Die Anzahl C_n der binären geordneten Bäume mit n Knoten.

Ein binärer geordneter knotenmarkierter Baum,

- dessen Knotenmarkierung Elemente einer geordneten Menge sind und
- in dem für jeden Knoten gilt, dass seine Markierung größer als alle Markierungen im linken Unterbaum ist und kleiner oder gleich zu allen Markierungen im rechten Unterbaum ist,

heißt (binärer) Suchbaum.

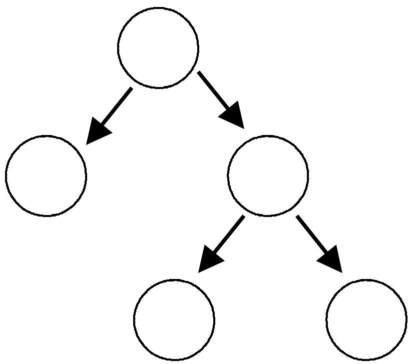
Beachte: Jede sortierte Folge $a_1\ a_2\ a_3 \dots a_n$ kann man in jeden binären geordneten Baum mit n Knoten auf genau eine Weise so einfügen, dass die sortierte Folge exakt die inorder-Reihenfolge ist.

Beispiel (für n=5):

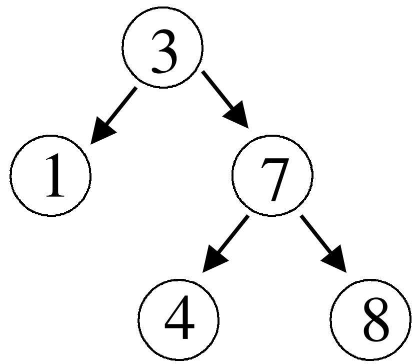
Folge: 1, 3, 4, 7, 8

+

binärer geordneter Baum:



$\Rightarrow$



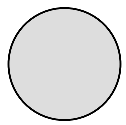
Wenn man einen besonders guten (z.B. einen "optimalen") Suchbaum zu einer Folge mit n Elementen (und eventuell weiteren Attributen) ermitteln will, dann braucht man also nur alle binären geordneten Bäume mit n Knoten durchzuprobieren und das Maximum auszuwählen, allerdings nur, wenn deren Anzahl nicht allzu groß ist.

Wieviele solcher binärer geordneter Bäume gibt es?

Die Anzahl C_n der Suchbäume mit n Knoten.

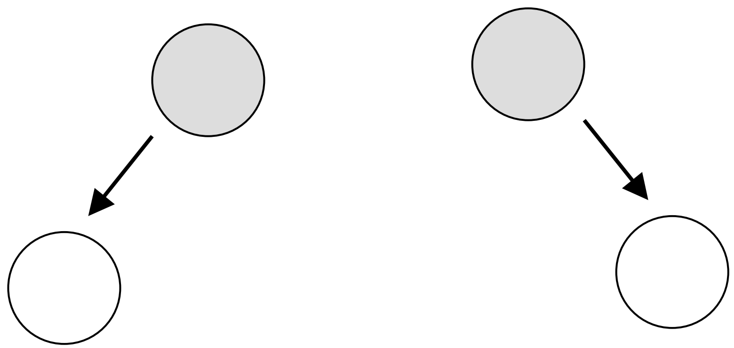
Die Wurzel des binären geordneten Baums (= Suchbaums) ist hier grau dargestellt.

