- Mots de Dyck et chemins de Dyck
- Bijection mots de Dyck - ABC
- Principe de réflexion et formule pour les nombres de Catalan

Exemples d'ABC



On a représenté en bleu les feuilles et en noir les nœuds internes

ABC

Soit $\mathcal{B}$ l'ensemble des ABC.
Cet ensemble peut être décrit **inductivement** :

- un arbre t peut être composé d'un unique nœud racine ; ou
- un arbre t peut être un triplet (r, t_1, t_2) composé d'un nœud racine r, et de deux arbres t_1 et t_2 ; t_1 est alors le sous-arbre gauche, et t_2 le sous-arbre droit.

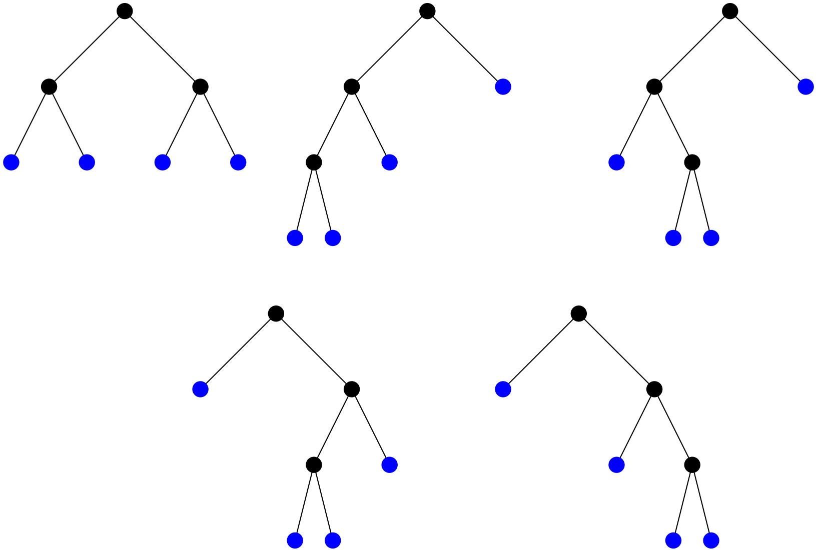
ne portent pas d'étiquettes)
On pourrait prendre comme notion de taille diverses quantités :

- le nombre total de nœuds (forcément, strictement positif)
- le nombre de nœuds internes (peut être nul)
- le nombre de feuilles
- la hauteur

On prend comme **taille** le nombre de **nœuds internes** ; la taille peut donc être n'importe quel nombre entier positif ou nul.

Exercice : ABC de taille 3

Exercice : Trouvez les ABC de taille 3.



Il y en a 5 ABC de taille 3.

Vers le comptage des arbres binaires complets

Prenons un entier $n \geq 0$ quelconque ; on note C_n le nombre d'arbres de taille n.

- Pour $n = 0$, il y a un seul arbre sans nœuds internes : $C_0 = 1$.
- Pour $n > 0$, tout arbre t a un sous-arbre gauche g ; soit k sa taille (son nombre de nœuds internes).
  - k peut valoir n'importe quel entier de 0 à $n - 1$;
  - Si on fixe k, le sous-arbre droit est forcément de taille $n - 1 - k$;
  - Si on fixe k, on a donc C_k choix pour le sous-arbre gauche, et C_{n-1-k} pour le sous-arbre droit.
  - Donc à k fixé, il y a $C_k \cdot C_{n-1-k}$ choix pour l'arbre entier.
  - On fait la somme pour toutes les valeurs de k : pour $n > 0$,

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k}$$

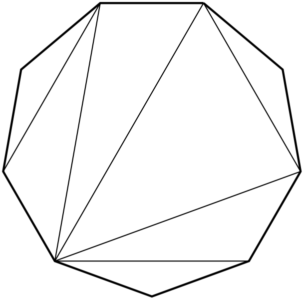
- Ça permet de calculer efficacement les premiers termes **(suite de Catalan)** : $1, 1, 2, 5, 14, 42, 132, 429, 1430 \ldots$ `https://oeis.org` **A000108**

