Catalan-Zahlen

Wir haben schon gesehen, dass die größten Binomialkoeffizienten die der Form $\binom{2n}{n}$ sind. Aus diesen werden die *Catalan-Zahlen* auf zwei Arten gebildet:

Definition: Die n-te Catalan-Zahl C_n ist definiert als

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

Bitte überzeugen Sie sich, dass beide Definitionen tatsächlich übereinstimmen!

Mit Hilfe der Stirling-Formel kann man aus der Definition folgende asymptotische Abschätzung für die Catalan-Zahlen gewinnen:

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

Dyck-Wörter

Besondere Bedeutung haben Catalan-Zahlen im Zusammenhang mit den sogenannten *Dyck-Wörtern:*

> *Dyck-Wörter sind Wörter über einem Alphabet, das aus zwei Zeichen (einer „öffnenden Klammer“ a und einer „schließenden Klammer“ b) besteht, wobei das gesamte Wort jeweils einen korrekten Klammerausdruck darstellt.*

Einige Beispiele für Dyck-Wörter:

aababbab *ababab* *aabaabaababbbb* *abaabbabab*

Dagegen ist *abba* kein Dyck-Wort. Warum nicht?

Wir brauchen eine formale Beschreibung, die genau festlegt, welche Wörter Dyck-Wörter sind!

Dyck-Wörter und die Menge D_n

Ein Dyck-Wort ist ein Wort w, das aus öffnenden und schließenden Klammern gebildet ist und die folgenden Bedingungen erfüllt:

1) w enthält gleich viele öffnende wie schließende Klammern.
2) In jedem Präfix von w sind mindestens so viele öffnende wie schließende Klammern.

Aus 1) folgt, dass Dyck-Wörter immer gerade Länge haben.

Mit D_n wird die Menge aller Dyck-Wörter der Länge $2n$ bezeichnet. Es gilt:

Satz: $|D_n| = C_n$ für alle $n \geq 1$

Beweis

Wir definieren zunächst zwei weitere Mengen, E_n und W_n:

$$E_n = D_n b = \{wb \mid w \in D_n\}$$

$$W_n = \{w \in \{a, b\}^{2n+1} \mid |w|_a = n \wedge |w|_b = n+1\}$$

Dann gilt $|E_n| = |D_n|$ und $E_n \subseteq W_n$.

Außerdem hat W_n genau $\binom{2n+1}{n} = (2n+1) \cdot C_n$ Elemente.

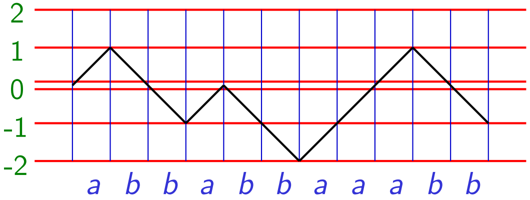
Wir wollen jetzt zeigen, dass die folgende Gleichung gilt:

$$|W_n| = (2n+1) \cdot |E_n|$$

Daraus folgt dann direkt $|D_n| = |E_n| = \frac{1}{2n+1}|W_n| = C_n$.

Klammergebirge für $|W_n| = (2n+1) \cdot |E_n|$

Ein Klammergebirge für ein Wort in $\{a, b\}^*$ erhalten wir, wenn wir graphisch jeden Buchstaben a als Aufstieg um 1, und jedes b als Abstieg um 1 darstellen:



Klammergebirge für das Wort *abbabbaaabb*, das zu W_n gehört, aber nicht zu E_n.

Die Klammergebirge für Wörter aus W_n beginnen immer auf der Null-Linie und enden auf der (-1)-Linie.

Bei Wörtern aus E_n wird dabei die Null-Linie erst beim letzten Buchstaben erstmals unterschritten!

Beweis (Forts.)

Wir behaupten nun, dass zu jedem Wort $w \in W_n$ genau ein Wort aus E_n existiert, das aus w durch eine zyklische Shift-Operation hervorgeht. Da $|w| = 2n + 1$ gilt, folgt hieraus die Behauptung!

Sei $d \leq -1$ die minimale Höhe, die im Klammergebirge zu w erreicht wird. Sei u der Teil des Wortes w bis zum erstmaligen Erreichen dieser Höhe, v der Rest. Das heißt, $w = uv$.

Das Klammergebirge zu vu kann im v-Teil die Null-Linie nicht unterschreiten, und im u-Teil auch erst ganz zum Schluss!

Beginnen wir das geshiftete Wort an irgendeiner anderen Stelle, dann unterschreitet das Klammergebirge die Null-Linie in dem Moment, in dem das Ende von u erreicht wird.

Daher ist vu ein Wort aus E_n, aber die übrigen $2n$ aus w durch zyklischen Shift erhältlichen Wörter sind nicht in E_n.

