Разложение числа Каталана на сегменты Чебышева

Геннадий Еремин

argenns@gmail.com

Март 24, 2016

Аннотация. Предлагается «нарезка» разложения числа Каталана на так называемые сегменты Чебышева. Множители сегментов не рассчитываются, а выбираются из ряда простых чисел, исходя из границ сегментов. Для факторизации n-го числа Каталана достаточно рассчитать кратность простых чисел $p < \sqrt{2n}$. Рассматривается легкое число Каталана, разложение которого представляет собой миниатюрное ядро – область кратных множителей соответствующего «тяжеловеса».

Ключевые слова: факторизация, разложение целых чисел, числа Каталана, сегменты Чебышева.

1. Введение

Задача разложения натуральных чисел имеет большую вычислительную сложность. Часто факторизация специальных чисел таких, как числа Ферма, Мерсенна, Каталана, Моцкина и др., используется для «обкатки» методов и алгоритмов разложения произвольных целых чисел. В настоящей работе рассматриваются числа Каталана, которые возникают в многочисленных комбинаторных приложениях (см. [Stan15]).

Формула Кэли. Специальные числа характеризуются наличием взаимосвязей между элементами соответствующих последовательностей, что существенно упрощает разложение таких чисел. Взаимосвязи реализуются рекуррентными и аналитическими зависимостями. Общеизвестна формула n-го числа Каталана

$$c_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!}, \quad n \geq 0. \tag{1}$$

В формуле (1) можно уменьшить количество операндов, отделим в числителе четные множители от нечетных:

$$(2n)! = 2\cdot4\cdot6\cdots2n \times 1\cdot3\cdot5\cdots(2n-1) = 2^n \times n! \times (2n-1)!!$$

После сокращений получим зависимость

$$c_n = 2^n \times (2n-1)!!\,/\,(n+1)! \tag{2}$$

Формулу (2) можно найти на сайте [Wei16]. Идентичное выражение в 1859 году опубликовал Артур Кэли, подсчитывая количество бинарных деревьев (см. исторический обзор [Pak14)]), поэтому назовем зависимость (2) *формулой Кэли*. Сочетание в формуле Кэли обычного факториала и двойного нечетного факториала упрощает отбор простых множителей.

В формуле (2) можно дополнительно уменьшить число операндов. Выделим и во втором факториале четные множители:

$$(n+1)! = 2\cdot 4\cdot 6\cdots 2\left\lceil\frac{n}{2}\right\rceil \times 1\cdot 3\cdot 5\cdots(2\left\lfloor\frac{n}{2}\right\rfloor+1) = 2^{\lceil n/2\rceil}\times\left\lceil\frac{n}{2}\right\rceil! \times (2\left\lfloor\frac{n}{2}\right\rfloor+1)!!$$

Очевидно, $2\lceil ½\, n\rceil$ и $2\lfloor ½\, n\rfloor+1$ – соответственно максимально возможные четное и нечетное числа, не превышающие $n+1$. Поскольку $\lceil ½\, n\rceil + \lfloor ½\, n\rfloor = n$, в конечном итоге получаем еще одну формулу

$$c_n = 2^{\lfloor n/2\rfloor}\times(2\lfloor n/2\rfloor+3)\cdot(2\lfloor n/2\rfloor+5)\cdots(2n\text{–}1)/\lceil n/2\rceil! \qquad (3)$$

Зависимость (3), несмотря на громоздкий внешний вид, практична для разложения чисел Каталана. Например, для подсчета числа двоек в разложении n-го числа Каталана достаточно разложить факториал $\lceil n/2\rceil!$

В разложении числа Каталана с индексом n много мелких множителей, и все меньше $2n$. Обозначим $\mathbb{P}$ – ряд простых чисел. *Факторной базой* n-го числа Каталана назовем *простой интервал* $B_n = {}_{\mathrm{p}}(1, 2n) \subset \mathbb{P}$, т.е. открытое множество простых чисел в указанных границах, но без граничных значений. Вне базы B_n нет простых множителей n-го числа Каталана. Мультимножество простых факторов разложения c_n обозначим $F_n = \{x_i\}$. Очевидно, в общем случае F_n не является подмножеством B_n ввиду кратных множителей.

Сегмент Чебышева. Состав простых чисел в разложении c_n во многом определяется биномиальным коэффициентом $\binom{2n}{n}$. В середине XIX века Чебышев обратил внимание на интервал простых чисел ${}_{\mathrm{p}}(n, 2n) \subset \mathbb{P}$, который полностью входит в разложение центрального биномиального коэффициента [Pom15]. Аналогичным образом все простые числа из открытого интервала ${}_{\mathrm{p}}(n+1, 2n) \subset B_n$ включены в разложение c_n. Простые множители интервала, назовем его *сегментом Чебышева*, не могут повториться в F_n. Наличие сегмента упрощает факторизацию чисел Каталана, поскольку вдвое сужается область отбора остальных простых множителей. Вышесказанное оформим в виде очевидной теоремы.

