Коточигов Александр Михайлович

УДК 21.7.12

АЛГОРИТМ НУМЕРАЦИИ ЭЛЕМЕНТОВ ВЫБОРКИ

Аннотация

Заметка содержит описание алгоритма, нумерации элементов выборки m элементов из n. Благодаря использованию биномиальных разложений натуральных чисел, алгоритм становится универсальным и удобным для приложений.

Ключевые слова: выборка m элементов из n, биномиальные коэффициенты, биномиальное разложение.

Музыкант: « Вам понравился камерный концерт?»
Математик: «Тривиальный случай, «к» равно трем.»

.

ВВЕДЕНИЕ

Цель заметки – описать алгоритм, позволяющий упорядочить (занумеровать) элементы множества, представляющего собой выборку из n элементов по m. Интерес к этой задаче возник при попытке анализировать структуру многогранника, являющегося выпуклой оболочкой заданного множества точек.

Поясним задачу и рассмотрим вначале случай, когда размерность равна трем. Алгоритм описания граней прост. Надо перебрать все наборы из трех точек, образующих плоскость, и выбрать те плоскости, у которых все заданные точки лежат в одном полупространстве. Следующий этап – описание ребер многогранника – более сложен. Фиксируем одну вершину. Далее будем перебирать все остальные вершины и выяснять, лежит ли средняя точка отрезка, соединяющего вершины, внутри многоугольника. Здесь проблем с упорядочением выборки еще не возникает. Они появляются с повышением размерности. Предположим, вам надо по поверхности многогранника провести гусеницу из точки A в точку B. Размерность многогранника 20, размерность гусеницы 10. Части поверхности, имеющие размерность меньше 10, гусеница может только пересекать. Как проложить оптимальный и, главное, безопасный путь? Задача может иметь и более серьезную интерпретацию. Допустим, вы должны контролировать некий процесс (перевести его из состояния A в состояние B), зависящий от 20 параметров. Контроль заключается в том, что значения параметров должны постоянно «находиться» на поверхности заданного многогранника. Условия контроля таковы, что в любой момент десять параметров должны оставаться «свободными», то есть найдется такое 10-мерное пространство, что при небольших движениях в нем точка не уходит с поверхности многогранника.

Для описания m-мерных элементов границы многогранника надо фиксировать вершину и перебирать все наборы из m вершин, не содержащие фиксированной вершины. Работа с такими объектами значительно упростится, если будет предъявлен достаточно простой алгоритм, нумерующий элементы выборки (изоморфизм множества элементов выборки и подмножества натуральных чисел $\{1, 2, ..., N\}$). Без такой нумерации вы не сможете, например, составить матрицу инцидентности, то есть таблицу, указывающую, какие вершины принадлежат заданному m-мерному элементу границы и какие m-мерные элементы границы содержат заданную вершину.

Конкретизируем нашу задачу. Допустим, имеется n элементов, из которых выбирается m. Множество выборок можно представить в следующем виде $\Delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$, где $\delta_j = 0$ или 1, $\delta_1 + \ldots + \delta_n = m$. Алгоритм нумерации таких объектов будет построен в два этапа. Сначала будет описано биноминальное разложение натуральных чисел, с которым естественно связаны наборы нулей и единиц несколько иной природы, чем Δ. На втором этапе будет описан изоморфизм этих множеств нулей и единиц, который позволит занумеровать элементы множества Δ.

БИНОМИНАЛЬНОЕ РАЗЛОЖЕНИЕ НАТУРАЛЬНЫХ ЧИСЕЛ

Напомним базовое для всех последующих конструкций определение числа сочетаний из n элементов по m, то есть количества способов выбрать m различных элементов из n имеющихся, причем порядок следования элементов в выборке безразличен. Для этого числа имеется стандартное обозначение и формула вычисления

$$C_n^m = \frac{n!}{m!(n-m)!}.$$

Здесь $m = 0, 1, \ldots, n$, при этом возникает стандартное обозначение $0! = 1$, позволяющее единообразно описывать все возможнвые ситуации. Эти числа появляются во многих математических моделях, но самая известная среди них – формула бинома Ньютона. Поэтому их называют еще биномиальными коэффициентами. Здесь будет приведен алгоритм разложения натурального числа в сумму биномиальных коэффициентов.

