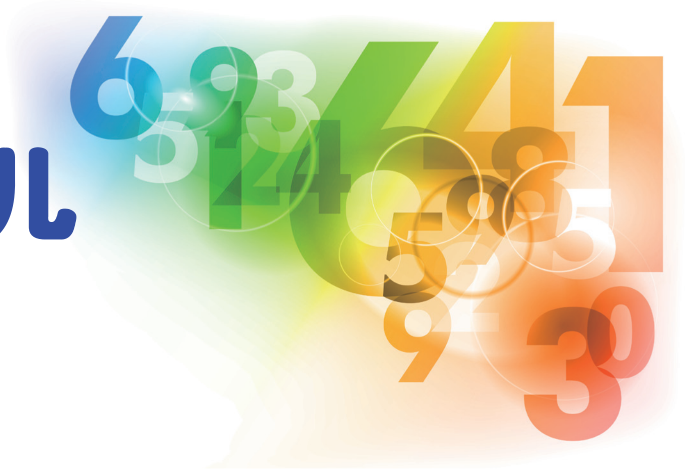
素数的那些事儿



陆俊

1 引子

近几年开设《初等数论》课程，我总是要抽出一定的时间专门给学生科普一下关于素数的故事。这篇科普文章正是基于这些讲稿整理出来的。

素数是整个数论的灵魂。然而多数学生对素数的了解非常少。很多人不明白：为什么我们要研究素数？素数如何与众不同？素数到底有趣在哪里？素数对数学很重要吗？……如果学生在上完一个学期的数论课后，却仍然对素数茫然无知，那无疑是一种讽刺——这就好比你看完一场戏，不知道主角做了些什么。

写这篇文章的另一目的也是为了给那些依然执着于证明哥德巴赫猜想的民科们做一次扫盲的尝试——尽管他们中的大多数会继续执着下去。然而我们不得不承认这样一个现实：民科们对素数的热情与执着确实远远超过很多数学系的本科生——这多少会让我们这些老师感到沮丧。

2 素数有多少?

我们说一个整数 a 能被另一个整数 b 整除，就是指 $\frac{a}{b}$ 是整数。有时我们也把 b 称作 a 的因子。

一个素数（Prime Number）是指这样一种正整数：除了 1 和它本身之外，其它任何正整数都不可能整除它。我们也可以这么定义素数：它不能写成两个大于 1 的正整数的乘积。有时我们也将素数称作质数。通常我们不承认 1 是素数。这样做的好处下面会介绍。除了 1 和素数之外，其他的正整数统称为合数。

最初的几个素数是 2, 3, 5, 7, 11,…。显然 6 不是素数，因为 $6=2\times3$，所有的素数中只有 2 是偶数！这件看似平凡的事，其实很重要。在许多数学研究中，2 和其他素数会对我们所考虑的问题产生不同的影响。你可能会问：为什么我们把这样的数命名为“素数”呢？这实际上来自于素数最基本的结论——**算术基本定理**：

> 任何大于 1 的正整数 n 都可以唯一地分解成一些素数的乘积 $n=p_1\cdots p_s$，这里 $p_1\le p_2\le\cdots\le p_s$ 都是素数（允许相同）。

无论如何，素数本身不能再进一步分解成一些更小的正整数乘积，因此它在此意义下是最基本或最本质的数——类似于朴素的原子论——从而命名它们为“素数”或者“质数”。容易看到，假如我们承认 1 是素数，那么算术基本定理就不能保证分解是唯一的了。 因为 1 可以写为任意多个自身的乘积。

接下去，一个最自然不过的问题当然是：究竟有多少素数？无限多个还是仅有有限个？这个问题的答案早由欧几里德在两千多年前解决了。他用初等方法巧妙地证明：存在无限多个素数！具体言之，我们假设所有正整数中只有有限个素数 $p_1,\cdots,p_n$，那么可以构造一个正整数

$$N=p_1p_2\cdots p_n+1\ .$$

很容易发现，左边的 N 分解成素数乘积的话，不可能包含任何素数 p_i，因此它的分解式中必定含有这些 p_i 之外的新素数。这就和我们的假设矛盾！

这个证明包含了富有启发性的思想。事实上，证明本身并没有提供构造出所有素数的具体方法。但是它却能告诉我们素数有无限个！这就是数学中所谓的“存在性证明”：它告诉你某些对象存在，但是却没有具体构造出来。“存在性”是数学哲学中的一个深刻话题，涉及到数学大厦的根基。数学史上曾经关于这类问题有过广泛而激烈的争论，有人反对这种类型的证明，有人却支持它们。这场争论涉及了许多重要的数学家，产生了许多和数学、逻辑、哲学相关的理论。有兴趣的读者可以参考相关书籍，此处不再赘述。