Теорема 1.1. *Разложение n-го числа Каталана полностью включает сегмент Чебышева ${}_{\mathrm{p}}(n+1, 2n) \subset \mathbb{P}$. Простые числа из сегмента не повторяются в разложении.*

В статье рассматривается семейство непересекающихся сегментов Чебышева в разложении числа Каталана. Количество сегментов фиксировано, интервал ${}_{\mathrm{p}}(n+1, 2n)$ наиболее протяженный, назовем его *главным сегментом*. Мощность других сегментов уменьшается по мере продвижения вниз по факторной базе.

Пространства между сегментами Чебышева – это простые замкнутые множества, своего рода, *запретные зоны*, свободные от элементов F_n. Запретные зоны также сокращаются при спуске по факторной базе. Самой протяженной зоной запрета условно можно считать бесконечное множество ${}_{\mathrm{p}}[2\mathrm{n}, \infty)$. Границы сегментов Чебышева и запретных зон определяются, исходя из индекса числа Каталана. Семейство сегментов составляет *шлейф числа Каталана*.

Сегменты Чебышева содержат подавляющую часть элементов F_n, в то время как область кратных множителей или *ядро числа Каталана* сжимается до размеров интервала ${}_{\mathrm{p}}(1, \sqrt{2n})$. Таким образом, практически вся факторная база может быть «нарезана» на сегменты Чебышева, которые переносятся фрагментами из ряда простых чисел, исходя из границ сегментов.

Пример 1.1. Для факторизации числа Каталана с индексом 1000 достаточно обсчитать лишь 14 простых чисел в диапазоне от 2 до 43 включительно. Множители, превышающие $\sqrt{2000} \approx 44{,}7$ (от 47 до 1999), выбираются из сегментов. При этом из главного сегмента ${}_{\mathrm{p}}(1001, 2000)$ в мультимножество F_{1000} переносятся 135 простых чисел от 1009 до 1999 включительно; из следующего по протяженности сегмента ${}_{\mathrm{p}}(1001/2, 2000/3)$ выбираются 26 множителей от 503 до 661 включительно и т.д. В зоне запрета ${}_{\mathrm{p}}[2000/3, 1001]$ (47 простых чисел) нет множителей разложения, т.е. $F_{1000} \cap {}_{\mathrm{p}}[2000/3, 1001] = \emptyset$. Всего в разложении 214 множителей, включая кратные. При этом из сегментов выбираются 197 простых чисел шлейфа, а ядро 1000-го числа Каталана включает лишь 17 расчетных множителей.

Сделаем небольшое отступление. Границы сегментов Чебышева и запретных зон, как правило, чисто условные, поскольку часто не принадлежат ни факторной базе, ни ряду простых чисел; в большинстве случаев это вещественные числа. Например, в сегментах верхние открытые границы и, соответственно, нижние закрытые границы запретных зон, будучи целыми, всегда четные числа.

В последнем 4-ом разделе рассматривается *легкое число Каталана*, разложение которого представляет собой миниатюрное ядро – область кратных множителей соответствующего «тяжеловеса».

2. Факторная база числа Каталана

В мультимножестве факторов F_n в общем случае есть и кратные, и неповторяющиеся простые множители из факторной базы B_n. Все простые числа из сегмента ${}_{\mathrm{p}}(n+1, 2n) \subset B_n$ не повторяются в F_n. Очевидно, числовое пространство факторной базы можно условно разделить на две непересекающиеся части: *верхняя область* B_n' содержит простые числа, которые чисто теоретически не могут повториться в F_n, и *нижняя область* B_n'', из которой выбираются остальные простые множители, включая и кратные. Границей между областями или *порогом кратности* g_n могла быть нижняя грань главного сегмента $n+1$, если бы этот сегмент был единственным. Ниже мы покажем, что это не так, и в общем случае $g_n \ll n$.

Порог кратности можно обнаружить и в разложении произвольного натурального числа. Для факторизации натурального m достаточно перебрать делители от 1 до $\sqrt{m}$. Именно в этом диапазоне простой делитель теоретически может повториться, но кратных делителей, превышающих $\sqrt{m}$, не существует.

Запретная зона. Рассмотрим в факторной базе B_n простые числа, которые непосредственно снизу примыкают к сегменту Чебышева $_{\mathrm{p}}(n+1, 2n)$. Распишем в формуле Кэли (2) оба факториала

$$c_n = 2^n \times \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2n-3)(2n-1)}{1 \cdot 2 \cdot 3 \cdot 4 \cdots n(n+1)}. \tag{4}$$

В знаменателе последний множитель $n+1$ может оказаться простым в случае четного n. Но такой множитель есть в числителе и тоже в единственном экземпляре, поэтому в разложении вероятностный простой множитель $n+1$ невозможен, т.е. $n+1 \notin F_n$. Простой нечетный множитель n также в одном экземпляре в обоих факториалах, следовательно, и $n \notin F_n$.

Ситуация повторится, если мы и дальше будем спускаться по натуральному ряду пока не доберемся до величины $\frac{2}{3}n$. Если это число целое, то четное, составное и также не может быть в разложении. В результате определилась первая (и, возможно, не последняя) *запретная зона* в верхней области факторной базы, ни один простой множитель из запретной зоны не может попасть в разложение n-го числа Каталана. Дополнительно мы снизили порог кратности, т.е. $g_n < \frac{2}{3}n$.