До того как выписать формулу разложения, приведем некоторые наводящие соображения. Начнем с известного, легко проверяемого тождества

$$C_n^m = C_{n-1}^m + C_{n-1}^{m-1}.$$

Последовательно применяя тождество к последнему слагаемому, легко получить:

$$C_n^m = C_{n-1}^m + C_{n-2}^{m-1} + C_{n-3}^{m-2} + \ldots + C_{n-m}^1 + 1,$$
$$(C_{n-m+1}^1 = C_{n-m}^1 + C_{n-m}^0 = C_{n-m}^1 + 1).$$

Это и есть разложение, которое мы будем называть биноминальным для числа $k = C_n^m$.

Оказывается, эта формула допускает значительное обобщение.

Предложение 1

Для любого натурального числа $k \in (1, C_n^m]$ существует единственный набор натуральных чисел

$$n_i,\ 1 \le i \le j \le m,\ n > n_1 > n_2 > \ldots > n_j > m - j$$

таких, что

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \ldots + C_{n_j}^{m-j+1} + 1.$$

Эта формула известна. Например, ее можно найти в книге [1: с. 22], где она приводится в качестве задачи. Здесь приведено подробное доказательство, по существу, алгоритм разложения, поскольку дальнейшие конструкции опираются на этот алгоритм.

Техническую основу доказательства составляют два совершенно очевидных утверждения.

Предложение 2

Полуоткрытые интервалы $(C_p^m, C_{p+1}^m]$, $p = m, \ldots, n-1$ образуют дизъюнктное покрытие интервала $(1, C_n^m]$.

Для доказательства достаточно убедиться, что $C_p^m < C_{p+1}^m$, что верно, так как

$$\frac{C_{p+1}^m}{C_p^m} = \frac{p+1}{p+1-m} \ge 1.$$

Пример.

$$n = 7,\ \ m = 4,\ \ C_7^4 = 35,$$
$$p = m, C_4^4 = 1,\ p = m+1, C_5^4 = 5,$$
$$p = m+2, C_6^4 = 15,\ p = m+3, C_7^4 = 35$$
$$(1,35] = (1,5] \cup (5,15] \cup (15,35].$$

Следствие 1

Если число $K \in (1, C_N^M]$, то существует число $N_1 \in [M, N-1]$, такое что $K \in (C_{N_1}^M, C_{N_1+1}^M]$.

Предложение 3

Если число $K \in (C_{N_1}^M, C_{N_1+1}^M]$, то число $K - C_{N_1}^M \in [1, C_{N_1}^{M-1}]$.

Это утверждение является простым следствие упомянутого выше тождества

$$C_{N_1+1}^M = C_{N_1}^M + C_{N_1}^{M-1},\ \ C_{N_1}^M < K \le C_{N_1+1}^M,$$
$$1 \le K - C_{N_1}^M \le C_{N_1+1}^M - C_{N_1}^M = C_{N_1}^{M-1}.$$

Оно очень важно, так как оно поддерживает самоподобие системы, в терминах которой происходит разложение.

Чтобы обосновать формулу предложения 1, опишем алгоритм разложения.

Предположим, что $K \in (1, C_N^M]$.

Шаг 1

Числу K соответствует единственное число N_1, такое что $K \in (C_{N_1}^M, C_{N_1+1}^M]$ (следствие 1).

Шаг 2

Получаем число $K_1 = K - C_{N_1}^M \in [1, C_{N_1}^{M-1}]$ (предложение 3).

Шаг 3

Если $K_1 = 1$, то разложение закончено $K = C_{N_1}^M + 1$.

Если $K_1 > 1$, то возвращаемся к шагу 1, полагая $K = K_1, N = N_1, M = M - 1$.

Алгоритм сводит доказательство предложения 1 к анализу его действия на число $k \in (1, C_n^m]$. После применения к числу k шага 1 получится число n_1 такое, что

$$k \in (C_{n_1}^m, C_{n_1+1}^m], \quad n > n_1 > m-1.$$

Шаг 2 дает число k_1, такое, что

$$k = C_{n_1}^m + k_1, \quad 1 \le k_1 \le C_{n_1}^{m-1}.$$

Шаг 3 остановит алгоритм, если $k_1 = 1$ (разложение завершено) или перейдет к шагу 1 и вычислит число n_2, такое что

$$k_1 \in (C_{n_2}^{m-1}, C_{n_2+1}^{m-1}], \quad n > n_1 > n_2 > m-2.$$

После второго применения шага 2 получится число k_2, такое что

$$k = C_{n_1}^m + C_{n_2}^{m-1} + k_2, \quad 1 \le k_2 \le C_{n_2}^{m-2}.$$

Продолжая процесс, на r-м проходе цикла получим числа k_r и n_r:

$$k_{r-1} \in (C_{n_r}^{m-r+1}, C_{n_r+1}^{m-r+1}], \quad n > n_1 > \ldots > n_r > m-r$$

$$k = C_{n_1}^m + \ldots + C_{n_r}^{m-r+1} + k_r, \quad 1 \le k_r \le C_{n_r}^{m-r}.$$

Если $k_r = 1$, то разложение окончено.

