

Mots de Dyck et nombres de Catalan

Marc Lorenzi

20 octobre 2025

Résumé

La suite des nombres de Catalan intervient dans de nombreux problèmes combinatoires. Dans cet article, nous nous intéressons à l'un de ces problèmes, celui du dénombrement des mots de Dyck. Nous en déduisons les principales propriétés des nombres de Catalan.

Table des matières

1	Introduction	1
1.1	Notations	1
1.2	Le langage des parenthèses	2
2	Mots de Dyck	3
2.1	Introduction	3
2.2	Écriture unique	4
2.3	Une relation de récurrence	5
2.4	Mots mal formés	5
2.5	Quelques formules	8
3	La série génératrice des nombres de Catalan	9
3.1	Une série entière	9
3.2	Convergence aux bornes	10

1 Introduction

1.1 Notations

Étant donné un alphabet A et un mot $w \in A^*$, la longueur de w est $|w|$. Pour toute lettre $x \in A$, $|w|_x$ désigne le nombre d'occurrences de x dans le mot w .

Soit $n = |w|$. Pour tout $j \in \llbracket 0, n-1 \rrbracket$ nous noterons w_j la j^e lettre de w . On a donc

$$w = w_0 \dots w_{n-1}$$

Pour tout $k \in \llbracket 0, n \rrbracket$, nous noterons $w[k]$ le k^e préfixe de w :

$$w[k] = w_0 \dots w_{k-1}$$

On a en particulier $w[0] = \varepsilon$ (le mot vide) et $w[n] = w$.

1.2 Le langage des parenthèses

On se place sur l'alphabet $A = \{ (,) \}$. Pour des raisons de clarté, nous notons parfois ℓ (pour « left ») et r (pour « right ») la parenthèse ouvrante et la parenthèse fermante. Le langage A^* est le *langage des parenthèses*.

On définit les deux morphismes de monoïdes π_ℓ et $\pi_r : A^* \longrightarrow \mathbb{N}$ en posant pour tout $w \in A^*$, $\pi_\ell(w) = |w|_\ell$ et $\pi_r(w) = |w|_r$. L'entier relatif

$$\pi(w) = \pi_\ell(w) - \pi_r(w)$$

est le *poids* de w . La fonction $\pi : A^* \longrightarrow \mathbb{Z}$ est elle aussi un morphisme de monoïdes.

Pour tous $p, q \in \mathbb{N}$, posons

$$\mathcal{L}(p, q) = \{ w \in A^* : \pi_\ell(w) = p, \pi_r(w) = q \}$$

Exemple. Le langage $\mathcal{L}(2, 4)$ contient 15 mots qui sont

$$\begin{aligned} & (()) \quad (()) \quad (()) \quad (()) \quad (()) \\ &)((\quad)((\quad)((\quad)((\quad)((\\ &))(\quad))(\quad))(\quad))(\quad))(\end{aligned}$$

Proposition 1. Pour tous $p, q \in \mathbb{N}$, le cardinal de $\mathcal{L}(p, q)$ est

$$|\mathcal{L}(p, q)| = \binom{p+q}{p} = \binom{p+q}{q}$$

Démonstration. Un mot de $\mathcal{L}(p, q)$ est caractérisé par les p positions de ses parenthèses ouvrantes (ou les q positions de ses parenthèses fermantes) parmi ses $p+q$ lettres. Plus mathématiquement, l'application $\Phi : \mathcal{L}(p, q) \longrightarrow \mathcal{P}_p(\llbracket 0, p+q-1 \rrbracket)$ définie par

$$\Phi(w) = \{ k \in \llbracket 0, p+q-1 \rrbracket : w_k = \ell \}$$

est une bijection. $\square$

Pour visualiser un mot de A^* , il est intéressant de représenter son *profil*, c'est à dire la liste des poids de ses préfixes. Par exemple, soit

$$w = (())(())(())(())(())(())(())(())(())(())(()) \in \mathcal{L}(11, 11)$$

La figure ci-dessous est le profil de w .