类似欧几里德的证明，你也可以轻松断言：所有被 4 除余数为 3 的素数有无限个！换言之，就是等差数列 3, 7, 15,

19, …中包含无穷多个素数。这就产生了一个有趣的问题：

一个等差数列 $a, a+b, a+2b, \cdots, a+nb, \cdots$ 中是否包含无限多个素数？

数学家狄利克雷回答了这一问题（**狄利克雷定理**）：

> 假如 a 和 b 是互素的（就是说它们不能同时被一个大于 1 的正整数整除），那么答案是肯定的！

不要以为欧几里德的方法可以轻松解决这一问题哦。事实上，除了少数情形之外，这个问题是不可能用它来简单解决的。

如果我们把等差数列换成其他数列，结论会怎样呢？比如考虑以下的数列：

$$2, 5, 10, 17, 26, \cdots, n^2+1, \cdots$$

其中是否有无限多个素数呢？让人颇为失望的是，这至今仍是一个未解决的难题。

3 素数是怎么分布的？

知道“素数有无限多个”仅仅是个开始。我们还想知道更多！比如，素数在所有自然数中所占的比率多大？当然，我们首先要说明“比率”在这里意味着什么。对任何正实数 x，我们用 $\pi(x)$ 表示不超过 x 的素数的个数。比如 $\pi(1)=0$, $\pi(4)=2$, $\pi(2.5)=1$ 等等。我们用 $\frac{\pi(x)}{x}$ 来反映所有不超过 x 的正整数中，素数所占的比率——也称作平均分布密度。

一个简单的结论告诉我们：当 x 非常非常大时，$\frac{\pi(x)}{x}$ 几乎就等于 0。换句话说，素数在所有正整数中极为罕见，可以说少得几乎没有——尽管我们知道它们有无穷多个！这就好比宇宙中有生命的星球也许有无限多个，但是它们相隔得太远，相对整个宇宙来说实在是十分稀疏罕见的。

对一般人来说，这个结论似乎已经让我们走到了问题的尽头。但是天才数学家高斯却不这么认为。在那个没有计算机的年代（1792-1793 年间），他通过大量的手工计算，单凭超人的直觉，竟然得到了一个让人吃惊的猜测（但其本人并未证明）：

> 当 x 非常大时，素数出现的比率 $\frac{\pi(x)}{x}$ 约等于 $\frac{1}{\log x}$。换言之，$\frac{\pi(x)}{x/\log x}$ 约等于 1，这里 $\log x$ 是 x 的对数函数。

高斯原始的猜测要比上面的表述式更为精确。在高斯之后，数学家勒让德实际上也通过数值计算得到过类似的猜测公式（1800 年左右），但没有高斯的精确。证明这一结论是极其困难的工作。直到 19 世纪中叶，俄国数学家切比雪夫才有了突破性进展，他证明了：

$$C_1 \le \frac{\pi(x)}{x/\log x} \le C_2,$$

这里 C_1 和 C_2 是确定的常数。此猜想大约到 19 世纪末，才由法国数学家阿达玛和 Paussin 几乎同时独立证明。人们将它称作**素数定理**。阿达玛等人的证明是建立在天才数学家黎曼的研究基础上的，用到了极为高深的函数理论。到了 1949 年前后，才由数学家爱尔特希和塞尔伯格给出了初等证明。请注意，这里所谓的“初等”只是说没有用到太多高深的数学理论，但是证明本身是很复杂的，也较为难懂。数学中有很多这样的问题（比如哥德巴赫猜想），它们表面上很简单，但实际上要证明它们往往是极其困难的。

素数定理只是在大样本范围内描述了一种统计规律。素数本身的分布位置极不规则。当你确定一个素数之后，很难预测在它之后的下一个素数是多少。尽管如此，我们仍有一些猜测和结论来描绘素数在整数集中分布性态。有趣的是，猜想要比结论多得多。

