

孪生素数问题

为什么只差 2 的素数对可能有无穷多？

孪生素数问题的陈述非常简单：像 $(3, 5)$ 、 $(5, 7)$ 、 $(11, 13)$ 这样相差 2 的素数对，会不会永远出现下去？截至本文整理时，人们相信答案是“会”，但还没有证明。

这份笔记先补齐两个基础结论，再回答孪生素数问题：

1. 欧几里得怎样证明素数有无穷多个；
2. 素数定理怎样描述素数的总体密度；
3. 孪生素数猜想精确地说了什么；
4. 为什么概率直觉强烈支持这个猜想；
5. 已有结果离间隔 2 还差在哪里。

目录

1. 两个关于素数的基础事实	2
1.1. 欧几里得：素数有无穷多个	2
1.2. 素数定理：素数有多稀疏	2
2. 孪生素数问题	3
2.1. 为什么最小间隔是 2	3
2.2. 两个基础结论为什么仍不够	3
3. 为什么人们相信它	3
3.1. 最粗略的概率直觉	3
3.2. Hardy-Littlewood 的精细预测	4
4. 为什么证明很难	4
4.1. 筛法能做什么	4
4.2. 只差一个几乎素数	4
5. 有界素数间隔的突破	4
6. 几个常见误解	5
7. 应该怎样理解这个问题	5
8. 延伸阅读	5

1. 两个关于素数的基础事实

一个大于 1 的整数，如果正因数只有 1 和它本身，就称为素数。前几个素数是

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \dots \quad (1)$$

1.1. 欧几里得：素数有无穷多个

定理

素数有无穷多个。

欧几里得给出了一条很短的反证法。假设素数只有有限多个，把它们全部列成

$$p_1, p_2, \dots, p_k. \quad (2)$$

构造一个新整数

$$N = p_1 p_2 \dots p_k + 1. \quad (3)$$

对每个已列出的素数 p_i ，都有

$$N \bmod p_i = 1, \quad (4)$$

所以没有一个 p_i 能整除 N 。另一方面， $N > 1$ ，因此 N 本身是素数，或者 N 至少有一个素因子 q 。这个 q 不在原来的清单中，与“已经列出了全部素数”矛盾。于是素数不可能只有有限多个。

这个证明告诉我们：无论已经找到多少素数，总能证明还有新的素数存在。但它没有告诉我们新素数离已有素数多远，也没有描述素数总体上有多稀疏。

1.2. 素数定理：素数有多稀疏

定义素数计数函数

$$\pi(x) = |\{p \leq x : p \text{ 是素数}\}|, \quad (5)$$

也就是不超过 x 的素数个数。例如 $\pi(10) = 4$ ，因为这些素数是 2, 3, 5, 7。

素数定理

当 x 趋向无穷时，

$$\pi(x) \sim \frac{x}{\ln x}, \quad (6)$$

等价地，

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1. \quad (7)$$

符号 $f(x) \sim g(x)$ 表示二者的比值趋近于 1。它不是说 $\pi(x)$ 在每一点都等于 $\frac{x}{\ln x}$ ，而是说后者抓住了前者的主要增长尺度。

x	$\pi(x)$	$\frac{x}{\ln x}$
10	4	4.34
100	25	21.71
1000	168	144.76
1 000 000	78 498	72 382.41

因此，在大小约为 x 的位置附近，素数的平均密度约为 $\frac{1}{\ln x}$ ，相邻素数的平均间隔约为 $\ln x$ 。这里的“平均”很重要：素数定理描述宏观分布，并不保证每个长度为 $\ln x$ 的区间都有素数，更不控制某个指定的短间隔。

2. 孪生素数问题

孪生素数

若 p 和 $p+2$ 都是素数，就称 $(p, p+2)$ 为一对孪生素数。

前几对孪生素数为

$$(3, 5), (5, 7), (11, 13), (17, 19), (29, 31), (41, 43), (59, 61), \dots \quad (8)$$

孪生素数猜想

存在无穷多个素数 p ，使 $p+2$ 也是素数。

设 p_n 是第 n 个素数。猜想也可以写成

$$\liminf_{n \rightarrow \infty} (p_{n+1} - p_n) = 2. \quad (9)$$

这里用下极限，是因为素数间隔会忽大忽小。猜想并不是说相邻素数最终总相差 2，而是说间隔 2 会在任意大的位置反复出现。

2.1. 为什么最小间隔是 2

除了 2 以外，所有素数都是奇数。两个奇数的差一定是偶数，所以足够大的相邻素数不可能只差 1；最小的可能间隔就是 2。

模 3 还能解释一个有趣现象。在三个奇数

$$n, \quad n+2, \quad n+4 \quad (10)$$

中，必有一个能被 3 整除。因此，除了包含素数 3 的 $(3, 5, 7)$ ，不可能再有三个连续奇数全是素数。换句话说， $(3, 5)$ 和 $(5, 7)$ 是唯一两对“首尾相接”的孪生素数。

2.2. 两个基础结论为什么仍不够

欧几里得定理只保证素数不会用完，并不控制它们之间的距离。例如，一个无穷集合完全可能越来越稀疏，相邻元素的距离也越来越大。

素数定理虽然给出了平均密度，却仍是宏观结论。它允许局部间隔剧烈波动，也没有说明 n 与 $n+2$ 的素性如何关联。孪生素数猜想要求的强得多：它要求在素数整体越来越稀疏时，仍有无穷多次恰好出现最小间隔 2。