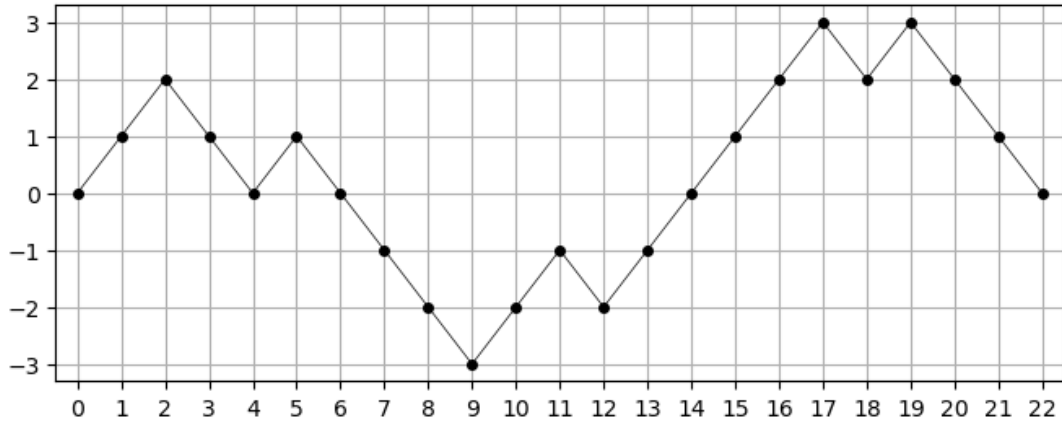


FIGURE 1 – Le profil d'un mot de $\mathcal{L}(11, 11)$

2 Mots de Dyck

2.1 Introduction

Définition 1. Un mot $w \in A^*$ est un *mot de Dyck* lorsque pour tout préfixe p de w , $\pi(p) \geq 0$, et $\pi(w) = 0$.

Intuitivement, les mots de Dyck représentent les parenthésages bien formés dans les expressions algébriques : lorsqu'on parcourt l'expression de la gauche vers la droite, on ne rencontre jamais plus de parenthèses fermantes que de parenthèses ouvrantes.

Remarquons qu'un mot de Dyck w est de longueur paire. En effet, $\pi(w) = \pi_\ell(w) - \pi_r(w) = 0$ et donc

$$|w| = \pi_\ell(w) + \pi_r(w) = 2\pi_\ell(w) = 2\pi_r(w)$$

Nous noterons $\mathcal{D}$ l'ensemble des mots de Dyck et, pour tout $n \in \mathbb{N}$, $\mathcal{D}_n$ l'ensemble des mots de Dyck de longueur $2n$.

Définition 2. Pour tout $n \in \mathbb{N}$, le n^{e} *nombre de Catalan* est

$$C_n = |\mathcal{D}_n|$$

Exemples. $\mathcal{D}_0 = \{\varepsilon\}$, $\mathcal{D}_1 = \{()\}$, $\mathcal{D}_2 = \{()(), (())\}$ et

$$\mathcal{D}_3 = \{()()(), (())(), ()(()), (()), ((()))\}$$

On a donc $C_0 = 1$, $C_1 = 1$, $C_2 = 2$ et $C_3 = 5$. Le langage $\mathcal{D}_4$ contient 14 mots qui sont

$$\begin{array}{cccccc} ()()() & ()()() & ()()() & ()()() & ()(()) \\ (())() & (())() & (())() & ((())) & ((())) \end{array}$$

((())) ((())()) ((()())) (((())))

