

第一章 数和数的表示形式

陈诗峰 2026/9/25

一、从不同的集合中抽象出共性——数

数和数的表示形式是两个很容易混淆的概念。我们先不要急着讨论数字 1、2、3、4、5 是怎样写出来的，而先想一个更基本的问题：

人为什么会有“数”这个概念？

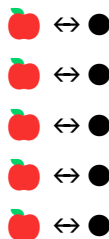
人具有一种认识数量的能力。我们能够把不同的对象一个一个地区分出来，也能够比较不同对象集合的多少。例如，这里有 5 个苹果：



这里有 5 块石头：



虽然苹果和石头完全不同，但是我们可以让每一个苹果恰好对应一块石头：



没有苹果剩下来，也没有石头剩下来。再拿来 5 个鸡蛋，也可以与这 5 块石头一一对应。于是我们发现了一件很有意思的事情：

5 个苹果、5 块石头、5 个鸡蛋，虽然是完全不同的对象集合，却有一个共同的数量特征。

我们把这个共同的数量称为：**5**

因此，“**5**”并不等于 5 个苹果，也不等于 5 块石头或 5 个鸡蛋。

它是我们从所有具有 5 个对象的集合中抽象出来的一个共同的数量。

二、传递信息的需要——数的表示方法

但是，我们还需要把这个数量记录下来、告诉别人。

于是，人类可以使用各种符号：**5、五、V、five、|||||.....**

这些写法看起来完全不同，却都可以用来表示同一个数量。所以我们必须区分两个概念：**数，是我们所认识的数量；数字和记数符号，是我们用来表示这个数的方法。**

这就像一个人的名字并不是这个人本身一样：“5”这个符号，并不是数 5 本身。

思考题

1. 如果世界上从来没有人发明符号 5，5 个苹果还会是 5 个苹果吗？
2. 如果你认为“是”，那么：数 5 和符号“5”，究竟哪一个更加基本？

三、语言学分析——区分“数”及其表示形式

我们每天都会说很多关于数的话。例如：

$$3 + 5 = 5 + 3$$

又例如：

偶数的个位是 0、2、4、6、8。

这两句话看起来都在谈“数”。但是，前者谈论的“数”本身的性质，而后者则是谈论 **10 进制**表示法下偶数的表示形式具有的特点，一旦更换数的表示形式，这句话将不再成立或者没有意义。

实验一：句子分组

请阅读下面 8 句话。

- A) $3 + 5 = 5 + 3$
- B) 两个奇数相加，结果是偶数。
- C) 任何一个数乘 1，仍然等于它自己。
- D) 8 比 5 大。

E) 偶数的个位是0、2、4、6、8。

F) 25 的最后一个数字是 5。

G) 100 有三个数字。

H) 一个数如果能被 10 整除，它的个位是 0。

请把你认为“性质相似”的句子分成两组。并且解释你为什么这样划分？

第一组：

第二组：

以上句子可以分为两类，一类谈论“数”本身的性质，一类是跟数的表示形式相关的性质，可以把数的表达方法换为其它方法，例如 2 进制，3 进制，如果发现跟表示方法无关，就是“数”本身的性质，否则就是依赖特定表示形式的性质。

例如 $5 + 3 = 3 + 5$ ，是在表达加法的交换律，无论 3，5 用什么形式表达，等式依然成立，它是谈论“数”的性质的句子。而“偶数的个位是 0、2、4、6、8。”则是 10 进制表示法特有的，因为 10 进制中的 4 如果写成 3 进制（逢 3 进 1）就是 11，个位是 1。

实验二：句子分组

在每句话后面写：

N—— 主要描述数本身 (Number)

R—— 主要描述数的表示 (Representation)

① $7 + 8 = 15$

② $7 + 8 = 8 + 7$

③ 12 能被 2 整除。

④ 12 的个位是 2。

⑤ 9 比 7 大。

⑥ 100 有三个数字。

⑦ 一个偶数加一个偶数仍然是偶数。

- ⑧ 25 的最后一个数字是 5。
- ⑨ 任何一个数加 0 仍然等于它自己。
- ⑩ 一个十进制整数如果个位是 0，就能被 10 整除。

思考题：

- 3. 辨析“如果一个整数各位数字之和能被 3 整除，那么这个数就能被 3 整除。”这句话是谈论是数本身还是数的表示形式。
- 4. 辨析“100 比 99 大。”和“100 比 99 多一个数字。”这两句话的区别。

