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.

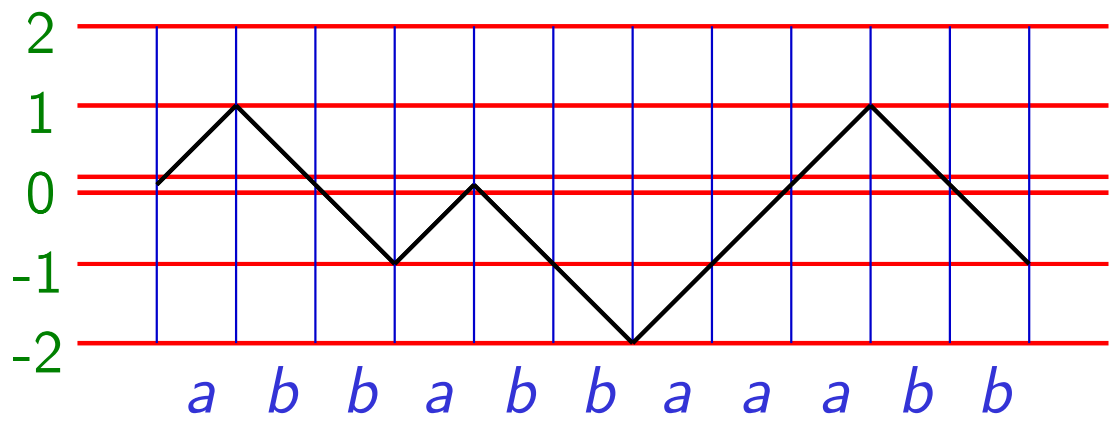
Wir wollen jetzt zeigen, dass die folgende Gleichung gilt:

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

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.