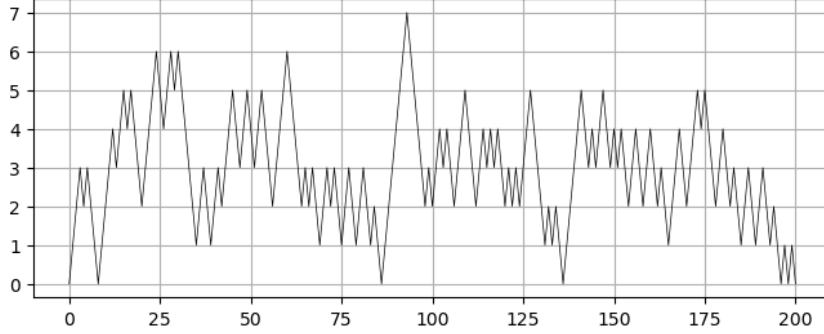


FIGURE 2 – Le profil d'un mot de Dyck de longueur 200

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
C_n	1	1	2	5	14	42	132	429	1430	4862	16796	58786	208012	742900

n	14	15	16	17	18	19	20
C_n	2674440	9694845	35357670	129644790	477638700	1767263190	6564120420

FIGURE 3 – Les 21 premiers nombres de Catalan

2.2 Écriture unique

Dans ce paragraphe, nous montrons que tout mot de Dyck s'écrit de façon unique sous une certaine forme, faisant intervenir d'autres mots de Dyck.

Proposition 2. *Pour tout $w \in \mathcal{D} \setminus \{\varepsilon\}$, il existe $u, v \in \mathcal{D}$ tels que $w = (u)v$.*

Démonstration. Montrons ce résultat par récurrence forte sur n , où $w \in \mathcal{D}_n$. C'est immédiat pour $n = 0$ puisqu'il n'existe aucun mot non vide appartenant à $\mathcal{D}_0$. Soit $n \geq 1$. Supposons la propriété vérifiée pour tous les mots de Dyck de longueur strictement inférieure à $2n$. Soit $w \in \mathcal{D}_n$. Soit $j \in \llbracket 1, 2n - 1 \rrbracket$ le plus petit entier tel que $\pi(w[j]) = 0$. En fait $j \geq 2$ puisque $w_0 = \ell$. On a $w_{j-1} = r$. Écrivons $w = (u)v$ où $u = w_1 \dots w_{j-2}$. On vérifie alors facilement que u et v sont des mots de Dyck. $\square$

Montrons maintenant l'unicité.

Proposition 3. *Soient $u, v, u', v' \in \mathcal{D}$. On suppose que $(u)v = (u')v'$. Alors, $u = u'$ et $v = v'$.*

Démonstration. Pour fixer les idées, $|u| \leq |u'|$. Le mot u est donc un préfixe de u' . Supposons, par l'absurde, que $u \neq u'$. Le mot ur est alors un préfixe de u' . Or,

$$\pi(ur) = \pi(u) - 1 = -1$$

ce qui contredit le fait que $u' \in \mathcal{D}$. Ainsi, $u = u'$. En simplifiant, il vient $v = v'$. $\square$

2.3 Une relation de récurrence

Proposition 4. *Pour tout $n \in \mathbb{N}^*$,*

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

Démonstration. Soit $n \in \mathbb{N}^*$. Posons

$$E = \bigcup_{k=0}^{n-1} \mathcal{D}_k \times \mathcal{D}_{n-1-k}$$

Pour tout couple $(u, v) \in E$, le mot $w = (u)v$ appartient à $\mathcal{D}_n$. On définit ainsi une application $\Phi : E \longrightarrow \mathcal{D}_n$ en posant $\Phi(u, v) = (u)v$.

Nous avons vu que tout $w \in \mathcal{D}_n$ s'écrit de façon unique $w = (u)v$ où $u, v \in \mathcal{D}$. Soit $2k = |u|$. On a $k \in \llbracket 0, n-1 \rrbracket$ et $|v| = 2(n-k-1)$. Ainsi, $(u, v) \in E$. La fonction Φ est donc une bijection. De là, par union disjointe,

$$C_n = |\mathcal{D}_n| = \sum_{k=0}^{n-1} |\mathcal{D}_k| \times |\mathcal{D}_{n-1-k}| = \sum_{k=0}^{n-1} C_k C_{n-1-k}$$

$\square$

La relation de récurrence de la proposition 4 ressemble à une convolution. Cette remarque permet effectivement de déterminer C_n en fonction de n en utilisant des séries. Nous allons choisir une toute autre méthode : au lieu de déterminer le cardinal de $\mathcal{D}_n$, nous allons étudier son complémentaire, l'ensemble des mots de $\mathcal{L}(n, n)$ qui ne sont pas des mots de Dyck. Nous reviendrons à la fin de l'article sur la série génératrice de la suite $(C_n)_{n \in \mathbb{N}}$.