四、最直观的表达法——一进制

我们现在已经知道一个数量可以写成：**5**，也可以写成：**五**、**V** 或者 **IIII**。那么：哪一个才是真正的 5？也许这个问题本身就问错了。因为：**5**、**五**、**V**、**IIII** 都只是表示方法。它们所指向的那个共同的数量，才是我们真正研究的：**数 5**。

1 进制是最自然的数的表示法，就是用一条竖线(I)表示数 1，两条竖线(II)表示数 2，依次类推，5 可以写成IIII，6 可以写成IIIII.....

思考题

- 5. 如果只允许使用 | 这一个符号，如何表示 0？
- 6. 如果只允许使用 | 这一个符号，能不能表示所有正整数？

1 进制的优点是直观，大数的表示形式是在小数的表示形式上加以扩展，所以适合动态计数，特别在没有电子屏幕的时代，不用来回擦写，例如我们在统计选票的时候，实质上用的就是 1 进制（但是做了一些优化，写“正”字）。另外，用 1 进制做加法、减法也特别直观和简单。

但是 1 进制的缺点也是明显的，首先，0 只能以空白来表示，这就很容易产生歧义，例如问“你有多少钱？”你在纸上留下空白，如果不加额外说明，就无法区分是你不愿意告诉我还是你一分钱也没有。其次，大数的表示过长，例如 1 百万，写 1 百万条竖线几乎不可能，而写 10 进制的 1000000 则轻而易举。再次，相邻的数表示方法过于接近，很容易混淆，而 10 进制则相邻的数字符号差别明显，不易

出错。这就是既然一个符号就已经足够表示所有自然数，人类为什么还需要发明0、1、2、3、4、5、6、7、8、9的根本原因。

五、改进一进制，分组、压缩，加法组合表示

我们已经发现，如果只使用一个符号|，也可以表示很多数量：

1 = |

2 = ||

3 = |||

4 = ||||

5 = |||||

.....

这种方法非常简单，它本质上就是一个加法式子 $1 + 1 + \dots + 1$ 。但是它有一个严重的问题：数越大，需要写的符号就越多。例如表示100，需要写100个|。表示1000，需要写1000个|。那么有没有办法：用更少的符号，表示更大的数？

5.1 第一次改进：把“1”分组

假设我们每看到一张选票就在候选人名字下面画一笔：|||||||。由于数起来很麻烦。于是可以规定：每5个|，组成一组。

例如中文环境中常见的“正”字计数，就是把五次记录组织成一组。

这种方法仍然没有改变一个基本原则：每发生一次，就增加一个笔画。

但是通过分组，我们已经更容易看出总数。例如：三个完整的“正”再加两笔。我们很快就知道是： $5 + 5 + 5 + 2 = 17$ 。

这里已经出现了一个非常重要的思想：**加法组合**。

我们不再把17看成17个毫无组织的“一”，而把它看成： $5 + 5 + 5 + 2$ 。

5.2 第二次改进：给“一组”发明新符号——压缩

既然每次都写5个符号还是麻烦，为什么不直接发明一个新符号，代表“一组”？如果我们把“正”用V来代替。6可以写成VI，7可以写成VII，而VIII就是 $5 + 1 + 1 + 1 = 8$ ，与|||||||相比，明显短了很多。

5.3 继续打包压缩，减少表示长度

如果数越来越大，我们还可以继续发明新符号。例如：

$$\begin{aligned}
 I &= 1 \\
 V &= 5 \\
 X &= 10 \\
 L &= 50 \\
 C &= 100 \\
 &\dots \dots
 \end{aligned}$$

这就是古罗马数字的基本思想。例如：XXVII可以理解为：10 + 10 + 5 + 1 + 1 = 27。

这里最基本的思想仍然是：把每个符号所代表的数量加起来。这就是一种典型的加性记数思想。

5.4 古埃及数字——离 10 进制一步之遥

古埃及的记数方法也采用了很强的“分组”思想。而且是按照不停扩大 10 倍进行分组。这与现代的 10 进制位值表示法已经非常接近。

他们为不同大小的数量设计不同符号如下表：

表示的数量	1	10	100	1,000	10,000	100,000	1,000,000
数字符号							
象形含义	一根竖线 (可能表示一根绳结 或一根棍子)	牛鞭或脚镣 (表示十)	盘绕的绳索 (表示一百)	莲花 (表示一千)	手指 (表示一万)	蝌蚪或青蛙 (表示十万，象征大量)	举起双臂的人 (表示一百万，象征无数)