Saturierte Binärbäume

Informell sind saturierte Binärbäume Bäume, bei denen jeder Knoten entweder zwei Nachfolgeknoten hat (*innere Knoten*) oder gar keinen Nachfolgenknoten (*Blatt*). Formaler:

1) Ein einzelner Knoten ist *saturierter Binärbaum der Höhe 0*. Er hat keine inneren Knoten, aber ein Blatt (zugleich die Wurzel!).
2) Sind T_1 und T_2 saturierte Binärbäume mit Knotenmengen V_1 bzw. V_2, wobei $V_1 \cap V_2 = \emptyset$ und $v \notin V_1 \cup V_2$, dann bildet das Tripel (v, T_1, T_2) einen neuen saturierten Binärbaum mit der Wurzel v, linkem Teilbaum T_1 und rechtem Teilbaum T_2. Innere Knoten des neuen Baums sind v und die inneren Knoten von T_1 und T_2, die Höhe ist 1 plus das Maximum der Höhen beider Teilbäume. Blätter sind diejenigen beider Teilbäume.

Achtung: Die Reihenfolge der Teilbäume ist hier von Bedeutung.

Allgemeine Binärbäume

Ein *Binärbaum* ist ein Baum, bei dem jeder Knoten maximal zwei Nachfolgeknoten hat. Formal kann man das so definieren:

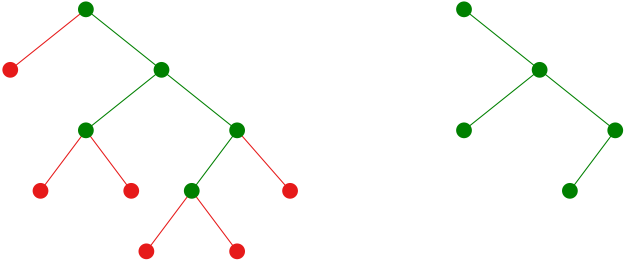
Ein Binärbaum entsteht aus einem saturierten Binärbaum durch Weglassen aller Blätter. Insbesondere ist auch der leere Graph ein Binärbaum.

Geht man von den saturierten Binärbäumen auf diese Weise zu allgemeinen Binärbäumen über und fügt dann wieder an jeden „freien Platz“ ein Blatt an, so erhält man den ursprünglichen Binärbaum. Umgekehrt geht das ebenso. Daraus erhalten wir eine 1:1 Korrespondenz zwischen Binärbäumen mit n Knoten und saturierten Binärbäumen mit n inneren Knoten.

Beachte zusätzlich: Ein saturierter Binärbaum mit n inneren Knoten hat $n+1$ Blätter, insgesamt also $2n+1$ Knoten.

Beispiele

Wir zeigen einen saturierten Binärbaum mit zugehörigem allgemeinem Binärbaum:



Bitte beachten: Die folgenden Bäume zählen wir als 4 verschiedene Binärbäume.

Anzahl verschiedener Binärbäume

Wir wollen eine Formel, die besagt, wieviele verschiedene Binärbäume mit n Knoten es gibt. Wegen der 1:1 Korrespondenz ist es hinreichend zu untersuchen, wieviele saturierte Binärbäume mit n inneren Knoten (und $n+1$ Blättern) es gibt.

Satz: Die Anzahl saturierter Binärbäume mit n inneren Knoten ist C_n.

Hierbei ist C_n die n-te Catalan-Zahl, d.h. $C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{1}{2n+1}\binom{2n+1}{n}$.

Für den Beweis benutzen wir die Tatsache, dass $C_n = |E_n|$ gilt, wobei E_n die Menge der Wörter der Länge $2n+1$ ist, die aus Dyck-Wörtern durch Anhängen eines b entstehen.

Beweis des Satzes

Wir definieren induktiv eine Abbildung von der Menge E_n in die Menge der saturierten Binärbäume mit n inneren Knoten. Auf Grund der Bijektivität dieser Abbildung ist damit der Beweis komplett.

$n = 0$: Das einzige Wort in E_0 ist das Wort b, dem wir den einzigen saturierten Binärbaum ohne innere Knoten, bestehend nur aus der Wurzel, zuordnen.

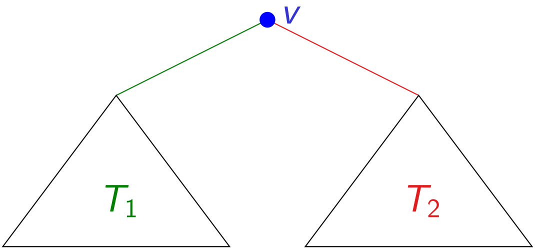
Nun sei also $n > 0$:

Jedes Wort in E_n ist von der Form $aubvb$ für Dyck-Wörter u und v; und diese Zerlegung ist eindeutig. (Bitte überprüfen Sie die Richtigkeit dieser Behauptung!)