Проверим для профилактики следующее по нисходящей число $\frac{2}{3}n-1$, которое будучи целым и, следовательно, нечетным, может оказаться еще и простым. Такой простой множитель в разложении знаменателя по-прежнему в одном экземпляре, но в числителе мы получим второй экземпляр из составного нечетного числа $2n-3 = 3 \cdot (\frac{2}{3}n-1)$. В результате, простое число $\frac{2}{3}n-1$ обязательно окажется в разложении, причем в единственном экземпляре, и этот множитель принадлежит уже другому сегменту Чебышева. Вышеизложенное можем считать доказательством следующей теоремы.

Теорема 2.1. *В разложении n-го числа Каталана нет простых множителей из отрезка* $_{\mathrm{p}}[\frac{2}{3}n, n+1] \subset \mathbb{P}$, т.е. $F_n \cap {}_{\mathrm{p}}[\frac{2}{3}n, n+1] = \emptyset$

В теореме 2.1 для произвольного индекса n нижняя граница отрезка либо дробное число, либо натуральное четное число, поэтому логично закрыть обе границы отрезка. Главный сегмент и запретная зона в три раза сужают числовое пространство для отбора остальных множителей разложения.

Порог кратности. Спускаясь по факторной базе, мы, отталкиваясь от верхнего сегмента и пройдя запретную зону, оказались в следующем сегменте Чебышева, ниже которого, видимо, располагается еще одна запретная зона. Очевидно, так мы нескоро доберемся до порога кратности. Следующий пример поможет ускорить этот процесс.

Пример 2.1. Выберем простое число 101 и с помощью формулы Кэли подберем минимальный индекс числа Каталана, в разложении которого множитель 101 станет кратным. Если у нас получится, то это будет означать, что в факторной базе найденного числа Каталана порог кратности равен 101, либо очень близко. Такой эксперимент поможет в дальнейшем определиться с порогом кратности для произвольного числа Каталана. Распишем некоторые составные числа числителя и знаменателя, в которых встречается множитель 101:

числитель $1, 3, 5, 7, \ldots, 101, \ldots, 101{\times}3, \ldots, 101{\times}5, \ldots, 101{\times}101, \ldots$
знаменатель $1, 2, 3, 4, \ldots, 101, \ldots, 101{\times}2, \ldots, 101{\times}3, \ldots, 101{\times}101, \ldots$

При значениях индекса $n \le 50$ в обоих факториалах нет множителя 101, и только при $n = 51$ в числителе появится множитель 101. Отметим, первые 51 чисел Каталана, включая c_0, некратны 101. В диапазоне $51 \le n < 100$ множитель 101 будет в единственном числе в мультимножестве F_n. Можно сказать, что обнаружена зона благоприятных индексов для множителя 101 или блок из 49 чисел Каталана кратных 101. При $n = 100$ в знаменателе появится симметричный множитель 101, в результате простое число 101 исчезнет из F_{100}. Такая ситуация не изменится в диапазоне $100 \le n < 152$, и это будет зона запретных индексов для множителя 101 или блок чисел Каталана некратных 101.

Увеличивая индекс n, мы обнаружим в диапазоне $152 \le n < 201$ еще одну благоприятную зону, а в диапазоне $201 \le n < 253$ очередную запретную зону. Дальнейшее чередование благоприятных и запретных зон становится привычным, в каждой благоприятной зоне индексов множитель 101 включается в разложение в единственном экземпляре, а в каждой запретной зоне множитель 101 удаляется из разложения. Но, когда после очередной зоны запрета мы доберемся в числителе до составного числа $2n-1 = 101{\times}101$, то сразу получим два множителя 101 в разложении числа Каталана с индексом $n = (101^2+1)/2 = 5101$. Таким образом, мы добрались до наименьшего индекса числа Каталана, в разложении которого множитель 101 встречается дважды.

В примере 2.1 получено наименьшее 5101-ое число Каталана, в разложении которого есть кратный простой множитель 101. Для следующего простого числа 103 получим уже бóльший индекс 5305. Поскольку в разложении c_{5101} нет кратных множителей $p > 101$, то можно принять в качестве порога кратности четное число $g_{5101} = 102$. Выше этой величины в разложении c_{5101} нет кратных простых множителей. Ниже порога 102 расположены кратные множители $2^8, 3^6, 5^2, \ldots, 101^2$, но есть и некратные множители 17, 19, 29, …, 97. На основании вышеизложенного можем сформулировать еще одну теорему.

Теорема 2.2. *В разложении n-го числа Каталана кратные простые множители меньше* $\sqrt{2n}$.

В теореме 2.2 подкоренное выражение для компактности увеличено на 1, но при этом исключается возможное равенство для кратных множителей. Рассмотренный пример позволяет сформулировать более общую теорему.

Теорема 2.3. *В разложении n-го числа Каталана простые множители, имеющие кратность k, меньше* $\sqrt[k]{2n}$.

Порог кратности n-го числа Каталана удобно принять равным $g_n = \sqrt{2n}$. Эта величина, будучи целой, всегда четна (при этом и n четно), поэтому $g_n \notin B_n$. Такой порог логично использовать в качестве границы областей факторной базы. В следующем 3-ем разделе рассматривается «нарезка» верхней области факторной базы сегментами Чебышева, т.е. формирование шлейфа числа Каталана. Этот метод успешно применяется при факторизации больших чисел (с индексами 10^6 и более).