2.4 Mots mal formés

Définition 3. Pour tout $n \in \mathbb{N}$, un *mot mal formé* de longueur $2n$ est un élément de

$$\overline{\mathcal{D}}_n = \mathcal{L}(n, n) \setminus \mathcal{D}_n$$

Nous allons dénombrer les mots mal formés. Il s'avère qu'il existe une bijection entre l'ensemble des mots mal formés et un ensemble dont nous connaissons le cardinal. Prenons un exemple. Considérons le mot

$$w = (()())(((((())()))) \in \mathcal{L}(11, 11)$$

En gras, la première lettre du mot dont le préfixe correspondant est de poids strictement négatif. L'idée est d'échanger, après cette lettre, les parenthèses ouvrantes et les parenthèses fermantes.

On obtient le mot

$$w' = ((\))(\))((\))((\))((\))((\))((\))((\))((\))((\)) \in \mathcal{L}(10, 12)$$

Voici les profils de w et w' .

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 1 & 0 & -1 & -2 & -3 & -2 & -1 & -2 & -1 & 0 & 1 & 2 & 3 & 2 & 3 & 2 & 1 & 0 \\ \begin{pmatrix} 1 & 2 & 1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & -2 & -3 & -4 & -5 & -4 & -5 & -4 & -3 & -2 \end{pmatrix} \end{pmatrix}$$

Si nous effectuons la même opération sur w' , nous revenons à w . La figure ci-dessous montre en noir le profil de w et en vert celui de w' .

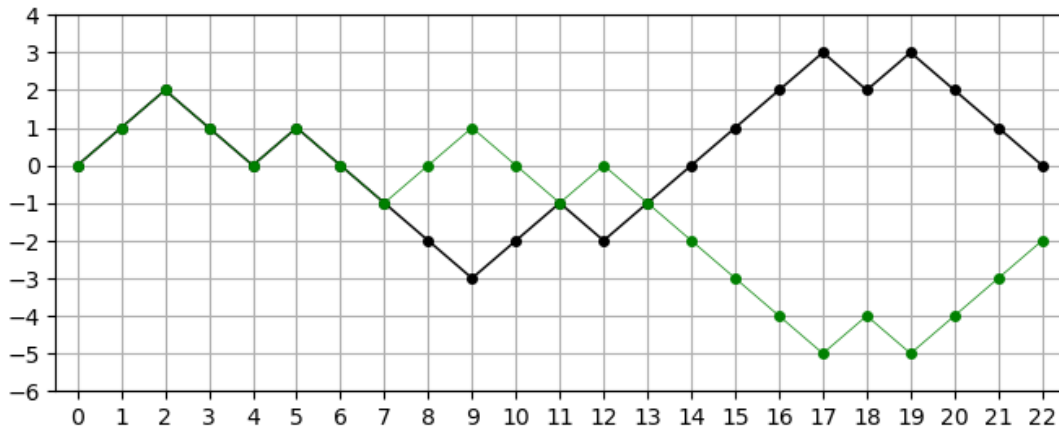


FIGURE 4 – Un mot mal formé et son symétrisé

Nous allons montrer que ce résultat est général : l'application qui à un mot $w \in \overline{\mathcal{D}}_n$ associe son « symétrisé » w' est une bijection de $\overline{\mathcal{D}}_n$ sur $\mathcal{L}(n-1, n+1)$.

Définissons le morphisme de monoïdes $w \mapsto \bar{w}$ de A^* vers A^* par $\bar{\ell} = r$ et $\bar{r} = \ell$. Par exemple,

$$\overline{((()()((=))()()))}$$

Remarquons que pour tout mot $w \in A^*$, $\pi_\ell(\bar{w}) = \pi_r(w)$ et $\pi_r(\bar{w}) = \pi_\ell(w)$.