首先是著名的**伯特兰猜想**（后被切比雪夫证明），它断言：

对任何大于 1 的正整数 n，必定有素数落在 n 和 $2n$ 之间。

比如 n 取 4，那么在 4 到 8 之间我们可以找到素数 5 和 7。当 n 非常大时，这一结论显然是素数定理的直接推论。

你可以随手举出很多类似伯特兰 - 切比雪夫定理的猜想，比如在 n^2 和 $(n+1)^2$ 之间是否必有素数存在？这一看似简单的问题实际上至今仍未解决！

其次是著名的**孪生素数猜想**：

是否存在无限多个素数 p，使得 $p+2$ 也是素数？

我们将这样的一对素数 $(p, p+2)$ 称为孪生素数对。比如 (3, 5), (5, 7), (11, 13) 等等都是孪生素数对。类似地，你也可以定义三生素数对 $(p, p+2, p+6)$，亦即要求这三个数同时为素数。三生素数猜想就是问：是否存在无限个三生素数对？回答仍是“不知道”。我们也可以定义 n 生素数对，并提出类似的猜测。有趣的是，有人证明：n 生素数猜想和以下的三角不等式猜测互为矛盾——也就是说不可能同时正确：

$$\pi(x+y) \le \pi(x) + \pi(y),$$

这里 $\pi(x)$ 定义同前。

另一个著名的猜想就是在国内广为人知的**哥德巴赫猜想**：

> 任何大于等于 6 的偶数必定能写成两个奇素数之和；任何大于等于 9 的奇数都是三个奇素数之和。

这个猜想和陈景润的名字联系在一起，带有很多现代历史的色彩。许多民科投身于哥德巴赫猜想的证明也与此有关。哥德巴赫只是一个普通的数学家，除了提出这个猜想之

外没有什么数学贡献。他将这一猜测告诉了天才数学家欧拉。遗憾的是，后者未能证明它，但是该猜想却得以被很多人知道。容易看到，哥德巴赫猜想第二部分只不过是第一部分的简单推论。但有趣的是，第二部分反而先被证明了（称作**三素数定理**），第一部分却迟迟得不到解决。目前最好的结果是陈景润的“1+2”定理，即充分大的偶数都可以写成一个素数和一个不超过两个素数乘积的数之和。哥德巴赫猜想的研究是十分艰难的，它本质上涉及到十分深刻的函数论知识，不可能如那些民科所妄想的那样，拍拍脑袋就能用初等方法做出来。

虽然我们无法彻底证实这个猜想，但是却可以退而求其次，用所谓的*密率方法*得到以下的有趣结论：

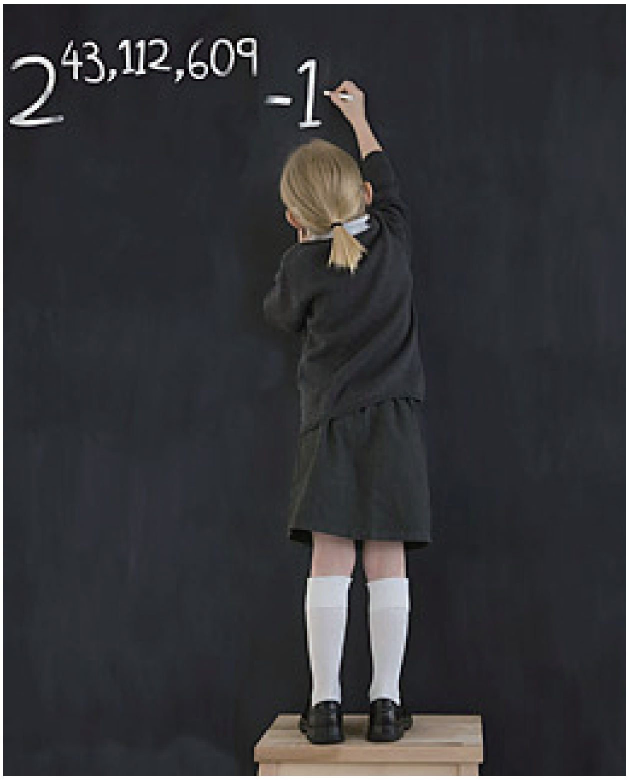
> 任何大于1的正整数必可写成不超过26个素数之和。

4 如何构造素数？

上面的讨论只是介绍了素数在整数中的分布情况，但是我们至今还没有具体构造出这些素数来。一个基本的问题就是：如何构造素数？最原始的办法就是古典筛法。比如我们要找出所有不超过 100 的素数，那么首先将所有从 $4 = 2^2$ 开始的偶数全部从这 100 个数中去除掉；接着将所有从 $9 = 3^2$ 开始的 3 的倍数全部去除掉；再将所有从 $25 = 5^2$ 开始的 5 的倍数全部去除掉……以此类推，最终通过筛选剩下的数恰好就是所有不超过 100 的素数。