3. Шлейф числа Каталана

Порог кратности $g_n = \sqrt{2n}$ разбивает факторную базу $B_n = {}_{\mathrm{p}}(1, 2n)$ на два непересекающихся подмножества: из нижней области $B''_n = {}_{\mathrm{p}}(1, g_n)$ отбираются все кратные множители разложения, в верхней $B'_n = {}_{\mathrm{p}}(g_n, 2n)$ содержатся элементы шлейфа n-го числа Каталана. Поскольку $g_n \notin B_n$, то $B_n = B'_n \cup B''_n$.

Сделаем еще одно отступление. Когда мы говорим, например, о четном числе $⅔n$ или о натуральном $(n+1)/2$, то имеются в виду вероятные (возможные) величины при определенных значениях индекса n, а именно: в первом случае n кратно 3, во втором n нечетное число.

Верхняя грань сегмента Чебышева. Шлейф числа Каталана представляет собой группу сегментов $S_i \subset B'_n$, разделенных зонами запрета. Главный сегмент очевидным образом проявляется в аналитических формулах, примыкает к верхней границе факторной базы, самый протяженный, определяется первым, и логично присвоить ему номер 1, т.е. $S_1 = {}_{\mathrm{p}}(n+1, 2n)$. К сегменту примыкает запретная зона ${}_{\mathrm{p}}[2n/3, n+1] \subset B'_n$, в которой нет элементов мультимножества F_n. Логично и удобно использовать открытые множества для обозначения сегментов (границы множества не включены в сегмент), а закрытые множества – для зон запрета (границы включаются в запретную зону); будем и далее следовать таким обозначениям.

Простое число $⅔n-1 \in F_n$ (в этом случае n кратно 3) принадлежит 2-му сегменту, а величина $⅔n$ является верхней открытой границе этого сегмента. Если записать верхнюю открытую грань первого сегмента в виде $2n/1$, то прослеживается закономерность: верхняя граница следующего 3-го сегмента, очевидно, равна $2n/5$, для 4-го сегмента – $2n/7$ и т.д. Проверим на примере.

Пример 3.1. Пусть $n=7500$, тогда верхняя граница 3-го сегмента должна быть равна $2×7500/5 = 3000$. Рассмотрим ближайшие простые числа 2999 и 3001 по обе стороны границы. Обратимся к зависимости (4).

В двойном факториале 14999!! (числитель дроби) множитель 2999 обнаруживается в трех экземплярах, а именно: 2999, 3×2999, 5×2999. В факториале знаменателя 7501! только два экземпляра – это 2999 и 2×2999, поэтому $2999 \in F_{7500}$. Но для следующего множителя 3001 составное число 5×3001 выходит за пределы факторной базы, поэтому и в числителе и в знаменателе по два экземпляра простого числа 3001, следовательно, $3001 \notin F_{7500}$.

Очевидно, аналогичным образом можно вычислить значение верхней границы для произвольного сегмента Чебышева. Обобщим полученные результаты в нижеследующей теореме (доказательство см. в разделе 3а).

Теорема 3.1. *В шлейфе числа Каталана с индексом* n *верхняя открытая граница* s*-го сегмента Чебышева равна* $2n/(2s-1)$, $s > 0$.

Нетрудно оценить количество сегментов в шлейфе числа Каталана, поскольку $2n/(2s_{max}-1) > g_n$, или $s_{max} < (1+g_n)/2$. Нижняя грань последнего приграничного сегмента также превышает g_n, поэтому неравенство можно упростить

$$s_{max} < g_n/2 = \sqrt{n/2}\ .$$

Можем считать, что доказана следующая теорема.

Теорема 3.2. *Количество сегментов Чебышева в шлейфе числа Каталана с индексом* n *меньше* $\sqrt{n/2}$.

Реальное число сегментов Чебышева часто существенно меньше величины, указанной в теореме 3.2. Для больших индексов по мере приближения к порогу кратности верхние грани соседних сегментов быстро сходятся, и многие приграничные сегменты (а также запретные зоны) практически «схлопываются», т.е. оказываются пустыми, без простых чисел.

Нижняя грань сегмента Чебышева. В сегменте $S_1 = {}_{\mathrm{p}}(n+1, 2n)$ открытая граница $n+1$ часто не принадлежит ни базе, ни ряду простых чисел $\mathbb{P}$, например, в случае нечетного n. К сегменту снизу примыкает запретная зона ${}_{\mathrm{p}}[⅔n, n+1] \subset B_n$ – закрытый отрезок, в котором нет множителей разложения c_n. В следующем сегменте S_2 верхняя граница равна $2n/(2s-1) = ⅔n$, максимально возможный элемент – простое число $⅔n-1$. Найдем нижнюю открытую границу 2-го сегмента Чебышева и, соответственно, минимальный элемент сегмента.