n = 1



1 Suchbaum

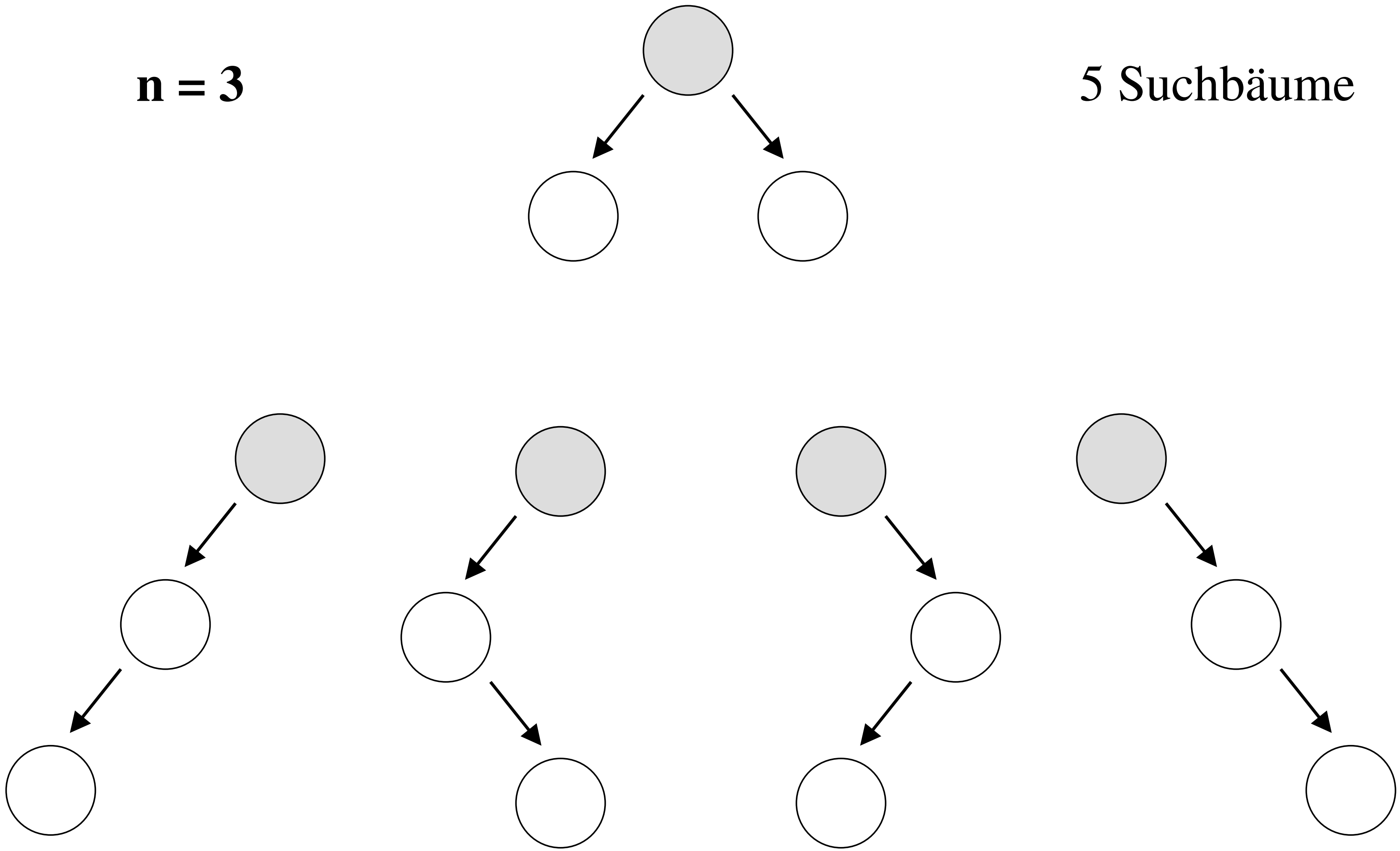
n = 2



2 Suchbäume

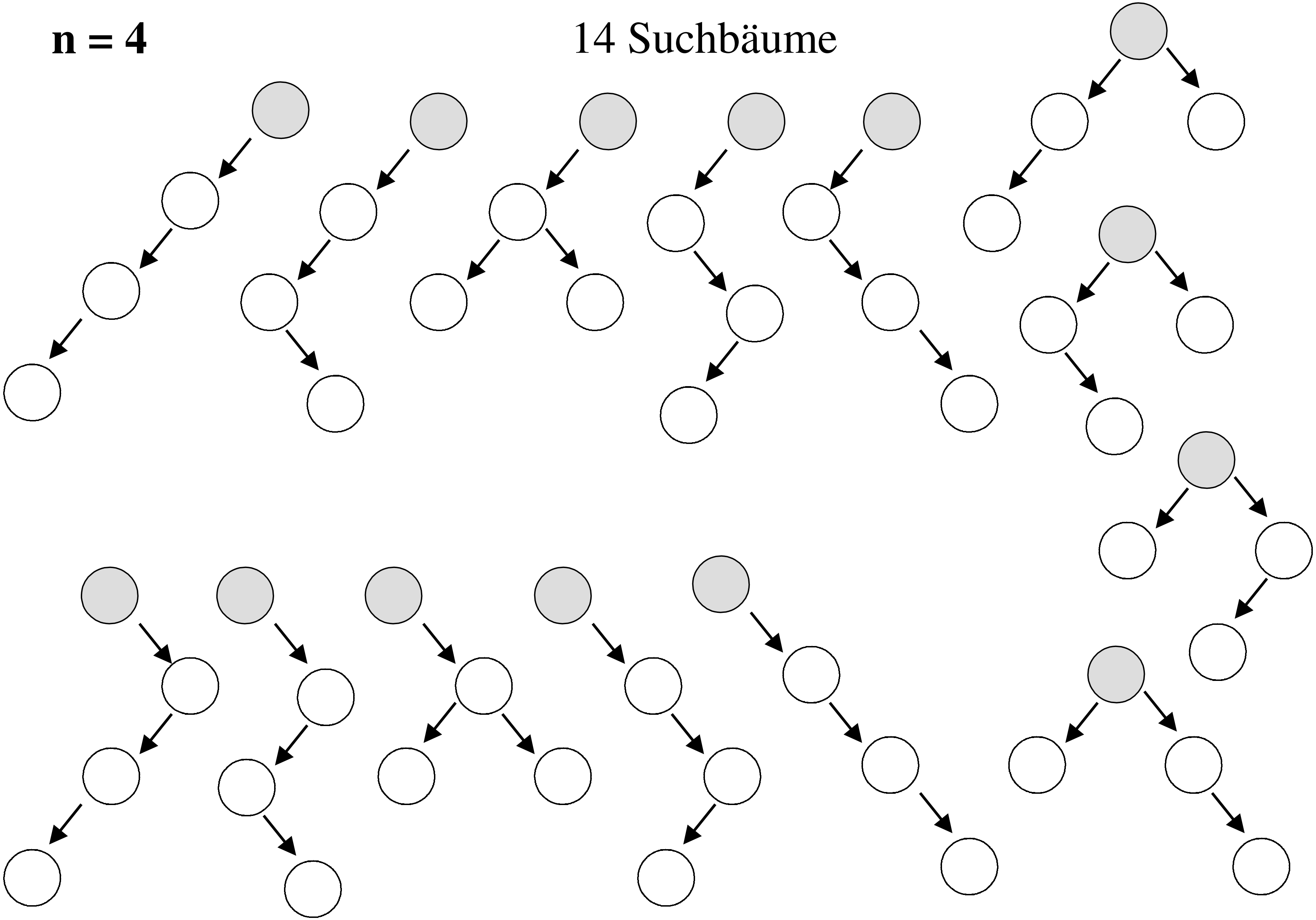
n = 3

5 Suchbäume



n = 4

14 Suchbäume



Durch Ausprobieren erhält man einige Werte:

