

张益唐与孪生素数猜想

孪生素数猜想

王元

王元，数论学家，主要从事解析数论研究。原籍江苏镇江，出生于浙江兰溪。1952年毕业于浙江大学。中国科学院院士，曾担任中国科学院数学所所长。由于在哥德巴赫猜想研究中取得的卓越成果与陈景润、潘承洞获国家自然科学一等奖（1982）。提出数论方法在近似分析中应用的“华-王”方法。从事原创性科学工作的同时，十分重视并参与科普写作。

编者按：本文主要参考了 J. Pintz 的文章 Landau's problems on primes, *Journal de Théorie des Nombres de Bordeaux*, 2009, pp. 357-404.

一

自然数可以分成三类：1) 1；2) 素数 p ，即仅含 1 与 p 为因子的大于 1 的自然数，如 2, 3, 5, 7, $\dots$ ；3) 复合数 n ，它含有两个大于 1 的素因子，如 4, 6, 8, 9, $\dots$ 。

首先是 Euclid 证明了素数有无穷多，接下来的一个自然的问题是孪生素数问题。在素数集合中，我们观察到有些素数对，如

$$3, 5; 5, 7; 11, 13; 17, 19; \dots$$

它们都是素数，而其差为 2。一般说来，当

$$p, p+2$$

都是素数，我们就称它们为一对孪生素数。

所谓孪生素数猜想是说：

$$\text{孪生素数对有无穷多。} \quad (1)$$

这是数学中迄今尚未解决的著名难题之一。实际上，这个猜想表明，素数不是孤独的，它永远会以 2 为距离成对地出现！

猜想是什么时候被提出来的呢？若以文献记载为准，则可以追溯到法国数学家 de Polignac (1849)。有人将它（或隐含地）归属于 Euclid 或 Eratosthenes，以纪念 Euclid 证明素数有无穷多及 Eratosthenes 用筛法构造素数表。

二

1900 年，Hilbert 在第二届国际数学家大会上，向 20 世纪数学家提出了 23 个待解决的数学问题，其中第八问题，包括黎曼 (Riemann) 猜想 RH¹⁾ 及两个变量 x, y 的线性代数方程

$$ax + by = c \quad (2)$$

在素数集中的求解问题，其中 a, b, c 为满足某必要条件的整数。

1) 当 $a = 1, b = -1, c = 2$ ，方程 (2) 有无穷多对素数解，即对应于孪生素数猜想。

2) 当 $a = b = 1$ ，偶数 $c \geq 4$ 时恒有解，即 Goldbach 猜想。

1912 年，Landau 在国际数学家大会的报告中提出了四个重大的数论问题供研究，孪生素数猜想与 Goldbach 猜想是其中的两个。

1920 年前后，Hardy, Ramanujan, Littlewood 创立了堆垒数论的新方法——圆法。Hardy 与 Littlewood 利用圆法对孪生素数猜想与 Goldbach 猜想重新表述如下：

命 $\pi_2(x)$ 表示不超过 x 的孪生素数对数， $r_2(n)$ 表示将偶数 n 表示为两个素数之和的表示法个数，则有

$$\pi_2(x) \sim 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \frac{x}{\log^2 x} \quad (3)$$

与

$$r_2(n) \sim 2 \prod_{\substack{p|n \\ p>2}} \frac{p-1}{p-2} \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \frac{n}{\log^2 n}, \quad (4)$$

¹⁾关于 RH、GRH、Siegel 零点、允许 k 数组可以在解析数论书中找到。例如，《王元谈求学之路》，大连理工大学出版社，2010。

其中 p 表示素数, 及

$$\prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) = 0.6601 \dots \quad (5)$$

显然 (4) 和 (5) 分别比孪生素数猜想与 Goldbach 猜想更深刻。至今尚不能证明 (4) 和 (5), 困难在于无法估计圆法中的“劣弧”上的积分的上界, 即使假定了广义黎曼假设 GRH (见第 9 页的脚注) 亦然。