Остается проверить, что разложение не может продолжаться более m шагов.

Допустим, алгоритм отработал $m-1$ раз, тогда

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \ldots + C_{n_{m-1}}^2 + k_{m-1},$$
$$1 \le k_{m-1} \le C_{n_{m-1}}^1.$$

Если и на этот раз $k_{m-1} > 1$ (процесс разложения не завершился), то далее надо применять шаг 1 и искать число n_m, такое что

$$k_{m-1} \in (C_{n_m}^1, C_{n_m+1}^1].$$

Остается заметить, что $C_R^1 = R$ и, следовательно, полагая $n_m = k_{m-1} - 1$, получим

$$k_{m-1} \in (C_{n_m}^1, C_{n_m+1}^1] = (k_{m-1} - 1, k_{m-1}].$$

В итоге

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \ldots + C_{n_{m-1}}^2 + C_{n_m}^1 + 1,$$

то есть разложение завершено. Формула полностью доказана.

Доказанная формула позволяет взаимно однозначно сопоставить каждому числу $k \in (1, C_n^m]$ некоторый набор из 0 и 1.

Следствие 2

Каждому числу $k \in (1, C_n^m]$ взаимно однозначно соответствует набор $\{\gamma_1, \gamma_2, \ldots, \gamma_n\}$, $\gamma_j = 0$ или 1, $\gamma_1 + \ldots + \gamma_n \le m$, составленный на основе биномиального разложения

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \ldots + C_{n_j}^{m-j+1} + 1,$$

$$n > n_1 > n_2 > \ldots > n_j \ge m - j, \quad 1 \le j \le m$$

по следующему правилу: $\gamma_s = 1$, если $s = n$ при некотором i, иначе $\gamma_s = 0$. В частности, для любого k.

Обозначим множество всех таких наборов через Γ:

$$\Gamma = \{(\gamma_1, \gamma_2, \ldots, \gamma_n) : \gamma_j = 0 \vee 1, \ \gamma_n = 0,$$
$$\gamma_1 + \ldots + \gamma_n \le m\}.$$

Следствие 2 устанавливает правило нумерации элементов множества Γ.

Описание нумерации элементов выборки из *n* элементов по *m*

Наша цель занумеровать элементы множества Δ, которое было определенно во введении следующими соотношениями:

$$\Delta = \{\delta_1, \delta_2, \ldots, \delta_n\}, \ \delta_k = 0 \vee 1, \delta_1 + \ldots + \delta_n = m.$$

Напомним, что число элементов такого множества равно C_n^m. Чтобы сделать это, достаточно установить взаимно однозначное соответствие между множествами Γ и Δ.

Это можно выполнить, разбив каждое из множеств на дизъюнктные системы под-множеств, имеющих одинаковое число элементов.

Для множества Г имеется естественное разбиение:

$$\Gamma_{m-r} = \{(\gamma_1, \gamma_2, \dots, \gamma_n) : \gamma_n = 0, \sum_{i=1}^{n-1} \gamma_i = m - r\},$$
$$r = 0, 1, \dots, m-1.$$

Каждому элементу множества Γ_{m-r} можно сопоставить набор натуральных чисел

$$(n_1, \dots, n_r), \quad n > n_1 > \dots > n_r > m - r,$$
$$n_j = i \leftrightarrow \gamma_i = 1,$$
$$r = 0, 1, \dots, m-1.$$

Занумеруем элементы этого множества числами $k = C_{n_1}^m + C_{n_2}^{m-1} + \dots + C_{n_r}^{m-r+1} + 1$.

Из предложения 1 следует, что таким образом будет установлено взаимно однозначное соответствие между элементами множества Г и множеством чисел $\{1, 2, \dots, C_n^m\}$.