如果需要表示352就写成



也就是100 + 100 + 100 + 10 + 10 + 10 + 10 + 10 + 1 + 1，这与写352个|相比，已经方便了很多。

但是注意，一个表示100的符号^②，无论写在哪里，都是表示100。符号的位置并不影响数值，加性组合表示法中，352有很多种写法，只要100，10，1的个数分

为3, 5, 2即可。符号本身决定它的值, 位置并没有承担我们今天十进制中那么重要的作用。

思考题

7. 请写出432的罗马数字和古埃及数字表示法。
8. 为什么古埃及数字没有给0设计符号?

5.5 加性表示的共同特点和共同缺陷

虽然“正”字、古罗马数字、古埃及数字看起来完全不同, 它们背后有一个共同思想: 先给一些常用的数量规定符号, 再通过加法把这些数量组合起来。这和我们用现金支付的时候, 用固定面值的纸币去凑一个特定金额的思想完全一样。

例如要表示27, 可以把它分解成 $10 + 10 + 5 + 1 + 1$ 。我们把这一类方法称为: 加性组合。它比先进得多, 因为人类已经开始, 把许多个“1”打包。将符号进行压缩表示。

但是, 加性组合仍然有缺陷。假设我们不断遇到越来越大的数。为了缩短表示, 我们可能不断需要新的符号, 甚至出现符号爆炸现象, 使得使用者的记忆负担非常重。第二个问题是: 表示方法不够统一。例如27即可以分解成 $10 + 10 + 5 + 1 + 1$ 也可以分解成 $10 + 5 + 5 + 5 + 1 + 1$, 有些数字表示法, 还引入了减法组合, 表示方法就更多了。

六、思想的突破——位值表示法

加性组合最不令人满意的是如果要表示无限大的数, 需要创造无限个符号。现在, 我们提出一个新的要求。假设我们只准备有限的几个数字符号。例如只使用: 0、1、2、3、4、5、6、7、8、9。但是, 我们希望: 无论数有多大, 都不需要再发明新的数字符号。

一百万也好, 一亿也好, 再大的数也好, 永远只使用这十个数字。可能做到吗?

答案隐藏在一个新的想法中: 把符号的无限性, 转化为空间(位置)的无限性。不再只让“符号本身”表示数量, 还让“位置”表示数量, 而位置是无限的。

观察现代的10进制表示法, 1, 10, 100, 1000, 这里始终只有一个数字**1**。

但是，在 1 中，它表示 1 个一。在 10 中，它表示 1 个十。在 100 中，它表示 1 个百。在 1000 中，它表示 1 个千。符号没有改变，位置改变了。于是位置本身开始携带信息。

如果数字写成一行，从低位到高位，每个位置赋予一个数值，该位置的符号就代表多少个该位置的数值相加，这种表示法称为“位值法”。

从加性组合到位值法是一个巨大的飞跃。以 10 进制位值法为例，我们只保留 0、1、2、3、4、5、6、7、8、9 这 10 个数字符号。但是同一个数字放在不同位置，价值不同。

例如：555

第一个 5 表示：5 个百；第二个 5 表示：5 个十；第三个 5 表示：5 个一。所以 $555 = 5 \times 100 + 5 \times 10 + 5 \times 1$ 。

思考题

9. 写出 10 进制数 5607 对用的加性组合式子。
10. 为什么一个正整数的 10 进制表示法是唯一的？

6.1 最规整的位值表示法——进制表示法

如果从低位到高位，位值依次为 $1, N, N^2, N^3, \dots$ ，并且仅为 $0, 1, 2, \dots, N-1$ 创造独立符号，这样的位值表示法就称为 N 进制表示法，最典型的的就是 $N = 10$ ，10 进制表示法。

因为每个数位的数字都不能大于 $N-1$ ，而 $(N-1) + 1 = N$ ，在 N 进制表示法就是 $(N-1) + 1 = 10$ ，这就产生了逢 N 进 1 的加法规则。