Notons $\mathcal{E}$ l'ensemble des mots $w \in A^*$ possédant au moins un préfixe de poids strictement négatif. Soit $w = w_0 \dots w_{n-1} \in \mathcal{E}$. Soit $j \in \llbracket 1, n \rrbracket$ le plus petit entier tel que $\pi(w[j]) < 0$. Par minimalité de j , on a $\pi(w[j-1]) \geq 0$, d'où $\pi(w[j-1]) = 0$ et $\pi(w[j]) = -1$. Posons

$$\Phi(w) = w[j] \overline{w_j \dots w_{n-1}}$$

Remarquons que $\Phi(w)$ possède un préfixe de poids strictement négatif. Ainsi, $\Phi : \mathcal{E} \longrightarrow \mathcal{E}$.

Proposition 5. *Pour tout $w \in \mathcal{E}$,*

$$\pi_r(\Phi(w)) = \pi_\ell(w) + 1$$

$$\pi_\ell(\Phi(w)) = \pi_r(w) - 1$$

Démonstration. Soit $w \in \mathcal{E}$. En reprenant les notations ci-dessus, comme $\pi(w[j-1]) = 0$, j est impair. Posons $j = 2k + 1$. On a

$$\pi_\ell(w) = \pi_\ell(w[j]) + \pi_\ell(w_j \dots w_{n-1}) = k + \pi_\ell(w_j \dots w_{n-1})$$

$$\pi_r(w) = \pi_r(w[j]) + \pi_r(w_j \dots w_{n-1}) = k + 1 + \pi_r(w_j \dots w_{n-1})$$

On a également

$$\pi_\ell(\Phi(w)) = \pi_\ell(w[j]) + \pi_\ell(\overline{w_j \dots w_{n-1}}) = k + \pi_r(w_j \dots w_{n-1})$$

$$\pi_r(\Phi(w)) = \pi_r(w[j]) + \pi_r(\overline{w_j \dots w_{n-1}}) = k + 1 + \pi_\ell(w_j \dots w_{n-1})$$

De là,

$$\pi_r(\Phi(w)) = \pi_\ell(w) + 1$$

$$\pi_\ell(\Phi(w)) = \pi_r(w) - 1$$

□

Proposition 6. $\Phi \circ \Phi = id_{\mathcal{E}}$.

Démonstration. Soit $w \in \mathcal{E}$. Toujours avec les mêmes notations, le plus petit préfixe de $\Phi(w)$ de poids strictement négatif est $w[j]$. De là,

$$\Phi \circ \Phi(w) = w[j] \overline{\overline{w_j \dots w_{n-1}}} = w[j] w_j \dots w_{n-1} = w$$

□

La fonction Φ est ainsi une bijection de $\mathcal{E}$ sur $\mathcal{E}$ égale à sa réciproque.

Proposition 7. Pour tout $n \in \mathbb{N}^*$,

$$\Phi(\overline{\mathcal{D}_n}) = \mathcal{L}(n-1, n+1)$$

Démonstration. Soit $n \in \mathbb{N}^*$. Soit $w \in \overline{\mathcal{D}_n}$. On a $\pi_\ell(w) = \pi_r(w) = n$ et donc, par la proposition 5, $\pi_\ell(\Phi(w)) = n-1$ et $\pi_r(\Phi(w)) = n+1$. Ainsi, $\Phi(w) \in \mathcal{L}(n-1, n+1)$.

Inversement, soit $w \in \mathcal{L}(n-1, n+1)$. On a $\pi_\ell(w) = n-1$ et $\pi_r(w) = n+1$ et donc, par la proposition 5, $\pi_\ell(\Phi(w)) = \pi_r(\Phi(w)) = n$. Ainsi, $\Phi(w) \in \mathcal{L}(n, n)$. De plus, $\Phi(w) \in \mathcal{E}$, donc $\Phi(w) \in \overline{\mathcal{D}_n}$.