Разбиение множества Δ подстроим «под разбиение» множества Г:

$$\Delta_m = \{(\delta_1, \delta_2, \dots, \delta_n): \ \delta_n = 0\},$$
$$\Delta_{m-1} = \{(\delta_1, \delta_2, \dots, \delta_n): \ \delta_n = 1, \ \delta_1 = 0\},$$
$$\dots$$
$$\Delta_{m-r} = \{(\delta_1, \delta_2, \dots, \delta_n): \ \delta_n = 1, \ \delta_j = 1, \ 1 \le j < r, \delta_r = 0\},$$
$$r = 2, \dots, m-1.$$

Заметим, что если $(\delta_1, \delta_2, \dots, \delta_n) \in \Delta_{m-r}$, то $\delta_{r+1} + \delta_{r+2} + \dots + \delta_{n-1} = m - r$ при $r = 0, \dots, m-1$.

Следовательно, если $(\delta_1, \delta_2, \dots, \delta_n) \in \Delta_{m-r}$, то $\gamma_i = 0$, полагая при $1 \le i \le r$, $i = n$ и $\gamma_i = \delta_i$ при $r < i < n$, получим $(\gamma_1, \dots, \gamma_n) \in \Gamma_{m-r}$.

Табл. 1 иллюстрирует правило «перевода» элементов множества Δ_{m-r} в элементы множества Γ_{m-r} и обратно.

Приведенное рассуждение и таблица справедливы для всех r, больших нуля. Если $r = 0$, то положение только упрощается, так как в этом случае и $\delta_1 + \dots + \delta_{n-1} = m$, и $\gamma_1 + \dots + \gamma_{n-1} = m$ (см. табл. 2).

Это соответствие взаимно однозначно, Так как изменяемая часть одинакова для соответствующих элементов множества Δ_{m-r} и Γ_{m-r}, а неизменяемая часть элементов Δ_{m-r} не позволяет двум таким множествам иметь общие элементы, если они имеют разные индексы. Таким образом, установлено взаимно-однозначное соответствие между элементами множеств $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$ и $\Delta = \{\delta_1, \delta_2, \dots, \delta_n\}$.

Следствие 3 (правило нумерации элементов множества Δ).

Если $\delta \in \Delta_m$, то $\delta_n = 0$ и номер элемента

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \dots + C_{n_m}^1 + 1,$$

где $n_i, i = 1, \dots, m$ все те индексы, для которых $\delta_{n_i} = 1$.

Если $\delta \in \Delta_{m-1}$, то $\delta_n = 1, \ \delta_1 = 0$, и номер элемента

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \dots + C_{n_{m-1}}^2 + 1,$$

где $n_i, i = 1, \dots, m-1$ все те индексы, для которых $n_i < n$ и $\delta_{n_i} = 1$.

Если $\delta \in \Delta_{m-r}$, то $\delta_n = 1, \ \delta_i = 1, \ i = 1, \dots, r-1, \ \delta_r = 0$, и номер элемента

$$k = C_{n_1}^m + C_{n_2}^{m-1} + \dots + C_{n_{m-r}}^{m-r+1} + 1,$$

где индексы $n_i, i = 1, \dots, m-r$, такие что $r < n_i < n$ и $\delta_{n_i} = 1$.

В частности, если $\delta \in \Delta_m$ $(r = m-1)$, то $\delta_n = 1, \ \delta_i = 1, \ i = 1, \dots, m-2, \ \delta_{m-1} = 0$, и номер элемента

$$k = C_{n_1}^m + 1.$$

Табл. 1

	1	...	$r-1$	r	$r+1 \dots n-1$	n
$\delta \in \Delta_{m-r}$	1		1	0	совпадение	1
$\gamma \in \Gamma_{m-r}$	0		0	0	совпадение	0

Табл. 2

	$1 \dots n-1$	n
$\delta \in \Delta_m$	совпадение	0
$\gamma \in \Gamma_m$	совпадение	0

Так как $\delta_1 + \ldots + \delta_n = m$, то среди индексов s, $m-1 < s < n$ есть только один, для которого $\delta_s = 1$, $n_1 = s$.

Пример (реализации алгоритма).

Таблица 3 иллюстрирует формирование нумерации для выборки 4-х элементов из 7.

$$n = 7,\ \ m = 4,\ \ C_7^4 = 35.$$

Элементы выборки перечислены в столбце δ_j, номера элементов находятся в столбце k.