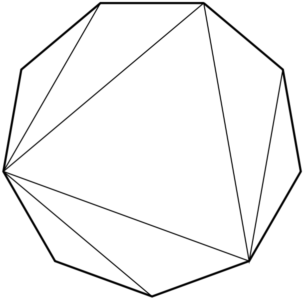
Bijections avec les arbres binaires complets

La suite de Catalan (suite de comptage des arbres binaires complets) est importante en combinatoire : on connaît des dizaines de classes combinatoires, en apparence très différentes, qui ont la même suite de comptage ; cela sous-entend qu'entre deux telles classes, on peut trouver une **bijection** qui **préserve la taille**.

- **Facile :** arbres binaires, non nécessairement complets (taille : nombre total de noeuds) ; la bijection consiste simplement à "effacer" les feuilles ;
- **Assez facile :** mots de Dyck (taille : la moitié de la longueur du mot) ;
- **Presque pareil :** chemins de Dyck (chemins dans une grille, avec des contraintes) ;
- **Moins facile :** "triangulations d'un n-gone régulier" (toutes les façons de découper un polygone régulier à n côtés en triangles, en traçant des diagonales qui ne se coupent pas ; taille : $n-2$, soit le nombre de triangles).

Où sont cachés les arbres ?





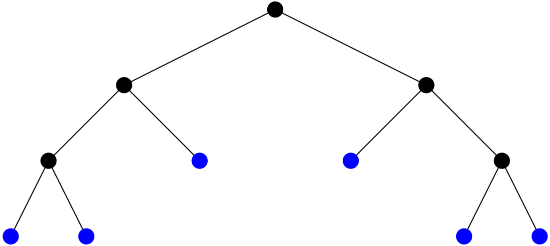
Mots et chemins de Dyck

- **Mots de Dyck :** l'idée est de représenter les mots de parenthèses bien formés : autant de parenthèses ouvrantes que fermantes, plus une condition "on ouvre une parenthèse avant de la fermer"
- **Plus formellement :** mots w sur l'alphabet $\{a, b\}$, avec $|w|_a = |w|_b$, et tels que, pour n'importe quel mot u qui est un "début" (facteur gauche) de w ($w = u.v$ avec v un certain mot), on ait $|u|_a \geq |u|_b$.
- $aababb$ est un mot de Dyck ; $abbaab$ n'en est pas un.
- **Chemin de Dyck :** on représente un mot de Dyck par un **chemin** faisant des pas Nord et Est : a code un pas Nord, b code un pas Est.
- Autant de a que de b : le chemin, partant de $(0, 0)$, se termine sur la diagonale, en un point (n, n).
- Condition de positivité : les points du chemin **ne sont jamais sous** la diagonale (ils ont le droit d'être pile sur la diagonale).

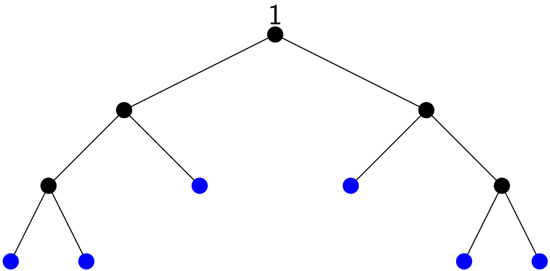
Correspondance : des arbres aux mots de Dyck

- On parcourt les **noeuds de l'arbre** dans l'**ordre préfixe**
- D'abord la racine, puis (récursivement) les noeuds de l'arbre gauche, puis (récursivement) les noeuds de l'arbre droit.
- On obtient un mot en écrivant une lettre par noeud, dans l'ordre du parcours : a pour un noeud interne, b pour une feuille.
- L'arbre a une feuille de plus que de noeuds internes : pour un arbre de taille n, on obtient un mot de longueur $2n+1$, qui se termine forcément par b (le dernier noeud est une feuille).
- On obtient le mot codant l'arbre, en effaçant le b final.
- Ce mot est forcément un mot de Dyck (la condition de positivité n'est pas évidente).
- Ce qui permet d'obtenir la positivité : quand on a déjà visité k noeuds (et écrit k lettres), la différence entre le nombre de a et de b déjà écrits, c'est **un de moins que le nombre de noeuds dont on a visité le parent, mais qu'on n'a pas visités**.

