質數愈大愈孤獨：談質數分布

作者　劉建亞

作者簡介

劉建亞現任山東大學教授，數學學院院長，專長為數論。《數學文化》主編。

本文首先簡單介紹數論的某些研究領域、研究方法。然後，著重介紹數論的重要研究領域——質數分布。鑒於本文的性質，我們的介紹只能從歷史和哲學的角度出發。從哲學的角度論述，就可能會摻雜作者的思辨甚至臆測；正如羅素（Bertrand Russell）[7] 指出的那樣，哲學乃臆測之學。然而，這種思辨或臆測只是針對數學的意義；臆測絕不能冒充數學內容，也不能冒充數學史。

數論

數論是研究整數的性質的，也被稱爲整數論或高等算術。數論有許多領域，下列是諸多重要領域中的兩個：

（1） 質數論；

（2） 丟番圖方程。

所謂質數論就是研究質數的分布。含有兩個或以上變量的方程稱爲不定方程，而丟番圖方程理論關心的是整係數不定方程的整數解的存在性及其分布。本文從第二節起專門介紹質數分布；關於丟番圖方程，只能另文介紹了。

我們必須強調，上述分類並不全面。事實上，除了上述兩個領域之外，數論還包含一些其他的領域，比如組合數論，即用組合的方法研究一些數列的性質；還有超越數論，例如證明圓周率π是超越數，自然對數的底e是超越數，等等。所謂超越數是指，它不是任何一個有理係數方程的根。這些領域，本文只能割愛了。

數論固然是研究整數的，但是數論的研究方法並不只局限於數手指頭、打算盤、或打高級算盤——計算機。現在，認爲數論只能利用手指或算盤來研究的人，在社會上並不在少數。用非整數的、甚至是連續的方法來研究整數的性質，這本身就有深刻的哲學；下文對此會再有論述。研究數論的方法，概括說來有四種：

（1） 初等方法；　　（2） 解析方法；

（3） 代數方法；　　（4） 幾何方法。

第一種辦法是初等方法。初等方法並不是僅僅使用初等數學，而是說，僅僅使用微積分及其以前的數學知識。換句話說，初等方法是牛頓

和萊布尼茲都知道的方法。初等方法還包含組合方法，近幾十年來，組合方法在數論中大有用武之地。第二個辦法是解析方法，也就是用分析的方法，包括實分析、複分析，以及更抽象層面上的泛函分析。第三種辦法是代數方法，要用到高等代數與高深的抽象代數，比如群、環、體、迦羅瓦理論，以及群表現論，等等。群表現論是一個很重要的數學話題。在群上附加拓樸結構，從而將代數和分析有機地融合在一起，這樣的群就叫做李群。一般來說，群的線性表現已經不是一件很容易的事情了，更不用說一個李群的表現論了。代數方法，就是要用到這樣的一些數學知識。第四個辦法，就是幾何方法；這裡的幾何方法是指代數幾何。算術代數幾何現在也是一個很熱門而且成果豐碩的研究方向。

離散分布於數軸上的整數，怎麼會跟分析、代數、幾何聯繫上呢？對於不研究數論的人士來說，這是個很神祕的話題。本文會盡量解釋。

質數論

第一大領域質數論研究質數在整數中的分布規律，比如質數定理和孿生質數猜想；大家可以從初等數論中瞭解質數分布的簡單結果。在更抽象的層面上，在代數數論中，我們也研究質理想（prime ideal）的分布。

當然，質數論還包括堆壘質數論（additive theory of prime number），即質數的加法理論。堆壘質數論涵蓋大家熟知的哥德巴赫猜想，即大偶數能夠寫成兩個質數之和，大奇數能寫成三個質數的和。高次的堆壘質數論問題還包括所謂的華林/哥德巴赫問題（Waring-Goldbach problem），即把滿足必要條件的大整數表示成質數的 k 次方之和，其中 k 是一個固定的正整數。這些問題都屬於質數論。

神祕的質數

畢達哥拉斯說，萬物皆數。克羅涅可（Leopold Kronecker）有一句名言，只有整數是上帝給的，其餘都是人類自己製造的。不過，大於1的整數都能唯一地分解爲質數的乘積，這說明整數可以進一步由質數製造出來，而質數是不可分解的。因此，質數就是構成整數的原子。質數表之於數學，就相當於元素週期表之於化學。沒有比質數更基本的東西了，所以質數分布規律的探求具有基本的重要性。然而，質數表與元素週期表有一個重大的區別：質數表是無限的！