Обратимся снова к формуле Кэли (2). Число Каталана с индексом n кратно простому числу $m = ⅔n-1$, поскольку в разложении факториала $(n+1)!$ множитель m будет в единственном экземпляре, а в разложении двойного факториала $(2n-1)!!$ множитель m дополнительно содержится в составном числе $3m < 2n$ (четное число $2m$ не подходит, т.к. двойной факториал нечетный). Очевидно, следующее по нисходящей нечетное число $⅔n-3$, будучи простым, также принадлежит разложению c_n. Ситуация не измениться пока не достигнем величины $m' = ½(n+1)$, которая может оказаться простым числом. В этом случае в разложении факториала $(n+1)!$ появится дополнительный простой множитель в составном числе $2m' = n+1$, и этот множитель обнуляет кратность m' в разложении C_n.

Таким образом, в сегменте 2 открытая нижняя граница равна $\frac{1}{2}(n+1)$, и, следовательно, ближайшее простое число, превышающее эту границу, принадлежит мультимножеству F_n. По-видимому, для 3-его сегмента получим нижнюю открытую границу $\frac{1}{3}(n+1)$ и т.д. Сформулируем полученный результат в следующей теореме (доказательство см. в разделе 3а).

Теорема 3.3. *В шлейфе числа Каталана с индексом* n *нижняя открытая граница* s*-го сегмента Чебышева равна* $(n+1)/s,\ s < \frac{1}{2}g_n$.

Таким образом, сегмент n-го числа Каталана имеет следующий вид:

$$S_i = {}_{\mathrm{p}}\big((n+1)/i,\ 2n/(2i-1)\big),\ i < \sqrt{n/2}\,. \tag{5}$$

Границы сегментов по большей части дробные числа, но удобнее работать с целыми величинами. При округлении открытые границы интервалов желательно сохранить, поэтому в нижней грани сегмента можно отбрасывать дробную часть (округление снизу – «пол»), а для верхней грани берется ближайшее целое сверху (округление «потолок»). Приведем два примера.

Пример 3.2. Рассмотрим разложение миллионного числа Каталана (читатель в этой программе может за приемлемое время факторизовать произвольное число Каталана с индексом до 10^7). Натуральный вид числа $C_{1000000}$ не интересен, массив из 600 тысяч десятичных знаков мало что может сказать. Другое дело простые множители, даже если их очень много. Мультимножество $F_{1000000}$ включает 101543 элементов (с учетом кратности множителей). Все простые числа выбираются из факторной базы ${}_{\mathrm{p}}(1,\ 2\cdot10^6)$. Минимальный множитель 2 имеет кратность 7, максимальный равен 1999993.

Факторная база разделена порогом кратности $g_{1000000} = 1414.21$ на две непересекающиеся части. Нижняя область – «вотчина» кратных множителей, неповторяющиеся простые числа верхней области распределены по 707 сегментам. Первые сегменты объединяют подавляющую часть элементов разложения. Например, сегменты $S_1 = {}_{\mathrm{p}}(1000001,\ 2000000)$ и $S_2 = {}_{\mathrm{p}}(500000,\ 666667)$ охватывают 82% множителей (70435+12531 элементов).

Первое простое число 1423 выше порога кратности попадает в сегмент $S_{703} = {}_{\mathrm{p}}(1422,\ 1424)$. Четыре сегмента с номерами 704-707 пустые, т.е. не содержат простых чисел. Следующее простое число 1433 также попадает в $F_{1000000}$ и выбирается сегментом $S_{698} = {}_{\mathrm{p}}(1432,\ 1434)$. В результате обнаружена еще одна группа пустых сегментов с порядковыми номерами 699-702. Как видим, вблизи порога кратности в данном случае лишь пятый сегмент «приносит добычу» – содержит простой множитель.

Пример 3.3. Для числа Каталана с индексом 10^8 порог кратности равен $g_{100000000} = 14142$, поэтому множители шлейфа выбираются из открытого интервала ${}_{\mathrm{p}}(14142,\ 2\cdot10^8)$. Определим кратность простых чисел p=463219 и q=543061. Чтобы решить эту задачу достаточно сопоставить оба числа с границами ближайших сегментов.

Выберем подходящий сегмент для первого числа, воспользуемся формулой нижней границы: $(10^8+1)/463219 = 215.88$. Следовательно, число p размещается между нижними границами сегментов 215 и 216. Сегмент 215 не подходит, поскольку его нижняя граница $\lfloor(10^8+1)/215\rfloor = 465116$ больше p. Остается сравнить p с верхней границей сегмента 216, и эта величина равна $\lceil 2\cdot 10^8/(2\cdot 216-1)\rceil = 464038$. Нетрудно видеть, число 463219 попадает в сегмент 216 и, следовательно, есть в разложении.

Обратим внимание, формула нижней границы могла дать целое значение. Случай редкий, но вероятный, например, для индекса 99592084 получим (99592084+1)/463219 = 215. Это означает, что заданное простое число совпадает с нижней границей 215-го сегмента (и одновременно с верхней границей запретной зоны между сегментами 215 и 216). В этом случае число 463219 отбраковывается, поскольку не попадает в разложение.

Аналогичные расчеты показывают, что второе число q размещается между верхней границей 185-го сегмента и нижней границей 184-го сегмента, т.е. в одной из запретных зон факторной базы. Следовательно, число 543061 не попадает в разложение числа Каталана с индексом 10^8.