Parcours préfixe et codage

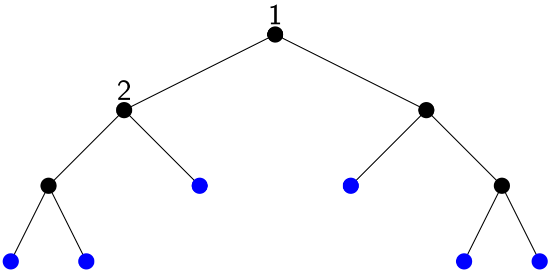


Parcours préfixe et codage



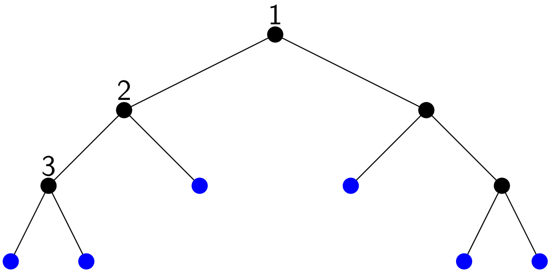
a

Parcours préfixe et codage



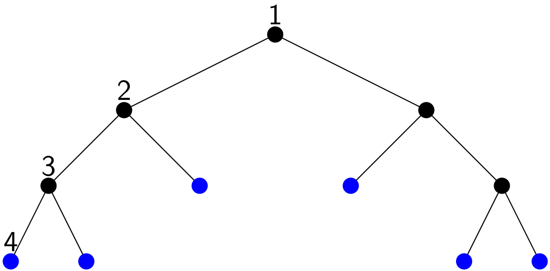
aa

Parcours préfixe et codage



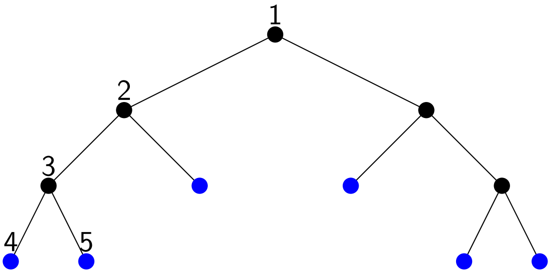
aaa

Parcours préfixe et codage



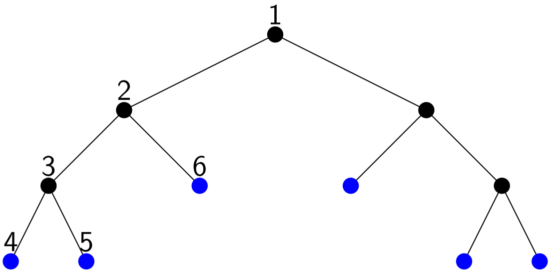
aaab

Parcours préfixe et codage



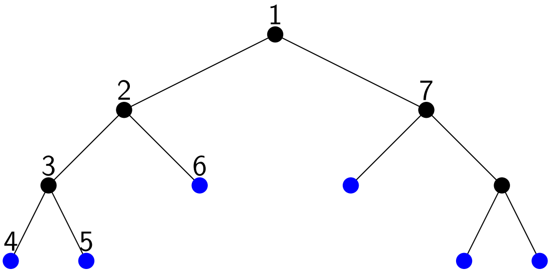
aaabb

Parcours préfixe et codage



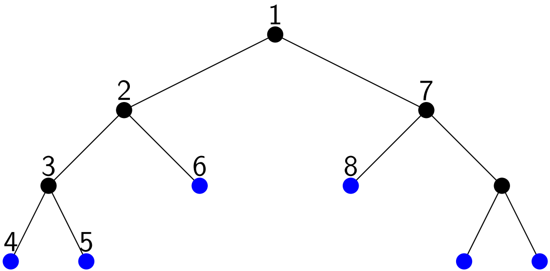
$aaabbb$

Parcours préfixe et codage



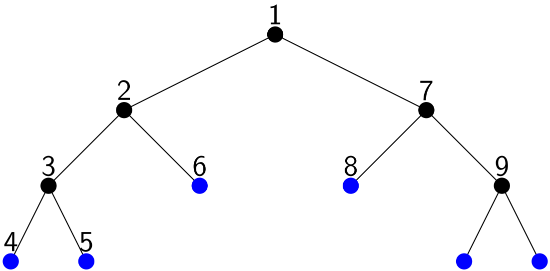
aaabbba

Parcours préfixe et codage



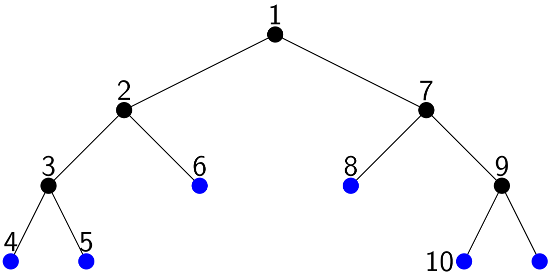
aaabbbab

Parcours préfixe et codage



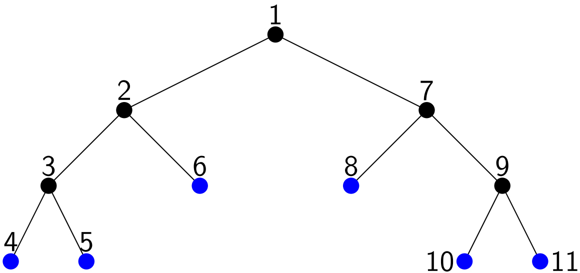
$aaabbbaba$

Parcours préfixe et codage