上面的筛法虽然可以逐一列出不超过某个上限 N 的全部素数，但是当 N 很大时，其工作量也是巨大的。因此人们开始寻找其他方法来构造素数。通常的思路是构造一个有规律的数列 $\{a_n\}_{n=1}^{\infty}$，使得数列中每一项都是素数。这样的数列称作素数公式。比如费马构造了以下数列（费马数），并

数学家群星图；排在最上面的是数论结果及高斯和欧拉等数论先驱



美国一家网络安全基金会悬赏超长素数；1 千万位素数的十万美元被加州大学的 Edson Smith 领走。他找到了第 46 个梅森数（见上图）

猜测它们都是素数：

$$F_n = 2^{2^n} + 1, \qquad n = 0, 1, 2, \cdots$$

他计算了前五项，即 3, 5, 17, 257, 65537，确实都是素数。然而欧拉以其高超的计算能力手算验证了 $F_5 = 641 \times 6700417$ 不是素数。事实上，由目前的计算机验算可知，从 F_5 到 F_{11} 都不是素数。是否存在无限多个费马素数？这是一个未解之谜。

尽管欧拉的计算粉碎了利用费马数构造素数公式的企图，但是这并不表示研究费马数没有意义。高斯在年少时期，证明了一个让人无比惊叹的奇妙结论，让费马素数声名大噪。这个结论（**正 m 边形尺规作图**）是说：

> 一个正 m 边形能用尺规作图得到的充分必要条件是：$m=2^e p_1 p_2 \cdots p_s$，这里诸 p_i 是费马素数，且两两不同。

比如 $m = 17$ 是费马素数，因此可以用尺规作图得到正十七边形！要知道，在高斯之前的几百年，有那么多人研究尺规作图问题，但谁也没有想到正十七边形居然可以尺规作图得到。这个结论的重要性在于，它将几何（代数）问题和数论问题这两个看似无关的领域奇妙地结合起来。

与费马数对应的是著名的梅森数列：

$$M_p = 2^p - 1, \qquad p = 2, 3, 5 \cdots$$

这里 p 依次取遍所有的素数。人们也曾猜测梅森数都是素数。比如前几项分别为 3, 7, 31, 127 都是素数。但是 $M_{11} = 23 \times 89$ 不是素数。一个有趣的结论断言：

> 如果素数 p 被 4 除余数为 3，并且 $2p+1$ 也是素数，那么 M_p 必定不是素数。

梅森数的遭遇与费马数类似，尽管没有能够达到原始的构造素数公式的目的，但是它却和另一个著名的定理联系起来。为了叙述该定理，我们做些准备工作。给一个正整数 n，我们把它的所有可能的因子（就是能整除 n 的那些正整数）加起来得到的总和记作 $\sigma(n)$。如果 n 满足 $\sigma(n) = 2n$，那么我们就称 n 是完全数。比如 $n = 6$ 就是完全数，因为 n 的因子只有 1, 2, 3, 6, 加起来正好是 12。同样地，$n = 28$ 也是完全数。我们有如下的**偶完全数定理：**

> 所有的偶完全数都可以写成 $\frac{1}{2}p(p+1)$，这里 p 是梅森素数。反之，这样的表达式得到的数也必定是偶完全数。

上面说的 $n = 6$ 恰好写成 $\frac{1}{2}3(3+1)$，其中 3 是梅森素数；28 可以写为 $\frac{1}{2}7(7+1)$，其中 7 是梅森素数。

由此产生另一个有趣的问题：奇完全数存在吗？这又是数论中一个至今悬而未决的著名猜想。人们借助计算机检验了 10^{300} 以内所有数，竟然都没能找到奇完全数！另外一个同样让人沮丧的事实是，至今我们还不知道是否有无穷多个梅森素数。

欧拉提出了另一类构造素数的方式。比如考虑多项式 $n^2 - n + 41$。当 n 从 0 取到 40 时，多项式的值皆素数。我们同样可以考虑多项式

$$n^2 - n + p \ ,$$

这里 p 是素数，使得当 n 从 0 开始直至某个数 N 为止逐一代入时，上述多项式取值始终为素数。对任意 N，我们是否总能找到这样的 p 满足上面的要求呢？这个有趣的问题也是未解决的难题之一。让我们在上述多项式中分别取 $n = 1, 2, 3$, 那么得到的三个值恰好为 $p, p+2, p+6$。如果上面的问题答案是肯定的话，这就立刻证实了前文所述的孪生素数猜想和三生素数猜想！因此很显然上面的问题要远远难于孪生或三生素数猜想。