3. 为什么人们相信它

3.1. 最粗略的概率直觉

由前面的素数定理，大小约为 x 的整数成为素数的“密度”大约是

$$\frac{1}{\ln x}. \quad (11)$$

如果暂时把“ n 是素数”和“ $n+2$ 是素数”当成近似独立事件，那么二者同时为素数的机会大约是

$$\frac{1}{(\ln x)^2}. \quad (12)$$

把 x 以内所有候选数的机会加起来，就会得到大约

$$\int_2^x \frac{dt}{(\ln t)^2} \quad (13)$$

对孪生素数的贡献。这个积分随 x 趋向无穷，因此粗略模型预测孪生素数有无穷多对。

但“独立”并不准确。比如对任意素数 q ，整数对 $(n, n+2)$ 能否避开被 q 整除，带有明确的同余约束。需要把这些局部效应逐个修正。

3.2. Hardy-Littlewood 的精细预测

记

$$\pi_2(x) = |\{p \leq x : p \text{ 和 } p+2 \text{ 都是素数}\}|. \quad (14)$$

Hardy-Littlewood 素数对猜想预测

$$\pi_2(x) \sim 2C_2 \int_2^x \frac{dt}{(\ln t)^2} \sim 2C_2 \frac{x}{(\ln x)^2}, \quad (15)$$

其中

$$C_2 = \prod_{q>2, q \text{ 为素数}} \left(1 - \frac{1}{(q-1)^2}\right) \approx 0.66016 \quad (16)$$

称为孪生素数常数。

这个公式比“有无穷多对”更强：它预测了 x 以内大概有多少对。计算实验与这一趋势非常吻合，但再多的有限计算也不能代替对所有整数的证明。

概率模型的地位

把一个整数“当作随机数”是启发式方法，不是证明。素数由整除性严格决定，各个事件之间也不是独立的；常数 C_2 正是在修正这些相关性。

4. 为什么证明很难

4.1. 筛法能做什么

筛法的基本思路是：从候选整数中排除能被小素数整除的数。若同时筛 n 和 $n+2$ ，就能很好地估计“二者都没有小素因子”的候选数量。

难点在于，没有小素因子不等于一定是素数。一个数还可能是两个大素数的乘积。许多筛法可以证明“素数或只有少量素因子的数”大量存在，却很难同时认定两个表达式都恰好只有一个素因子。

这通常称为筛法的**奇偶障碍**：经典筛法难以区分一个整数含奇数个还是偶数个素因子，而“恰好是素数”正需要这种更精细的信息。

4.2. 只差一个几乎素数

陈氏定理

存在无穷多个素数 p ，使 $p+2$ 要么是素数，要么是两个素数的乘积。

“两个素数的乘积”允许平方，例如 q^2 。陈氏定理已经非常接近孪生素数：左边确定是素数，右边至多有两个素因子。但从“至多两个”推进到“恰好一个”，正会撞上筛法最顽固的障碍。

5. 有界素数间隔的突破

令

$$d_n = p_{n+1} - p_n \quad (17)$$

表示相邻素数的间隔。孪生素数猜想等价于 $d_n = 2$ 出现无穷多次。

2013 年，张益唐证明存在某个固定常数 B ，使

$$d_n \leq B \quad (18)$$

出现无穷多次。关键不是最初的 B 有多大，而是它不随素数的位置增长：即使素数越来越稀疏，仍有无穷多对相邻素数保持在某个统一的有限距离内。

随后，Maynard 和 Tao 发展出新的方法，并与协作改进一起显著缩小了可取的界。这些工作证明了

$$\liminf_{n \rightarrow \infty} (p_{n+1} - p_n) < \infty, \quad (19)$$

而孪生素数猜想要求把右边的某个有限界一直压到 2。

“有界间隔”还不是“孪生素数”

已知结果保证无穷多次出现“不超过某个常数”的素数间隔，但没有指定这个间隔必须是 2。即使由抽屉原理可知，某个固定的偶数间隔会出现无穷多次，我们仍不知道它是不是 2。

这些突破的思想来自 GPY 方法：不再试图直接证明指定的 $(n, n+2)$ 同时为素数，而是考察一组平移

$$n + h_1, n + h_2, \dots, n + h_k \quad (20)$$

并设计权重，证明其中至少两个数经常同时为素数。这样可以先得到“某个短区间里有两个素数”，却无法精确控制究竟是哪两个位置胜出。

6. 几个常见误解

说法	问题在哪里
“已经验证到极大的范围，所以猜想成立。”	有限计算只能排除某个范围内的反例，不能证明无穷性。
“素数无穷多，所以总有相差 2 的素数。”	无穷多个数仍可能彼此越来越远。
“张益唐已经证明了孪生素数猜想。”	他证明的是无穷多个统一有界的素数间隔，没有把界降到 2。
“随机模型就是证明。”	随机模型解释了猜想为何可信，但必须由数论方法处理整除相关性。

7. 应该怎样理解这个问题

孪生素数问题位于两种力量的交界处：

- **总体稀疏性**：素数在 x 附近的密度只有约 $\frac{1}{\ln x}$ ；
- **局部重复性**：尽管越来越稀疏，间隔 2 似乎仍会反复出现。

我们已经能证明素数之间存在无穷多次的统一有界间隔，也能证明“素数加几乎素数”的间隔可以固定为 2；但要让间隔两端同时都是素数，现有筛法仍缺少关键的一步。

一句话总结

孪生素数猜想断言最小可能的奇素数间隔 2 会出现无穷多次；大量理论和计算证据都支持它，而证明的核心困难，是现有筛法难以同时确认两个相隔 2 的数都只有一个素因子。

8. 延伸阅读

可以沿着这条历史主线继续阅读：Brun 的筛法与孪生素数倒数和，Hardy–Littlewood 的素数对猜想，陈景润的“素数加几乎素数”定理，Goldston–Pintz–Yıldırım 的 GPY 方法，以及张益唐、James Maynard、Terence Tao 关于无条件有界素数间隔的工作。