由于 (4) 的右端与偶数 n 有关, 即使 (4) 成立, 也只能导致 Goldbach 猜想对充分大的偶数成立, 而 (3) 不跟某数相关, 所以有人认为 (3) 比 (4) 的提法更合理。

三

不少人用数值计算来检查孪生素数猜想。已知小于 10^5 的自然数中有 1224 对孪生素数, 即 $\pi_2(10^5) = 1224$ 。还有以下结果:

$$\pi_2(10^6) = 8164, \quad \pi_2(3.3 \times 10^7) = 152892, \quad \pi_2(10^{11}) = 224376048.$$

还知道一些大的孪生素数对, 例如

$$107570463 \times 10^{2250} \pm 1.$$

目前所知道的最大孪生素数对是 2011 年发现的

$$3756801695685 \times 10^{666669} \pm 1.$$

四

关于孪生素数猜想的重大成果首先是由筛法得到的。筛法起源于公元前 250 年的所谓 Eratosthenes 筛法, 在一定范围内的素数表就是以这个方法为基础造出来的。

1919 年, Brun 本质上改进了 Eratosthenes 筛法, 提出了他的新筛法并成功地应用到数论中许多极困难与重要的问题, 孪生素数猜想与 Goldbach 猜想就是其中最主要的两个, 而且关于这两个猜想的结果常常是相伴产生的, 所以从筛法的眼光看, 这两个问题是“姊妹问题”。

Brun 证明的定理为

存在无穷多的整数 n , 使 n 与 $n+2$ 都是不超过 9 个素数之积, 简记为 (9, 9). (6)

我们可以类似地定义 (a, b) , 孪生素数猜想就是要证明 $(1, 1)$ 。

关于 Goldbach 猜想的相应结果为“每个充分大的偶数都是两个素因子个数不超过 9 的整数之和”。我们以后就不再表述 Goldbach 猜想的相应结果了。

Euler 曾证明过

$$\sum_p \frac{1}{p} = \infty,$$

这里 p 表示素数, 从而推知素数有无穷多。

Brun 用他的方法证明了

$$\sum_{p^*} \frac{1}{p^*} < \infty, \quad (7)$$

此处 p^* 为所有的孪生素数。

如果级数 $\sum_{p^*} \frac{1}{p^*}$ 发散, 则孪生素数猜想就成立了。遗憾的是由 $\sum_{p^*} \frac{1}{p^*}$ 的收敛, 并不能得出孪生素数对数是有限还是无穷的结论!

不少数学家改进了 Brun 方法与他的结果, 现在将这些成果列于下:

$$\begin{aligned} &(7, 7), \text{Rademacher (1924),} \\ &(6, 6), \text{Estermann (1932),} \\ &(5, 7), (4, 9), (3, 15), (2, 366), \text{Ricci (1937),} \\ &(5, 5), \text{Buchstab (1938),} \\ &(4, 4), \text{Buchstab (1940).} \end{aligned} \quad (8)$$

1941 年, Kuhn 发表了 his 加权筛法, 从而可以得到

$$(a, b), \text{ 其中 } a + b \leq 6. \quad (9)$$

1947 年, Selberg 给出了 Eratosthenes 筛法另一个重大改进。他于 1950 年宣称用他的方法可以得出 $(2, 3)$, 但未给出细节, 基于他的方法, 出现了下面的结果:

$$\begin{aligned} &(3, 4), \text{王元 (1956),} \\ &(3, 3), \text{A. Vinogradov (1957),} \\ &(2, 3), \text{王元 (1957, 1958).} \end{aligned} \quad (10)$$