Биномиальное разложение числа k определяется набором чисел n_j. Элементы множества Γ строятся по набору n_j по простому правилу: единица на местах, номера которых попали в набор, на остальных позициях нули. Элементы выборки δ_j получены из соответствующих элементов γ_j добавлением некоторого количества единиц, выделенных в таблице жирным шрифтом. Правила добавления единиц сформулированы в следствии 3. Курсивом выделены позиции

Табл. 3

k	n_j	γ_j	δ_j
1			
2	4	(0,0,0,1,0,0,0)	(**1,1** ,0,1,0,0, **1**)
3	4,3	(0,0,1,1,0,0,0)	(**1** ,0,1,1,0,0, **1**)
4	4,3,2	(0,1,1,1,0,0,0)	(0,1,1,1,0,0, **1**)
5	4,3,2,1	(1,1,1,1,0,0,0)	*(1,1,1,1,0,0,0)*
6	5	(0,0,0,0,1,0,0)	(**1,1**, 0,0,1,0, **1**)
7	5,3	(0,0,1,0,1,0,0)	(**1** ,0,1,0,1,0, **1**)
8	5,3,2	(0,1,1,0,1,0,0)	(0,1,1,0,1,0, **1**)
9	5,3,2,1	(1,1,1,0,1,0,0)	*(1,1,1,0,1,0,0)*
10	5,4	(0,0,0,1,1,0,0)	(**1** ,0,0,1,1,0, **1**)
11	5,4,2	(0,1,0,1,1,0,0)	(0,1,0,1,1,0, **1**)
12	5,4,2,1	(1,1,0,1,1,0,0)	*(1,1,0,1,1,0,0)*
13	5,4,3	(0,0,1,1,1,0,0)	(0,0,1,1,1,0, **1**)
14	5,4,3,1	(1,0,1,1,1,0,0)	*(1,0,1,1,1,0,0)*
15	5,4,3,2	(0,1,1,1,1,0,0)	*(0,1,1,1,1,0,0)*
16	6	(0,0,0,0,0,1,0)	(**1,1**, 0,0,0,1, **1**)
17	6,3	(0,0,1,0,0,1,0)	(**1** ,0,1,0,0,1, **1**)
18	6,3,2	(0,1,1,0,0,1,0)	(0,1,1,0,0,1, **1**)
19	6,3,2,1	(1,1,1,0,0,1,0)	*(1,1,1,0,0,1,0)*
20	6,4	(0,0,0,1,0,1,0)	(**1** ,0,0,1,0,1, **1**)
21	6,4,2	(0,1,0,1,0,1,0)	(0,1,0,1,0,1, **1**)
22	6,4,2,1	(1,1,0,1,0,1,0)	*(1,1,0,1,0,1,0)*
23	6,4,3	(0,0,1,1,0,1,0)	(0,0,1,1,0,1, **1**)
24	6,4,3,1	(1,0,1,1,0,1,0)	*(1,0,1,1,0,1,0)*
25	6,4,3,2	(0,1,1,1,0,1,0)	*(0,1,1,1,0,1,0)*
26	6,5	(0,0,0,0,1,1,0)	(**1** ,0,0,0,1,1, **1**)
27	6,5,2	(0,1,0,0,1,1,0)	(0,1,0,0,1,1, **1**)
28	6,5,2,1	(1,1,0,0,1,1,0)	*(1,1,0,0,1,1,0)*
29	6,5,3	(0,0,1,0,1,1,0)	(0,0,1,0,1,1, **1**)
30	6,5,3,1	(1,0,1,0,1,1,0)	*(1,0,1,0,1,1,0)*
31	6,5,3,2	(0,1,1,0,1,1,0)	*(0,1,1,0,1,1,0)*
32	6,5,4	(0,0,0,1,1,1,0)	(0,0,0,1,1,1, **1**)
33	6,5,4,1	(1,0,0,1,1,1,0)	*(1,0,0,1,1,1,0)*
34	6,5,4,2	(0,1,0,1,1,1,0)	*(0,1,0,1,1,1,0)*
35	6,5,4,3	(0,0,1,1,1,1,0)	*(0,0,1,1,1,1,0)*

элементов δ_j, совпадающих с соответствующими элементами γ_j.

Отметим простую связь маркировки элементов δ_j и множеств Δ_{m-r} (разбиения множества Δ). Курсив отмечает элементы множества Δ_4, одна жирная единица – Δ_3, две жирных единицы – Δ_2, три жирных единицы – Δ_1.

Литература

1. *Ландо С.К.* Лекции о производящих функциях. М., 2004.

Abstract

The notes contain description of natural renumbering algorithm for units of sample of m elements from n. Algorithm based on presentation of natural numbers as sum of binomial coefficients.

Keywords: sample of m elements from n, binomial coefficients, binomial decomposition.

Коточигов Александр Михайлович,
доктор физико-математических наук, профессор СПбГЭТУ,
amkotochigov@gmail.com