Il reste à remarquer que $w = \Phi(\Phi(w))$ et donc $w \in \Phi(\overline{\mathcal{D}_n})$. □

Corollaire 8. Pour tout $n \in \mathbb{N}$,

$$|\overline{\mathcal{D}_n}| = \binom{2n}{n+1}$$

Démonstration. Soit $n \in \mathbb{N}$. Si $n = 0$ le résultat est immédiat. Sinon, Φ est une bijection de $\overline{\mathcal{D}}_n$ sur $\mathcal{L}(n-1, n+1)$ et donc

$$|\overline{\mathcal{D}}_n| = |\mathcal{L}(n-1, n+1)| = \binom{2n}{n+1}$$

□

Il est maintenant immédiat d'obtenir une expression explicite des nombres de Catalan.

Proposition 9. *Pour tout $n \in \mathbb{N}$,*

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

Démonstration. Soit $n \in \mathbb{N}$. On a

$$|\overline{\mathcal{D}}_n| = |\mathcal{L}(n, n)| - |\mathcal{D}_n| = \binom{2n}{n} - C_n$$

d'où le résultat par le corollaire 8. □

2.5 Quelques formules

Nous donnons dans ce paragraphe deux autres formules. La première est une petite simplification de la formule de la proposition 9. La seconde est une relation de récurrence reliant C_{n+1} à C_n .

Proposition 10. *Pour tout $n \in \mathbb{N}$,*

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

Démonstration. Soit $n \in \mathbb{N}$. On a

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

□

Proposition 11. Pour tout $n \in \mathbb{N}$,

$$C_{n+1} = \frac{4n+2}{n+2} C_n$$

Démonstration. Soit $n \in \mathbb{N}$. On a

$$\begin{aligned} C_{n+1} &= \frac{1}{n+2} \binom{2n+2}{n+1} \\ &= \frac{(2n+1)(2n+2)}{(n+1)(n+2)} \frac{1}{n+1} \binom{2n}{n} \\ &= \frac{(2n+1)(2n+2)}{(n+1)(n+2)} C_n \\ &= \frac{4n+2}{n+2} C_n \end{aligned}$$

□

Cette relation de récurrence permet d'obtenir les nombres de Catalan $C_0, \dots, C_n$ en $O(n)$ opérations sur des entiers.

3 La série génératrice des nombres de Catalan

3.1 Une série entière

Nous terminons cet article par l'étude de la série génératrice des nombres de Catalan, qui est la série

$$\sum_{n=0}^{\infty} C_n x^n$$

Nous nous restreindrons à considérer $x \in \mathbb{R}$. La première question qui se pose est celle du rayon de convergence de cette série entière.

Proposition 12.

$$C_n \sim \frac{4^n}{n\sqrt{\pi n}}$$

Démonstration. Il suffit d'utiliser la formule de Stirling sur l'expression de la proposition 10.

□

Le résultat précédent montre que le rayon de convergence de notre série est égal à $1/4$. En effet, C_{n+1}/C_n tend vers 4 lorsque n tend vers l'infini.

Proposition 13. Pour tout $x \in]-1/4, 1/4[\setminus \{0\}$,

$$\sum_{n=0}^{\infty} C_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x}$$

Remarquons que pour $x = 0$, la série est évidemment convergente et sa somme est $C_0 = 1$.

Démonstration. La fonction $x \mapsto \sqrt{1+x}$ est développable en série entière en 0, avec un rayon de convergence 1. Pour tout $x \in]-1, 1[$,

$$\sqrt{1+x} = \sum_{n=0}^{\infty} \binom{1/2}{n} x^n$$

Pour tout $n \geq 1$,

$$\begin{aligned} \binom{1/2}{n} &= \frac{\frac{1}{2} \left(\frac{1}{2} - 1\right) \dots \left(\frac{1}{2} - n + 1\right)}{n!} \\ &= (-1)^{n-1} \frac{1 \times 3 \times \dots \times (2n-3)}{2^n n!} \\ &= (-1)^{n-1} \frac{(2n-2)!}{2^{2n-1} (n-1)! n!} \\ &= 2 \frac{(-1)^{n-1}}{4^n} C_{n-1} \end{aligned}$$