N 进制表示法解决了加性组合的一个缺陷，就是表示的唯一性。以 10 进制为例，一个数 a ，如果用十进制表示分为 $a_n a_{n-1} \dots a_1 a_0$ ，那么 $a = a_n \times 10^n + a_{n-1} \times 10^{n-1} + \dots + a_1 \times 10 + a_0$ ，那么 a 除以 10 的余数就是 a_0 ，商就是 $a_n \times 10^{n-1} + a_{n-1} \times 10^{n-2} + \dots + a_1$ ，而正整数的带余数除法的商和余数都是唯一的，所以 a_0 被唯一确定， $a_n \times 10^{n-1} + a_{n-1} \times 10^{n-2} + \dots + a_1$ 也对应唯一确定的商 a' ，如果 $a' > 0$ 可以对 a' 进行同样的余数分析，依次类推，可以知道 $a_n, a_{n-1}, \dots, a_1, a_0$ 都是唯一的。

N 进制还有一个好处，就是比较 2 个数的非常直观，位数多的大于位数少的，位数相同从高位到低位开始比较，数字大的就大。

思考题

11. 为什么 N 进制中3位数一定大于2位数?
12. 为什么 10 进制中一个数是否为偶数, 只需要看个位数?
13. 讨论: 实际应用中如果加2个很大的数, 竖式加法是从低位算起更好还是从高位算起更好?

6.2 0 在位值制中的作用

我们已经知道, 在 N 进制中, 不同的位置代表不同大小的单位。

以十进制为例:

个位 $\rightarrow 1$

十位 $\rightarrow 10$

百位 $\rightarrow 100$

千位 $\rightarrow 1000$

现在我们遇到一个问题。怎样表示: 一百零五?

它有: 1 个百, 0 个十, 5 个一, 所以, $105 = 1 \times 100 + 0 \times 10 + 5 \times 1$ 。中间的“十位”什么都没有。怎么办?

假设我们把中间什么都没有的十位直接省略。那么 105 就会写成 15。可是 15 表示: 1 个十 + 5 个一, 而不是: 1 个百 + 5 个一。

所以, 在位值制中, 一个位置即使什么都没有, 这个位置也不能随便消失。

于是0承担了一个极其重要的任务: **占位**。在位值记数法中: 0 不仅表示“没有”, 还表示“这个位置存在, 但这一位有 0 个单位”。

6.3* 无 0 位值制

目前流行的 N 进制的 N 个基本数字为 $0, 1, 2, \dots, N-1$, 但是如果仅仅用于表达正整数, 0 不是必须的, 我们可以以 $1, 2, \dots, N$ 为 N 个基本数字, 位值依然是 $1, N, N^2, N^3, \dots$, 这样的表示法称为“无 0- N 进制”。以无 0-10进制为例, 我们给 10 分配一个符号 κ , 这样原来 10 进制的15就依然是15, 而100则表示为 $9\kappa = 9 \times 10 + \kappa$, 105则表示为 $\kappa 5 = \kappa \times 10 + 5$ 。

这种无 0-10 进制表示法，完全不需要 0，并且其表示法也是唯一的，因为如果把正整数 a 表示为这种无 0 的进制表示法， $a = a_n \times 10^n + a_{n-1} \times 10^{n-1} + \dots + a_1 \times 10 + a_0$ ，容易知道 a_0 除以 10 的余数和 a 除以 10 的余数一样，而 a_0 只能取 $1, 2, \dots, \kappa$ ，根据 a 除以 10 的余数就可以确定 a_0 ，此时 $\frac{a-a_0}{10} = a' = a_n \times 10^{n-1} + a_{n-1} \times 10^{n-2} + \dots + a_1$ ，若 $a' > 0$ 又可以用类似的方法确定 a_1 ，以此类推，可以知道 $a_n, a_{n-1}, \dots, a_1, a_0$ 都是唯一的。

既然 N 进制可以有 0 也可以无 0，为什么历史选择了有 0 的 N 进制呢，这是因为 0 不仅仅是占位符，它本身就是一个具有重要性质的数，有 0 可以使得加减乘除运算算法更加简便，优美，我们将在下一章讨论 0 和 1 的性质。

思考题

14. 标准十进制的 3020，用无 0 - 10 进制表示应该是什么？
15. 标准十进制的加法是逢十进一，请问无 0 - 10 进制的加法应该如何描述？