Selberg 关于 $(2, 3)$ 的证明细节是 1991 年发表的。

五

对于 (a, b) 中有一个等于 1 的情况, 匈牙利数学家 Rényi 首先于 1947 年证明了

$$(1, c), \quad (11)$$

此处 c 是一个大的未明确给定的常数。

过去, $(1, c)$ 型的结果都是建立在 GRH 之上的。

命 $\vartheta > 0$ 。若对于任何 $A > 0, \vartheta > \varepsilon > 0$, 我们有

$$\sum_{q \leq X^{\vartheta - \varepsilon}} \max_{\substack{a \\ (a, q) = 1}} \left| \sum_{\substack{p \equiv a \pmod{q} \\ p \leq X}} \log p - \frac{X}{\varphi(q)} \right| \ll_{\varepsilon, A} \frac{X}{\log^A X}, \quad (12)$$

其中 q, a 表示正整数, p 表示素数, $\varphi(q)$ 表示 Euler 函数, 即模 q 的既约剩余类个数, 则称 ϑ 是一个允许标 (admissible level)。

Rényi $(1, c)$ 的证明实际上依赖于他证明了存在某个小 ϑ 使 (12) 成立。潘承洞首先证明了 $\vartheta = 1/3$, 并由此推出 $(1, 5)$ 。以后潘承洞又与 Barban 独立地证明了 $\vartheta = 3/8$ 并得出

$$(1, 4), \text{ 潘承洞 (1962, 1963), Barban (1963)}. \quad (13)$$

Bombieri 与 A. Vinogradov 独立地证明了 $\vartheta = 1/2$ 是一个允许标, 从而得出:

$$(1, 3), \text{ Bombieri (1965), A. Vinogradov (1965)}. \quad (14)$$

最后, 应用 Selberg 筛法, 某种形式的 Kuhn 加权筛法, 允许标 $\vartheta = 1/2$ 及他自己的转换原理 (switching principle), 陈景润成功地证明了

$$(1, 2), \text{ 陈景润 (1966, 1973)}. \quad (15)$$

这是迄今为止, 基于筛法所能达到的最佳结果。

六

相邻素数差的估计是解析数论的中心问题之一, 命

$$d_n = p_{n+1} - p_n,$$

此处 p_n 表示第 n 个素数, 我们需估计 d_n 的上、下界。

d_n 的上界估计, 即相邻素数间的最大间隙估计。目前最好的估计为

$$d_n \ll p_n^{21/40}, \text{ Baker, Harman, Pintz (2001)}. \quad (16)$$

注意在黎曼假设 RH 之下, 可以证明

$$d_n \ll \sqrt{x} \log x, \text{ Cramer (1920)}, \quad (17)$$

此处 d_n 表示不超过 x 的最大相邻素数差。

现在我们来陈述关于 d_n 的下界估计问题，即寻找最小可能的间隙 d_n ，使之对于无穷多个整数 n 成立。

如果孪生素数猜想成立，则 d_n 的下界问题就完满地解决了，答案就是 2。

由素数定理

$$\pi(x) \sim \frac{x}{\log x},$$

此处 $\pi(x)$ 为不超过 x 的素数个数，可知 d_n 的平均值为

$$\log p_n.$$

因此我们可以考虑量

$$\Delta = \liminf_{n \rightarrow \infty} \frac{d_n}{\log p_n}$$

的上界估计。第一个非寻常的估计是 1926 年，Hardy 与 Littlewood 在 GRH 之下证明的，即

$$\Delta \leq 2/3, \text{ Hardy, Littlewood (1926), GRH.} \quad (18)$$

不假定未经证明的猜想，Erdős 首先证明了

$$\Delta \leq 1 - c, \text{ Erdős (1940),} \quad (19)$$

其中 c 是一个可计算的小常数。

基于允许标 $\vartheta = 1/2$ (见 (12))，则有

$$\vartheta = \frac{2 + \sqrt{3}}{8} = 0.4665 \dots, \text{ Bombieri, Davenport (1966).} \quad (20)$$