Процедура, описанная в примере 3.3, сводится к следующим двум последовательным операциям: а) вычисляется номер сегмента, нижняя граница которого максимально приближена снизу к простому числу p, и б) сравнение p с верхней границей найденного сегмента. Иначе говоря, простое число $p \in B_n''$ попадает в разложение n-го числа Каталана, если

$$p < \lceil 2n/(2k-1)\rceil,\ \text{где}\ k = \lceil (n+1)/p\rceil,\ kp \neq n+1.$$

Если p делит $n+1$, то такое простое число не принадлежит разложению C_n.

Решето Чебышева. Чтобы получить шлейф числа Каталана достаточно переместить содержимое сегментов Чебышева в пустое множество, начиная, например, с главного сегмента. Но есть более быстрый метод, который аналогичен алгоритму построения известного решета Эратосфена (когда из отрезка натурального ряда вычеркиваются последовательно составные числа, кратные простым числам 2, 3, 5 и т.д.). Очевидно, проще и быстрее очистить запретные зоны в верхней части факторной базы числа Каталана, сохраняя содержимое сегментов. В результате получим конструкцию, которую по аналогии можно назвать *решето Чебышева*.

3а. Доказательства теорем

Функция кратности. Введем несколько определений и обозначений, которые понадобятся в дальнейшем. В литературе [Pom13] можно встретить *функцию кратности* $v_p(m)$, которая равна кратности простого числа p в разложении натурального m. Например, $v_7(14)=1$, $v_7(98)=2$, $v_7(5!)=0$. Подробно о свойствах функции кратности можно посмотреть на главной странице ресурса [Ep15] в разделе «Функционал Куммера». Приведем несколько очевидных свойств функции для натуральных a, b, k:

$$v_p(ab) = v_p(a) + v_p(b),\ \ v_p(a^k) = kv_p(a),\ \ v_p(a/b) = v_p(a) - v_p(b).$$

Функция кратности используется для разложения специальных чисел. Например, в формуле Кэли (2) очевидным образом просматривается кратность нечетных простых множителей:

$$v_p(c_n) = v_p((2n–1)!!) – v_p((n+1)!), \quad p > 2. \tag{6}$$

Последний операнд в равенстве (6) можно вычислить с помощью известной формулы Лежандра [Pom13]:

$$v_p((n+1)!) = \sum_{j>0} \lfloor (n+1)/p^j \rfloor . \tag{7}$$

Нечетное округление «пол». Для положительного $a>1$ обозначим $\lfloor a \rfloor_{\text{odd}}$ – округление «пол» до ближайшего снизу нечетного натурального числа, т.е. при таком округлении отбрасываем дробную часть a с дополнительным уменьшением на 1, если результат окажется четным числом. Например, $\lfloor 23/7 \rfloor_{\text{odd}} = \lfloor 34/7 \rfloor_{\text{odd}} = 3$. Такую операцию назовем *нечетное округление «пол»*. Операция нечетного округления удобна для определения кратности простых множителей в двойных факториалах. Например, вычислить первый операнд в правой части равенства (6) можно по следующей формуле (см. в [Ер15]):

$$v_p((2n–1)!!) = ½ \sum_{j>0} (\lfloor (2n–1)/p^j \rfloor_{\text{odd}} + 1). \tag{8}$$

Доказательство теоремы 3.1. Необходимо показать, что в шлейфе n-го числа Каталана верхняя грань s-го сегмента равна $u = 2n/(2s–1)$. При $s=1$ мы получаем верхнюю открытую грань факторной базы $2n$, поскольку первый сегмент примыкает вплотную к открытой верхней границе B_n.

Будем полагать u – целое и, следовательно, четное число. Если это не так, и u – дробное, то округлим эту величину до ближайшего четного числа. Пусть значение u находится в натуральном ряде между двумя нечетными простыми числами-близнецами $p = u–1$ и $q = u+1$. Очевидно, $g_n < p < u < q$. Выбраны наиболее жесткие условия для простых чисел, и необходимо доказать, что из двух близнецов p и q только p делит c_n. Очевидно,

$$n = ½\, u(2s–1) = ½\,(p+1)(2s–1) = ½\,(q–1)(2s–1).$$

Поскольку $p^2 > g_n^2 = 2n$, то $p^2 \nmid (2n–1)!!$ и $p^2 \nmid (n+1)!$, т.е. оба факториала не кратны p^2 (свободны от квадратов). Очевидно, это справедливо и для $q = p+2$. Тогда суммы в формулах (7–8) содержат только по одному слагаемому:

$$v_p((n+1)!) = \lfloor (n+1)/p \rfloor, \; v_p((2n–1)!!) = ½\,(\lfloor (2n–1)/p \rfloor_{\text{odd}} + 1), \; p > \sqrt{2n} .$$

Для доказательства теоремы достаточно проверить два равенства

$$v_p(c_n) = 1, \; n = ½\,(p+1)(2s–1), \quad \text{и} \quad v_q(c_n) = 0, \; n = ½\,(q–1)(2s–1). \tag{9}$$

Рассмотрим сначала операнды с обычными факториалами.

$$v_p((n+1)!) = \lfloor (½(p+1)(2s–1) +1)/p \rfloor =$$

$$= \lfloor s + s/p – ½ + 1/2p \rfloor = \begin{cases} s-1, & \text{если } s < p/2. \\ s, & \text{если } s \geq p/2. \end{cases} \tag{10}$$