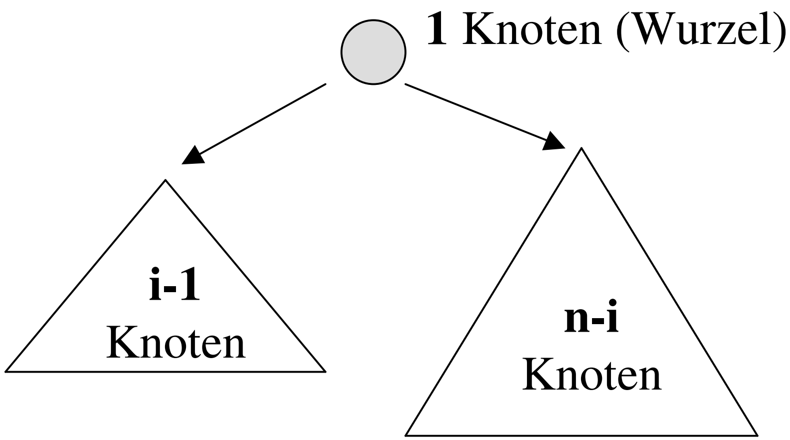
n	Anzahl
1	1
2	2
3	5
4	14
5	42
6	132
7	429
8	1410

Sei C_n die Anzahl dieser Suchbäume mit n Knoten, dann gilt:

$C_0 = 1$, und für alle $n \geq 1$:

$$C_n = \sum_{i=1}^{n} C_{i-1} \cdot C_{n-i}$$

wegen:



Satz: $C_n = \dfrac{1}{n+1}\dbinom{2n}{n}$ Catalansche Zahlen

Näherung: $C_n \approx \dfrac{4^n}{(n+1)\sqrt{\pi \cdot n}}$

Wir beweisen nur den Satz. Die Näherung ergibt sich direkt aus dem Satz unter Verwendung der Stirlingschen Formel für die Fakultät.

Die Anzahl C_n der Suchbäume ist gleich der Anzahl der Möglichkeiten, um n „Klammer auf" und n „Klammer zu" wie in korrekt geklammerten Ausdrücken aneinander zu reihen. Z.B. gibt es genau 5 korrekte Klammerungmöglichkeiten für n = 3, nämlich:

((())), (()()), (())(), ()(()), ()()().

Behauptung: Die Anzahl C_n der Suchbäume ist gleich der Anzahl der Möglichkeiten, um n „Klammer auf“ und n „Klammer zu“ wie in korrekt geklammerten Ausdrücken aneinander zu reihen:

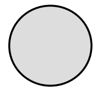
((())), (()()), (())(), ()(()), ()()().

Wir geben diesen Zusammenhang präzise an:

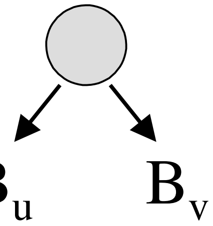
1. Jeder korrekt geklammerte Ausdruck fängt mit "(" an.

2. Es gibt zu dieser "(" genau eine zugehörige ")", nämlich die erste "Klammer zu", bei der die Anzahl der "Klammer auf" und "Klammer zu" gleich sind (von links nach rechts gezählt).

3. Also hat jeder Klammerausdruck die Form (u)v, wobei sowohl u als auch v korrekt geklammerte Ausdrücke sind. u und v sind eindeutig bestimmt.

4. Ordne () den einknotigen Baum zu: 

5. Ordne dann rekursiv dem Ausdruck (u) v folgenden Baum zu:



wobei B_u der zu u und B_v der zu v gehörige Baum ist.

Umgekehrt gewinnt man aus diesem Baum den Ausdruck (u) v.

Auf diese Weise lässt sich jedem korrekt geklammerten Ausdruck umkehrbar eindeutig ein binärer Baum zuordnen.

Folglich ist auch die Anzahl der korrekten Klammerungen aus n Klammerpaaren gleich C_n.

Damit ist die Behauptung bewiesen.

Wieviele korrekt geklammerte Ausdrücke gibt es?