Soit $x \in]-1/4, 1/4[$. On a $-4x \in]-1, 1[$ et donc (les termes constants disparaissent),

$$\begin{aligned} 1 - \sqrt{1 - 4x} &= -2 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{4^n} C_{n-1} (-4x)^n \\ &= 2 \sum_{n=1}^{\infty} C_{n-1} x^n \\ &= 2x \sum_{n=0}^{\infty} C_n x^n \end{aligned}$$

d'où le résultat. $\square$

3.2 Convergence aux bornes

Nous allons pour terminer montrer que l'égalité de la proposition 13 reste vérifiée si $x = \pm 1/4$.

Proposition 14. La série de terme général $C_n/4^n$ converge et

$$\sum_{n=0}^{\infty} \frac{C_n}{4^n} = 2$$

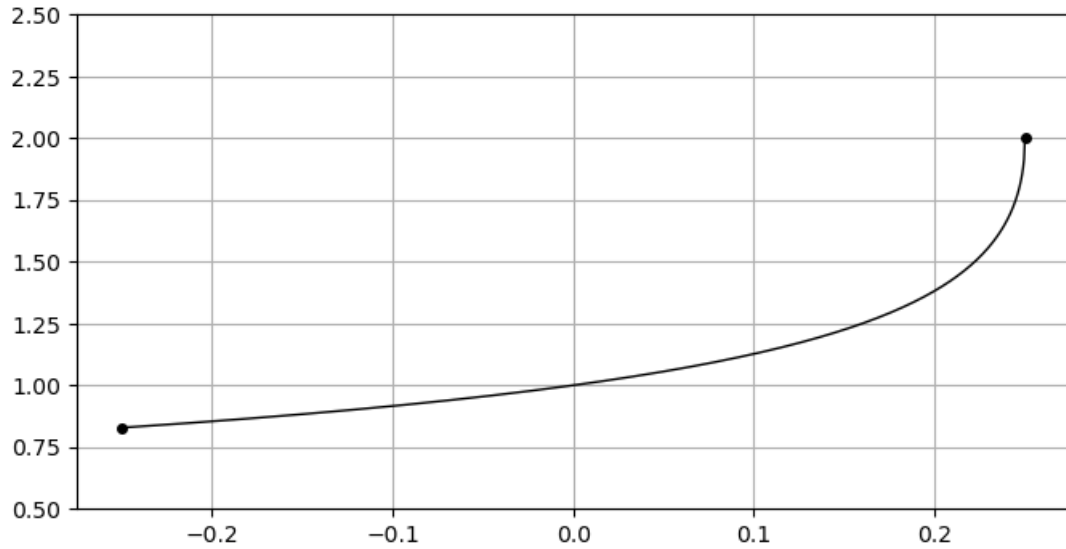


FIGURE 5 – La série génératrice des nombres de Catalan

Démonstration. Cette série est à termes positifs. L'équivalent de C_n donné à la proposition 12 montre que la série converge. Par ailleurs,

$$\frac{1 - \sqrt{1 - 4x}}{2x} \xrightarrow{x \rightarrow 1/4} 2$$

d'où le résultat par le théorème d'Abel. $\square$

Proposition 15. *La série de terme général $(-1)^n C_n / 4^n$ converge et*

$$\sum_{n=0}^{\infty} (-1)^n \frac{C_n}{4^n} = 2\sqrt{2} - 2$$

Démonstration. Par la proposition précédente, cette série est absolument convergente. En faisant tendre x vers $-1/4$ et en utilisant le théorème d'Abel, on obtient le résultat. $\square$

Bibliographie.

- Richard P. Stanley. *Catalan Numbers*. Cambridge University Press, 2015.
- Thomas Koshy. *Catalan Numbers with Applications*. Oxford University Press, 2009.
- Tom Davis. *Catalan Numbers*. Mathcircles.org, 2016.
- Karola Mészáros. *Catalan Numbers*. Cornell University, 2016.