質數分布是一個非常古老的問題。如果查看數軸，很難預測哪個大整數是質數。實際上，我們似乎不可能發現質數出現的任何具體規律，也就是說，我們還沒有辦法寫一個通項公式，使其每項都產生質數。所以，質數分布好像是隨機的一樣。更進一步，即使找到了一個質數，也很難知道下一個質數何時出現。相鄰質數之間的距離有時是2，比如3與5，11與13，等等。相鄰質數之間的距離有時也可以任意大，比如對於事先給定的任何大整數 N，數列 $N!+2, \cdots, N!+N$ 中都沒有質數。

機率統計的思想指出，對於隨機事件，我們只能觀察它的統計規律。比如，當扔一個硬幣時，我們雖然不知道它什麼時候正面朝上或是反面朝上，但是可以採取大量的重複試驗，以發現其統計規律。對於質數分布，同樣的思路依然有效。我們很難預測一個大整數到底是不是質數。但是，我們可以在數軸上取一個很大的 x，這個 x 可以趨向於無窮大，然後研究 x 前面有多少個質數，即不超過 x 的質數的個數。

記不超過 x 的質數的個數爲 $\pi(x)$，我們來研究 $\pi(x)$ 和 x 之間的關係。研究 $\pi(x)$ 要比推測具體哪個數是質數容易。

初中生或者小學生都知道一個定理，即自然數當中有無窮多個質數。這個定理最初發表於歐幾里得的《原本》。用數學的語言就是說，當 x 趨於無窮大的時候，$\pi(x)$ 也趨於無窮大。但是，職業數學家很自然地就會問下一個更精確的問題：$\pi(x)$ 是怎麼樣趨向於無窮大的，也就是 $\pi(x)$ 趨於無窮的速度是多少？有了趨於無窮的主項，數學家還得問餘項有多大。數學家或者潛在的數學家，必須關心這些更

加精確的數學問題。高斯和勒讓德（Adrien-Marie Legendre）經過了大量的計算，在1792 年左右分別提出了下面這個猜想。

質數分布猜想 (Gauss-Legendre, 約1792)

當 $x \to \infty$ 時，

$$\pi(x) \sim \frac{x}{\log x} \tag{1}$$

上述猜想指出，在 x 趨向於無窮大的時候，$\pi(x)$ 趨向於無窮大的速度是$x/\log x$。在數論中，習慣使用 $\log x$ 表示以 e 爲底的自然對數。這就是說，在不超過 x 的自然數中，質數出現的機率是 $1/\log x$。這就表明，質數在整個自然數集合中是零密度。用機率的語言來說，若在數軸上隨便拿一個自然數，則拿到質數的機率是0。在1792年之後，人類最具智慧的頭腦花了一百多年的時間，終於完成了這個猜想的證明。質數猜想被證明後，被稱爲質數定理。質數定理的證明是數學史上的璀璨篇章，這些將在稍後介紹。

結束本節之前，我們再敘述兩個著名猜想；這兩個猜想更能讓人體會質數之神祕。考察一對質數，如果它們之間的差是2，那麼就稱它們爲一對孿生質數。比如3與5，11與13，等等。大家最初的印象是，孿生質數的分布相當密集；實際上，稍微往後看幾個例子就會發現，孿生質數似乎相當稀疏。傳說古希臘人考察了這些孿生質數，提出了一個著名的孿生質數猜想。

孿生質數猜想：存在著無窮多對孿生質數。

下面的猜想涉及質數在稀疏集合中的分布，也是古希臘人提出的。

猜想：數列 n^2+1 中存在無窮多個質數，其中 n 是自然數。

這兩個猜想都是質數論中很根本、很困難的問題，到現在也沒有被證明。證明兩個猜想當中任何一個，一定會名垂青史，與數學共輝煌。1920年代，哈第（Godfrey H. Hardy）與李托伍德（John E. Littlewood）給出了上述兩個猜想的精確表達；最近沙納克（Peter Sarnak）提出了一個綱領性的猜想，將以往質數分布的結果與猜想網羅殆盡。以上兩個猜想的研究有很多進展，其中更有中國數學家的傑出貢獻。對此，另闢專文介紹爲宜。

英雄

必須強調這樣一個事實：高斯比勒讓德年歲小很多。高斯生於1777年，他提出質數猜想是在15歲左右，所用的工具只是一張對數表和一張質數表。「群眾創造歷史」固然不假，然而，總是有那麼幾個個別的群眾，他們在歷史上所發揮的作用和別人很不一樣。這幾個個別的群眾就被稱爲英雄。高斯就是一位英雄，他對數學的發展起了異常重要的作用。一個15歲的孩子在那麼兩張表上進行演算，從而提出這個偉大的猜想，這絕對是一件不同凡響的事。