$aaabbbabab$

Parcours préfixe et codage



$aaabbbabab$

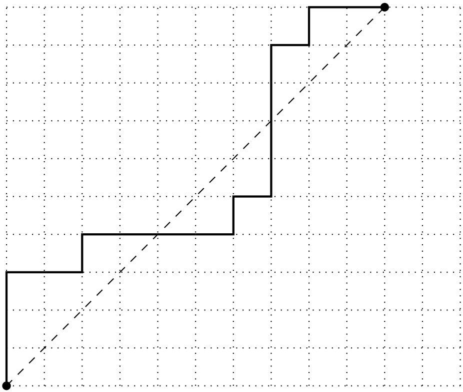
Correspondance inverse : des mots de Dyck aux arbres

- On part d'une "place" pour la racine de l'arbre.
- On lit le mot de gauche à droite, lettre par lettre.
- Quand on lit un a : on remplace la première "place" par un noeud (interne) avec deux "places" pour ses enfants (qui seront en première et deuxième positions). Le nombre de places augmente de 1.
- Quand on lit un b : on remplit la première "place" par une feuille ; le nombre de places diminue de 1.
- Quand le mot se termine, il reste forcément une seule place : on y met une feuille.
- La transformation se programme très bien avec une **pile** (par exemple des listes en OCAML)

Une formule pour les nombres de Catalan ?