Jetzt liegen die Teilwörter ub und vb auch jeweils in einem E_m mit $m < n$. Sei also $ub \in E_{m_1}$ und $vb \in E_{m_2}$. Dann gehört zu ub induktiv bereits ein saturierter Binärbaum T_1 mit m_1 inneren Knoten, zu vb ein T_2 mit m_2 inneren Knoten. Wir beachten noch, dass ub die Länge $2m_1 + 1$ und vb die Länge $2m_2 + 1$ hat. Das gesamte Wort $aubvb$ hat damit Länge $2(m_1 + m_2) + 3 = 2(m_1 + m_2 + 1) + 1$. Daher gilt $n = m_1 + m_2 + 1$.

Abschluss des Beweises

Dem Wort *aubvb* ordnen wir jetzt den saturierten Binärbaum (v, T_1, T_2) zu (mit einem neuen Wurzelknoten v und disjunkten Teilbäumen T_1 und T_2):



Die Wohldefiniertheit der konstruierten Abbildung folgt aus der Eindeutigkeit der Zerlegung des Ausgangswortes in *aubvb* mit Dyck-Wörtern u und v. Injektivität und Surjektivität sind leicht mit Induktion einzusehen.

Cliquen und unabhängige Mengen

Sei $G = (V, E)$ ein ungerichteter Graph, d.h. $E \subseteq \binom{V}{2}$. Für Teilmengen $U \subseteq V$ bezeichnet man mit G_U den *induzierten Teilgraph* $G_U = (U, E_U)$, wobei $E_U := E \cap \binom{U}{2}$ ist.

Eine *Clique* ist eine Teilmenge $U \subseteq V$ mit $E_U = \binom{U}{2}$, das heißt, es gilt $\binom{U}{2} \subseteq E$. In diesem Fall ist G_U ein vollständiger Graph!

Eine *unabhängige Menge* U im Graph $G = (V, E)$ ist eine Teilmenge $U \subseteq V$ mit $E_U = \emptyset$. Hier ist G_U ein kantenloser Graph.

Welches sind die größten Cliquen in P_n, C_n, K_n und $K_{m,n}$?

Antwort: In P_n höchstens Größe 2, in C_n meist auch, nur C_3 ist eine Ausnahme.
In K_n ist der ganze Graph eine Clique, in $K_{m,n}$ wieder nur Größe 2.

Und wie steht es mit unabhängigen Mengen in diesen Graphen?

Antwort bitte selbst überlegen...

Der Satz von Ramsey

Es gibt viele Versionen des Satzes von Ramsey, denn man kann zunächst den endlichen vom unendlichen Fall unterscheiden, aber zusätzlich gibt es auch die Unterscheidung, ob man nur über *normale* Graphen mit Kanten und „Nichtkanten“ spricht, oder ob man beliebige Kantenfärbungen betrachtet. Und schließlich kann man statt Kanten, also Elementen von $\binom{V}{2}$, auch sogenannte *k-Hyperkanten*, also Elemente von $\binom{V}{k}$ betrachten.

Wir beweisen hier die *Basisversion*, also den folgenden Satz:

Satz: Für alle $n \in \mathbb{N}$ gibt es ein $N \in \mathbb{N}$, so dass jeder Graph mit N Knoten eine Clique der Größe n oder eine unabhängige Menge der Größe n enthält.

Ramsey für $n = 3$

Zum Eingewöhnen beweisen wir den Satz jetzt für den Fall $n = 3$.
Wir können $N = 6$ wählen und zeigen, dass dieser Wert optimal ist:

Beh.: Jeder Graph mit 6 Knoten enthält entweder eine Clique der Größe 3 oder eine unabhängige Menge der Größe 3. Für Graphen mit 5 Knoten ist das nicht der Fall.

Für den Beweis nehmen wir zunächst an, dass ein Knoten v mit mindestens drei Kanten existiert. Die Nachbarknoten seien u_1, u_2, u_3. Falls diese eine unabhängige Menge bilden, sind wir fertig. Andernfalls gibt es eine Kante zwischen einem u_i und einem u_j (mit $i \neq j$). Dann bilden v, u_i, u_j eine Clique der Größe 3.

Gibt es keinen solchen Knoten, dann hat jeder Knoten drei nichtbenachbarte Knoten! Auch hier betrachten wir einen Knoten v mit drei Nichtnachbarn u_1, u_2, u_3. Bilden diese drei Knoten eine Clique, sind wir fertig. Andernfalls gibt es $i \neq j$ so, dass u_i und u_j nicht benachbart sind. Dann sind v, u_i, u_j eine unabhängige Menge.

Zur Optimalität: C_5 hat 5 Knoten und weder Clique noch unabh. Menge der Größe 3.