爲了進一步展現高斯的英雄本色，我們還必須強調，在人類歷史上，高斯似乎是第一個眞正理解質數神祕規律之人。勒讓德確實提出了猜測，在其猜測中只取首項，就蘊含公式（1）；但是勒讓德對其他較小項的猜測全是錯誤的。更糟糕的是，除了高斯與勒讓德之外，歷史上不乏智者猜測質數分布的規律，但是沒有一個是正確的。至於民間，研究質數分布之風一直很盛。陳景潤關於哥德巴赫猜想的著名工作發表之後，尤其是1978年徐遲的報告文學發表之後，我國有大量民間科學家群起加入哥德巴赫猜想研究。可以想見，研究哥德巴赫猜想，首先需要理解質數的分布規律，即理解質數定理。在作者所閱讀過的眾多民間科學家的稿件中，所得到的 $\pi(x)$ 的漸近性態都與質數定理矛盾！這更從另一個側面說明，提出質數猜想需要非凡的智慧和超常的洞察力。

初等方法

在高斯之後，許多人都想去證明質數猜想*。偉大的高斯自己也想證，但沒有找到證明的方法。質數猜想是當時數論的核心猜想之一，也是整個數學的核心猜想之一，是許多領袖數學家共同關心的問題。

*文後所述質數猜想通常指的都是質數分布猜想。

沿著高斯和勒讓德指明的方向，柴比雪夫（Pafnuty Chebyshev）證明了，當 x 充分大時，$\pi(x)$ 與 $x/\log x$ 的比值就落在與0.9與1.1之間；而且，若這個比值當 x 趨於無窮時有極限，其極限必爲1。柴比雪夫的上述不等式是說，$\pi(x)$ 的階應該是 $x/\log x$。但是，不等式中 $x/\log x$ 的上下界兩個係數與質數猜想中的1還有一定的差距。

這個不等式用初等方法就可以證明，證明也比較簡單，大概用幾頁紙就可以寫出來。柴比雪夫不等式的常數是可以改進的，我們可以將0.9做得更大一點，1.1做得更小一點，不過就是對不起來，得不到1。這說明，柴比雪夫的初等方法在質數猜想的證明中可能效果有限。

解析數論

分析方法如何用來研究數論，這是本文首先要破解的神祕。下面這個公式是用解析方法研究質數分布的出發點：

$$(2)\qquad \zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_p\left(1-\frac{1}{p^s}\right)^{-1}$$

這裡 s 是一個複數參數，它的實部大於1，也寫爲 $\mathrm{Re}(s)>1$。上式的第一個等式，就是黎曼（Bernhard Riemann）給出的著名的 ζ 函數的定義；第二個等式，由算術基本定理得來，最後無窮乘積的項 p 是所有質數。這裡需要注意，當 s 的實部大於1的時候，（2）中的無窮級數與無窮乘積都是絕對收斂的，所以 $\pi(x)$ 就在複半平面 $\mathrm{Re}(s)>1$ 是解析的。我們注意，（2）中的無窮乘積與質數緊密相關；也就是說，一個複變函數與包含質數信息的一個無窮乘積是相等的。

在黎曼之前，歐拉曾經對實參數 s 得到過公式（2），並由此出發證明了歐幾里得定理，即上文所述的質數無窮多。也是在黎曼之前，狄利克雷（Lejeune Dirichlet）將乘法特徵引入了歐拉版本的公式（2），證明了若一個等差數列的公差與首項互質，則其中必含有無窮多個質數。狄利克雷的這個著名的定理，將歐幾里得定理推廣到等差數列中。與歐拉的版本相比，黎曼的公式（2）的一個重大的區別是，s 是一個複數參數。

名詞解釋

算術基本定理：大於1的正整數有唯一的質因數分解。

極點：局部形如 $\frac{1}{z^n}$ 的奇點稱為極點。當 $n=1$ 時，稱為單極點。

零點：即函數之根。

1851年，黎曼在高斯指導下獲得博士學位。1859年對於黎曼來說是特殊的年份。這一年，他被聘爲哥廷根大學數學教授，繼任了高斯的職位；也是在這一年，他當選柏林科學院通訊院士。新當選的院士需要做一個主題演講，上述公式（2）就發表在黎曼的那一篇演講[6]當中。演講的題目是〈論不超過一個大整數的質數的個數〉，也就是研究 $\pi(x)$。