$v_q((n+1)!) = \lfloor(½(q-1)(2s-1)+1)/q\rfloor =$

$$= \lfloor s - {}^{s}\!/_{q} - ½ + {}^{3}\!/_{2q}\rfloor = \begin{cases} s-1, \text{ если } s \le q/2. \\ s-2, \text{ если } s > q/2. \end{cases} \qquad (11)$$

В выражениях мы отбросили величины ${}^{1}\!/_{2p}$ и ${}^{3}\!/_{2q}$, поскольку они практически не влияют на конечный результат (разве что в случае p=2s или q=2s). В выражениях (10–11) реально работают только верхние ветви, поэтому в факториале $(n+1)!$ кратность простых чисел p и q одинакова и равна s–1. Вычислим операнды с двойными нечетными факториалами.

$v_p((2n-1)!!) = ½(\lfloor((p+1)(2s-1)-1)/p\rfloor_{\text{odd}}+1) =$
$= ½(\lfloor 2s-1+{}^{2}\!/_{p}(s-1)\rfloor_{\text{odd}}+1) = ½((2s-1)+1) = s.$

$v_q((2n-1)!!) = ½(\lfloor((q-1)(2s-1)-1)/q\rfloor_{\text{odd}}+1) =$
$= ½(\lfloor 2s-1-2s/q\rfloor_{\text{odd}}+1) = ½((2s-3)+1) = s-1.$

Последние расчеты подтверждают равенства (9), что и доказывает теорему.

Доказательство теоремы 3.3. В шлейфе числа Каталана открытая нижняя граница сегмента в отличие от верхней грани вполне может оказаться нечетным и, соответственно, простым числом. В связи с этим, покажем сначала, что произвольное простое число $p = {}^{1}\!/_{s}(n+1) > g_n$ (натуральное $s < ½g_n$) не делит C_n. Или на основании формулы (6) следует доказать равенство

$$v_p((2n-1)!!) = v_p((n+1)!),\ n = sp-1. \qquad (12)$$

Вычислим факториалы (7) и (8), которые по-прежнему свободны от квадратов p.

$v_p((n+1)!) = \lfloor(sp)/p\rfloor = s,$
$v_p((2n-1)!!) = ½(\lfloor(2sp-3)/p\rfloor_{\text{odd}}+1) = ½(\lfloor 2s-3/p\rfloor_{\text{odd}}+1) = ½(2s-1+1) = s.$

Таким образом, n-ое число Каталана некратно натуральному числу $(n+1)/s$, т.е. $(n+1)/s \nmid C_n$. Далее проверим минимально возможный элемент сегмента с порядковым номером s.

Пусть в шлейфе n-го числа Каталана нижняя грань s-го сегмента $w=(n+1)/s$ представляет собой четное натуральное число, и пусть существует простое $q = w+1 < 2n/(2s-1)$. Покажем, что C_n кратно q. Очевидно, $n = s(q-1)-1$, $q > w > g_n > 2s$ или $q \ge 2s+3$. Вычислим снова факториалы (7) и (8).

$v_p((n+1)!) = \lfloor s(q-1)/q\rfloor = \lfloor s - {}^{s}\!/_{q}\rfloor = s-1.$

$v_p((2n-1)!!) = ½(\lfloor(2(s(q-1)-1)-1)/q\rfloor_{\text{odd}}+1) =$
$= ½(\lfloor 2s-(2s+3)/q\rfloor_{\text{odd}}+1) = ½(2s-1+1) = s.$

Таким образом, в сегменте Чебышева с номером s минимально возможное простое число $(n+1)/s+1$ гарантированно делит C_n. Теорема доказана.

4. Ядро числа Каталана

Множители разложения числа Каталана, превышающие порог кратности, не рассчитываются, а отбираются из фиксированных простых интервалов. Границы интервалов определяются индексом числа Каталана. Таким образом, подавляющая часть множителей заранее известна, и сегменты Чебышева можно не хранить. Шлейф – это балласт, его можно сбросить, а при необходимости легко восстановить. Ниже в таблице подобрана информация для чисел Каталана с индексами 3–20.

Индекс	Порог крат-ности	Легкое число Каталана	Разложение чисел Каталана		Число Каталана «тяжеловес»
			ядро	шлейф	
3	2,45	1		5	5
4	2,83	2	2	7	14
5	3,16	6	2, 3	7	42
6	3,46	12	2, 2, 3	11	132
7	3,74	3	3	11, 13	429
8	4,00	2	2	5, 11, 13	1430
9	4,24	2	2	11, 13, 17	4862
10	4,47	4	2, 2	13, 17, 19	16796
11	4,69	2	2	7, 13, 17, 19	58786
12	4,90	4	2, 2	7, 17, 19, 23	208012
13	5,10	100	2, 2, 5, 5	17, 19, 23	742900
14	5,29	360	2, 2, 2, 3, 3, 5	17, 19, 23	2674440
15	5,48	45	3, 3, 5	17,19, 23, 29	9694845
16	5,66	90	2, 3, 3, 5	19, 23, 29, 31	35357670
17	5,83	30	2, 3, 5	11, 19, 23, 29, 31	129644790
18	6,00	300	2, 2, 3, 5, 5	7, 11, 23, 29, 31	477638700
19	6,16	30	2, 3, 5	7, 11, 23, 29, 31, 37	1767263190