Man muss n "Klammer auf" auf 2n Positionen verteilen.
Hiervon gibt es genau $\binom{2n}{n}$ Möglichkeiten.

Von dieser Anzahl muss man die abziehen, die zu keinen korrekten Klammerungen führen. Diese besitzen eine erste Position, bis zu der mehr "Klammer zu" als "Klammer auf" stehen; sie haben also die Form:

$$x\ )\ y$$

wobei x korrekt geklammert ist und y genau eine "Klammer auf" mehr besitzt als "Klammer zu".

Ersetze nun in y jede "(" durch ")" und umgekehrt. So möge die Klammerfolge y' entstehen. Für $x\ )\ y'$ gilt dann:

x möge k Klammerpaare besitzen ($0 \leq k \leq n$).
Dann besitzt y n-k "Klammer auf" und n-k-1 "Klammer zu".
Dann besitzt y' n-k-1 "Klammer auf" und n-k "Klammer zu".
Also besitzt x)y' n-1 "Klammer auf" und n+1 "Klammer zu".

Weiterhin gilt: Geht man von zwei verschiedenen unkorrekten Klammerungen aus, so erhält man auch zwei verschiedene Ausdrücke der Form x)y'. Wäre nämlich x)y' = x_1)y_1' , dann muss x=x_1 sein (da x und x_1 beide korrekt geklammert sind und ")" an der ersten Position steht, an der die Anzahl der "Klammer zu" die Anzahl der "Klammer auf" übersteigt) und ebenso muss dann y' = y_1' sein (da die Längen der beiden Ausdrücke gleich lang sind, nämlich 2n) und folglich gilt auch y = y_1.

Wir haben also gezeigt: Jeder unkorrekten Klammerung von n Klammerpaaren lässt sich eindeutig eine Folge von n-1 "Klammer auf" und n+1 "Klammer zu" zuordnen.

Die Umkehrung gilt aber auch.

Wenn eine Folge aus n-1 "Klammer auf" und n+1 "Klammer zu" gegeben ist, so muss es genau eine erste Stelle geben, an der die Zahl der "Klammer zu" die Zahl der "Klammer auf" erstmals übersteigt. Die Folge hat also die Form x)y' , wobei in x überall die Anzahl der "Klammer auf" größer oder gleich der Anzahl der "Klammer zu" bis zu dieser Stelle ist. x ist also ein korrekt geklammerter Ausdruck. In y' gibt es dann eine "Klammer auf" weniger, als es "Klammer zu" gibt. Wandle nun y' in ein y um, indem jede "(" durch ")" ersetzt wird und umgekehrt. So erhält man eine Folge x)y, die gleich viele "Klammer auf" und "Klammer zu" besitzt und die nicht korrekt geklammert ist.

Diese Zuordnung ist offenbar ebenfalls eindeutig. Also gilt:

Daher gilt:

Jeder unkorrekten Klammerung von n Klammerpaaren lässt sich umkehrbar eindeutig eine Folge von n-1 "Klammer auf" und n+1 "Klammer zu" zuordnen. Und:

Die Anzahl der unkorrekten Klammerungen mit n Klammerpaaren ist gleich der Anzahl der Folgen von n-1 "Klammer auf" und n+1 "Klammer zu".

Deren Anzahl ist aber $\binom{2n}{n-1}$

Wir haben also gezeigt:

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

. Damit ist der Satz bewiesen.

Folgerung:

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

Dieser Formel kann man zum einen das exponentielle Wachstum entnehmen, das in der Näherungsformel ausgedrückt wird. Zum anderen lassen sich hiermit die Catalanschen Zahlen, ausgehend von $C_1 = 1$, leicht iterativ berechnen.

Hinweis: Aus dem Satz lässt sich schließen:

Der Binomialkoeffizient $\binom{2n}{n}$ ist stets durch n+1 teilbar.

Unter welchen Bedingungen auch durch n+2?

Man erhält oft solche "nebensächlichen" Resultate.