大家回憶，在黎曼那個時代，研究一個複變函數的性質是很困難的一件事情，因爲當時複變函數的理論尚未完全確立，甚至很多概念都是不清楚的。現在的複變函數教科書裡都有一個著名的柯西/黎曼方程（Cauchy-Riemann equation），黎曼也確實是複變函數論的創始人之一。他對 ζ 函數的研究表明，他對複變函數有著非常深入的洞察力。

在那篇偉大的演講裡，黎曼證明了 ζ 函數的函數方程，從而將 ζ 函數解析延拓到整個複平面。黎曼更研究了 ζ 函數的零點分布，得到了下面的結論（R1）與（R2）：

R1 **ζ 函數在 $s=1$ 處有單極點（simple pole）。ζ 函數有實零點 $-2,-4,\cdots$，但是這些實零點相對不重要；重要的是 ζ 函數有無窮多個非實的複零點。**

R2 **ζ 函數所有的複零點都落在無窮帶形 $0\le\mathrm{Re}(s)\le1$ 裡，並且對中線 $\mathrm{Re}(s)=1/2$ 對稱。上述帶形稱爲臨界帶形，中線稱爲臨界直線。**

從（R1）結合高斯所證明的代數基本定理，可以看出，ζ 函數是一個超越函數，而不是一個多項式

或有理函數。這是黎曼在那篇演講中所證明的結論。

一個自然的問題是：黎曼沒有直接去研究質數分布以證明高斯與勒讓德的猜想，而是定義了這樣一個 ζ 函數，研究它的零點分布，他這樣做意圖何在？下面的（R3）就回答這個疑問。黎曼進一步證明了：

R3 如果在臨界帶形的邊界上 ζ 函數沒有零點，那麼高斯和勒讓德的質數猜想就成立。

將質數的分布與一個半純函數（meromorphic function）的零點分布相聯繫，這就標誌著解析數論的誕生。黎曼本人也不能證明臨界帶形的邊界上沒有零點。他只能證明，如果邊界上沒有零點，那麼質數猜想就是正確的。

質數猜想的解析證明—黎曼的哲學

重大進展發生在世紀之交。1896年，阿達瑪（Jacques Hadamard）[3][4]和瓦里普桑（Charles de la Vallée-Poussin）[9] 沿著黎曼的思路，獨立地證明了質數猜想。實際上，他們證明了 ζ 函數在臨界帶形邊界上確實是沒有零點，因此根據黎曼的論述，$\pi(x)$ 就等價於 $x/\log x$。經過一百多年，阿達瑪和瓦里普桑使質數猜想變成了質數定理，這是一件了不起的大事。1958年，維諾格拉多夫（Ivan Vinogradov）更推進了質數定理的餘項。

爲了證明質數猜想，阿達瑪花費很長時間來研究有限階整函數（entire function of finite order）。他證明了，這類整函數可以寫成展布在其零點上的無窮乘積。這套理論相當深刻，當代數論當中的自守 L-函數（automorphic L-function）也都可以用有限階整函數來刻畫。

讓我們總結一下黎曼的哲學思路。假如有一個算術或者幾何的問題，從這個問題出發，定義一個複變函數，然後研究這個函數的解析性質，利用得到的解析結果，就能解決最初提出的問題。用一個圖形來表示，就是這樣：

對黎曼來說，這個算術問題就是高斯與勒讓德的猜想。從這個算術問題出發，他定義了黎曼 ζ 函數；而通過研究 ζ 函數的解析性質尤其是零點分布，最終質數猜想得到證明。這就是黎曼的哲學思想，這種思想對日後數論的發展有著極其深遠的影響。

黎曼假說

在同一篇演講[6]中，黎曼還提出一個著名猜想，現在被稱爲黎曼假說（Riemann hypothesis）。只要是學數學的學生，無論是不是學習數論，可能都聽說過這個猜想。

黎曼假說：ζ 函數的所有複零點都落在臨界直線 $\mathrm{Re}(s) = 1/2$ 上。

這個猜想的提出，既有美學的原因，也有科學依據。首先，黎曼肯定相信高斯和勒讓德的猜想是對的。要想證明這個猜想是正確的，臨界帶形邊界上應該不能有零點。既然零點越靠邊界其作用就越壞，那麼，肯定零點越靠臨界直線 $\mathrm{Re(s)} = 1/2$ 越好。黎曼本人也知道，如果複零點都落在臨界直線 $\mathrm{Re(s)} = 1/2$ 上，質數猜想不僅成立，而且其餘項也能夠達到最佳。因此，如果所有零點都落在臨界直線上，世界將何其完美！