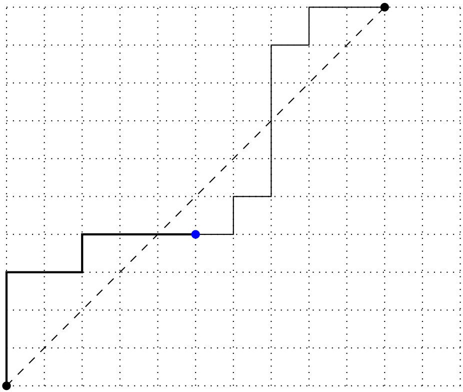
- On voudrait une formule pour C_n, le nombre d'arbres binaire complets de taille n (ou le nombre de mots de Dyck de longueur $2n$).
- On va voir en TD : le nombre de chemins faisant des pas Nord et Est, partant du point $(0,0)$ et finissant au point (a,b), est $\binom{a+b}{a}$.
- Donc les chemins de $(0,0)$ à (n,n) sont au nombre de $\binom{2n}{n}$.
- Mais pour compter les mots (chemins) de Dyck de longueur $2n$, il faut rajouter la contrainte "le chemin ne passe pas en dessous de la diagonale"...
- On va les compter par différence, en trouvant une formule pour le nombre de chemins de $(0,0$ à (n,n) qui **ne sont pas** des chemins de Dyck (ceux qui passent sous la diagonale).

Principe de réflexion



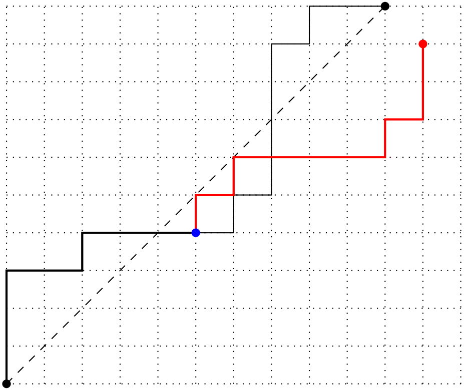
$w = aaabbabbbbabaaaababb$

Principe de réflexion



$w = aaabbabbbbabaaaababb$

Principe de réflexion



$$w = aaabbabbbbabaaaababb$$

$$w' = aaabbabbb\textcolor{red}{ababbbbabaa}$$

Le principe de réflexion

- On part d'un mot chemin C, qui va de $A = (0,0)$ à $B = (n,n)$, mais qui passe **au moins une fois** sous la diagonale.
- On repère le **premier point** D de C qui se trouve **sous** la diagonale ; ses coordonnées sont de la forme $(k, k-1)$.
- On fabrique un nouveau chemin C', qui va de $(0,0)$ à $(n+1, n-1)$:
  - on conserve le début du chemin C entre A et D ;
  - après D, on remplace la fin du chemin C par **son symétrique par rapport à la droite d'équation** $y = x - 1$ (les pas Nord deviennent des pas Est, les pas Est deviennent des pas Nord).
- **Inversement**, partant de n'importe quel chemin C' de $(0,0)$ à $(n+1, n-1)$, on retrouve le chemin C dont il est issu :
  - on repère le premier point D de C' sous la diagonale (il y en a forcément, puisque la fin du chemin est sous la diagonale)
  - on symétrise la partie du chemin qui se trouve après D
- **Donc** il y a autant de chemins de $(0,0$ à (n,n) qui ne sont pas des mots de Dyck, que de chemins de $(0,0)$ à $(n+1, n-1)$, soit $\binom{2n}{n+1}$.

Reprenons. . .

- Il y a $\binom{2n}{n}$ chemins de $(0, 0)$ à (n, n).
- Parmi eux, il y a les chemins de Dyck (il y en a C_n ; on cherche C_n) et ceux qui n'en sont pas.
- Grâce au principe de réflexion, on sait combien ne sont pas des mots de Dyck : $\binom{2n}{n+1}$.
- Et donc on a une équation pour C_n :

$$\binom{2n}{n} = C_n + \binom{2n}{n+1}.$$

La formule

D'après l'équation précédente, on a

$$\begin{aligned}
C_n &= \binom{2n}{n} - \binom{2n}{n+1} \\
&= \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n+1)!(n-1)!} \\
&= \frac{(2n)!(n+1)}{(n+1)!n!} - \frac{(2n)!n}{(n+1)!n!} \\
&= \frac{(2n)!}{(n+1)!n!} \\
&= \frac{1}{n+1}\binom{2n}{n}.
\end{aligned}$$