最佳纪录为

$$\Delta \leq e^{-r} \times 0.4425 \dots = 0.2484 \dots, \text{ Maier (1988),}$$

其中 r 为 Euler 常数。

七

关于孪生素数猜想，我们来介绍 Heath-Brown 的一个重要的条件结果 (1983)。他证明了，由 Siegel 零点（见第 9 页的脚注）的存在性可以推出孪生素数猜想。更一般些，对于每个偶数 k ，均存在无穷多对素数，其差为 k 。

很自然地，绝大多数数学家都不相信有 Siegel 零点。例如，由 GRH 成立就能推出 Siegel 零点不存在。因此，一般认为经过 Heath-Brown 的结果来证明孪生素数猜想是没有什么希望的。

现在我们来陈述关于孪生素数猜想较弱的版本：

$$\text{小间隔猜想: } \Delta = 0 \quad (21)$$

与

$$\text{有界间隔猜想: } \liminf_{n \rightarrow \infty} (p_{n+1} - p_n) < \infty. \quad (22)$$

显然 (22) 比 (21) 强得多，实际上，上一节的工作之理想即在于证明 (21)。

2005 年，Goldston, Pintz 与 Yildirim 出色地证明了小间隔猜想，即 (21)：

$$\Delta = 0, \text{ Goldston, Pintz, Yildirim (2005)}. \quad (23)$$

关于 (23) 的一个简化且自含的证明由 Goldston-Motohashi-Pintz-Yildirim (2006) 得到。

关于有界间隔猜想，Goldston-Pintz-Yildirim (2005) 有条件地加以证明了。详言之，他们证明了

$$\begin{aligned} &\text{若 (12) 存在允许标 } \vartheta > \frac{1}{2}, \text{ 则对 } k \geq C(\vartheta), \text{ 任意允许} \\ &k \text{ 数组 (admissible } k\text{-tuple) (见第 9 页的脚注) 必包} \\ &\text{含无穷多次含有两个素数, 此处 } C(\vartheta) \text{ 为仅依赖于 } \vartheta \\ &\text{的常数. Goldston-Pintz-Yildirim (2005).} \end{aligned} \quad (24)$$

特别他们证明了

$$\text{若存在允许标 } \vartheta > 0.971, \text{ 则 } k \geq 6. \quad (25)$$

6 数组 $(n, n+4, n+6, n+10, n+12, n+16)$ 是允许 6 数组，所以由 Elliott-Halberstam 猜想，即 $\vartheta = 1$ 推出

$$\liminf_{n \rightarrow \infty} (p_{n+1} - p_n) \leq 16. \quad (26)$$

我们知道由 GRH 只能推出 $\vartheta = 1/2$ ，所以 $\vartheta > \frac{1}{2}$ 就意味着在很多算术序列中，素数的分布是异常有规则的，所以要证明有允许标 $\vartheta > \frac{1}{2}$ 是极困难的！

2013 年，张益唐出人意料地取得了划时代的成就，他成功地证明了有界间隔猜想，详言之，他证明了

$$\liminf_{n \rightarrow \infty} (p_{n+1} - p_n) \leq 7 \times 10^7, \text{ 张益唐 (2013)}. \quad (27)$$

这说明有一个常数 $d (\leq 7 \times 10^7)$ ，存在无穷多对素数，其差为 d 。这该是多么美妙啊！

八

关于 Goldbach 猜想也有重要进展。最近，法国数学家 Helfgott 证明了

每个奇数 $n \geq 7$ 都是三个素数之和。 (28)

早在 1937 年，I. Vinogradov 就证明了，对于充分大的奇数 n ，即 $n \geq C$ ，结论 (28) 成立。I. Vinogradov 的方法允许有效地定出常数 C 。关于 C 的确定，有 Borozdkin、陈景润和王天泽等人的工作，但跟理想结果 $C=7$ 还相去甚远。1997 年在 GRH 之下，Deshouillers, Effinger, te Riele 与 Zinoviev 证明了 $C=7$ 。