然而，任何一個猜想，不管在美學上是多麼完美，如果沒有一定的事實來支持，也非常有可能是錯誤的。黎曼在演講裡並沒有提到他是如何得到這個猜想的，但是留下了這方面的手稿。黎曼去世以後，其手稿歷經輾轉、火災，現存哥廷根大學圖書館。黎曼的手稿顯示，他進行過大量計算，手算出了三個複零點，結論是這三個複零點的實部全都等於1/2。我們注意，ζ 函數是超越函數。不要說超越函數，就是一

個高次的一般多項式也是沒有根式解的。因此，手工計算ζ函數的零點相當困難。無論如何，黎曼得到了支持自己猜想的依據。一個好的猜想，第一要有美學，第二要有相當的數學依據，兩者缺一不可。

哈第早年曾經證明了，ζ函數有無窮多個零點落在臨界直線上，但這無窮多個零點佔整體零點是零密度，就像質數在整數當中的密度一樣。塞爾伯格（Atle Selberg）進而證明了，落在臨界直線上的零點具有正的密度。現在最好的結果是，超過40%的零點落在臨界直線上。[1]這爲人們進一步研究黎曼假說堅定了信心。換句話說，即使黎曼假說是錯的，也不會錯得很離譜。

有民間科學家「證明」了ζ函數在臨界直線上徹底沒有零點！若果眞如此，黎曼假說將是人類歷史上錯得最離譜的猜想，因爲連一個支持猜想的例子都沒有！我們回憶，ζ函數在半平面 $\mathrm{Re}(s) \le 1$ 上是由解析延拓定義的，而不是由（2）定義的。該民間科學家論證的錯誤在於，強迫（2）式在臨界帶形內也成立。

上個世紀初，黎曼假說被希爾伯特（David Hilbert）收入他著名的23個數學問題裡面。這個世紀初，黎曼假說又被收入七個千禧年問題之中，且帶有百萬美元懸賞。希爾伯特曾經說過，如果一百年後我復活了，我要問，黎曼假說被證明了嗎？黎曼假說並不是一個孤立的猜想，它對整個數學的發展都有廣泛、深遠的影響，它是數學的核心猜想之一。因此，我眞誠地希望，當黎曼假說最終被證明的時候，那百萬美元獎賞能在屢經金融風暴之後仍然所値不菲。

質數定理的初等證明

自從質數定理用解析方法證明了之後，數論界普遍相信質數定理不能用初等方法證明。其原因是，質數定理與ζ函數的某個非零區域等價，而ζ函數是個複雜的半純函數；用初等方法得到複雜的半純函數的零點分布信息似乎不大可能，因此，用初等方法證明質數定理也似乎不大可能。然而，這一普遍的認識並不正確。1949年，艾狄胥（Paul Erdös）[2]與塞爾伯格[8]用初等方法證明了質數定理，兩人的證明都依賴於塞爾伯格的一個著名不等式。

不朽

哈第[5]說，「不朽」可能是個愚蠢的詞彙，但是別管其具體意義如何，數學家最有可能不朽。哈第的話，顯然是哲學意義上的；猜測或證明質數分布規律的英雄們，自然永垂不朽於數學史。實際上，即使是在現實的紅塵世界之中，也未必永遠是「自古英雄多寂寥」；英雄有著各種各樣的人生與結局。本文所談到的對質數分布作出實質性貢獻的幾位英雄，除了黎曼外，都算特別健康長壽。下面列出幾位的生卒年份：

勒讓德，1752-1833;
高斯，1777-1855;
柴比雪夫，1821-1894;
阿達瑪，1865-1963;
瓦里普桑，1866-1962;
維諾格拉多夫，1891-1983;
艾狄胥，1913-1996;
塞爾伯格，1917-2007.

只有黎曼不幸英年早逝，其生卒年份是1826-1866，享年尚不到40歲。∞

本文參考資料請見〈數理人文資料網頁〉
http://yaucenter.nctu.edu.tw/periodical.php

延伸閱讀

+ Derbyshire, John《質數魔力》(上)(下) (2005) 天下文化。
+ Edwards, H . M. , *Riemann's zeta function* (1974) AP, (2001) Dover。從黎曼的八頁文章談起，本書也是經典，適合更進深入的數學閱讀。
+ Riemann Hypothesis, CMI（http://www.claymath.org/millennium/Riemann_Hypothesis/）克雷數學研究所七大名題之黎曼假說網頁（包括黎曼珍貴的手稿）