20	6,32	60	2, 2, 3, 5	7, 11, 13, 23, 29, 31, 37	6564120420

Натурализованное легкое число Каталана существенно короче своего прародителя «тяжеловеса» по двум причинам: во-первых, мощность ядра гораздо меньше мощности шлейфа и, во-вторых, множители шлейфа по величине неизмеримо больше множителей ядра. Разница особенно ощутима на больших индексах, например, 170-ое легкое число Каталана равно 4080, в то время как в соответствующем прародителе 99 разрядов:

c_{170} = 566 348 408 726 522 751 148 449 775 858 720 556 691 622 190 005 859 389 137 334 807 741 187 245 136 563 759 042 704 144 561 379 440.

Конечно, и легкие числа на больших индексах выглядят внушительно. Легкое число Каталана с индексом 50 000 имеет длину 163 знака, вот это число:

1 029 142 440 334 210 758 480 708 051 574 203 810 960 548 889 204 183 685 792 756 307 454 455 534 384 071 475 346 148 063 228 602 598 951 202 112 317 923 378 920 856 767 137 362 573 866 245 850 058 506 779 182 608 960.

Но «тяжеловес» с индексом 50 000 показать проблематично, поскольку «весит» 30 000 десятичных разрядов и больше в 10^{180} раз.

Каждому числу Каталана соответствует определенное легкое число, но обратное неверно, среди легких чисел есть «двойники». С ростом индекса доля ядра в разложении числа Каталана стремительно снижается. Например, для индекса 10^4 из 1563 множителей разложения в ядре содержится 51 простое число, т.е. 3%. В то время как для числа Каталана с индексом 10^7 доля ядра меньше 0.1% (из 867821 множителей в ядро попадает 738). Дополнительно следует отметить, что ядро состоит из младших множителей.

Небольшие размеры легких чисел дают возможность осваивать «глубины» последовательности Каталана. Причем чаще всего интересно не натурализованное легкое число, а непосредственно ядро. Дополним таблицу, выпишем первые 100 легких чисел Каталана (первые четыре числа приравнены к 1):

00-09: 1, 1, 1, 1, 2, 6, 12, 3, 2, 2;

10-19: 4, 2, 4, 100, 360, 45, 90, 30, 300, 30;

20-29: 60, 60, 120, 450, 36, 1764, 392, 28, 280, 56;

30-39: 112, 7, 294, 1470, 84, 14, 28, 28, 1400, 490;

40-49: 980, 3780, 7560, 18900, 2520, 2520, 35280, 4410, 900, 36;

50-59: 216, 108, 216, 840, 336, 12, 24, 24, 240, 72;

60-69: 1008, 121968, 11616, 45375, 18150, 1650, 3300, 11550, 1039500, 29700;

70-79: 59400, 4950, 108900, 544500, 2134440, 1067220, 27720, 83160, 831600, 20790;

80-89: 1540, 10780, 21560, 84700, 33880, 5725720, 34354320, 780780, 273273000, 18218200;

90-99: 400400, 200200, 400400, 2002000, 8808800, 34684650, 69369300, 1415700, 5577000, 111540.

Программный сервис. В заключение отметим программный комплекс для проведения тестовых расчетов в оперативном режиме, сервис подробно описан в [Ер15] на странице «Сервис on-line». Читателю уже предлагались отдельные программы, использовались эпизодические ссылки на сопутствующие расчеты по материалам статьи. Перечислим основные программы с кратким описанием выполняемых функций.

Небольшая программа позволяет получить список простых чисел в некотором диапазоне, начало диапазона не превышает 10^{10}. Факторизовать числа Каталана можно в нескольких вариантах. В первом варианте выводится натурализованное число Каталана (максимальный индекс 10^4), во втором варианте опускается вывод натурального числа (допустимый индекс 10^7). Есть еще третий вариант, в котором факторизуется легкое число Каталана (допустимый индекс 10^8). Отдельная программа позволяет получить ядро числа Каталана (величина индекса ограничена лишь разрядной сеткой компьютера).

Список литературы

[Pak14] Igor Pak. History of Catalan numbers (August 2014). http://www.math.ucla.edu/~pak/papers/cathist4.pdf

[Pom13] Carl Pomerance. *On numbers related to Catalan numbers*. Mathematics Department, Dartmouth College, Hanover (September 2013). https://math.dartmouth.edu/~carlp/catalan

[Pom15] Carl Pomerance. *Divisors of the middle binomial coefficient*. Amer. Math. Monthly **122** (2015), 636–644. https://math.dartmouth.edu/~carlp/amm2015.pdf

[Stan15] Stanley, R. P. (2015): Catalan Numbers. Cambridge (2015). http://www.cambridge.org/ro/academic/subjects/mathematics/discrete-mathematics-information-theory-and-coding/catalan-numbers?format=HB

[Wei16] Weisstein, Eric W. Catalan Numbers (Wolfram MarhWorld). http://mathworld.wolfram.com/CatalanNumber.html

[Ер15] Г. Еремин. Факторизация чисел Каталана (2015). http://eremin.magekit.com/catalan/