

Integer Density Distribution and Prime Density

Yunhua Cui

Unit 95899, Beijing
Email: cuiyunhuacyh@163.com

Received: Apr. 17th, 2018; accepted: May 1st, 2018; published: May 10th, 2018

Abstract

The concept of the Integer density distribution is put forward and preliminarily studied. When It is applied to analyse the effect of the Eratosthenes' sieve, the prime density D is proved to satisfy $\frac{0.99903591\dots}{\ln n} < D < \frac{1.00489426\dots}{\ln n}$. The prime density $D = \frac{1}{\ln n}$ has not been completely proved in this paper, while the train of thought and the method to approach the conclusion are found. The mathematicians who accept this train of thought are expected to further study and put forward the authoritative demonstrations. The prime density is the foundation of the prime distribution and all prime conjecture, based on which, most prime issues can be solved.

Keywords

Integer, Prime Density, Prime Distribution, Eratosthenes' Sieve, Asymptotic Series, Prime Theorem, Prime Twins Conjecture, Goldbach's Conjecture

正整数密度分布与素数密度

崔蕴华

95899部队, 北京
Email: cuiyunhuacyh@163.com

收稿日期: 2018年4月17日; 录用日期: 2018年5月1日; 发布日期: 2018年5月10日

摘要

提出了正整数密度分布概念并进行了初步研究, 将其应用于埃氏筛法的“筛余截首”步骤作用分析, 得

出了素数密度 $\frac{0.99903591\dots}{\ln n} < D < \frac{1.00489426\dots}{\ln n}$ 的结论。尚未完成素数密度 $D = \frac{1}{\ln n}$ 的证明，但提出了完成证明的思路和逼近结论的方法，期待认同这一思路的数学家继续研究并提出权威论证。素数密度是素数分布和所有素数猜想的基础，大多数素数问题可据此解决。

关键词

正整数，素数密度，素数分布，埃氏筛法，渐近级数，素数定理，孪生素数猜想，哥德巴赫猜想

Copyright © 2018 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

为研究素数密度和素数分布，作者曾在中文杂志《前沿科学》2016年第2期发表论文《素数分布研究的一种新思路》[1]，核心方法是用埃氏(Eratosthenes)筛法对素数生成过程进行建模与递推。依埃氏筛法，为生成不超过正整数 n 的全部素数，以素数序列 $\{P\} = 2, 3, \dots, p_0, p, \dots, p_n \leq \sqrt{n}$ ，在不含1的正整数列 $\{I\} = 2, 3, 4, \dots, n, \dots$ 初始背景中，由2, 3依次到 $p_n \leq \sqrt{n}$ 实施筛法。文献[1]规范了埃氏筛法的三步骤操作规则：1) 后续 p 筛在前面 p_0 筛除剩余中进行无公倍数的纯净筛除，除第一个被筛除数 p^2 外，对 p_0 筛余周期内全部筛余数的筛除率和对其中任一筛余数的筛除概率均为 $d = \frac{1}{p}$ ，避免了因素数之间公倍数引起的重复筛除，简称“纯净筛除”。2) $\{2, 3, \dots, p_0, p\}$ 筛剩余(简称 p 筛剩余)形成以 $p^2 + 1$ 为起点，以 $C = \prod_p p$ 为周期的移位周期重构，简称“周期重构”。3) p 筛以概率1固定筛除 p^2 点，从而截取 p_0 筛余周期 $C = \prod_{1 < p \leq p_0} p$ 之首 (p_0^2, p^2) ，确认区间 (p_0^2, p^2) 的生成素数，简称“筛余截首”。

基于步骤1)，步骤2)，论文证明了 $2, 3, \dots, p_n \leq \sqrt{n}$ 筛除剩余在其周期 $C = \prod_{1 < p \leq \sqrt{n}} p$ 内平均密度

$$\hat{D} = \prod_{1 < p \leq \sqrt{n}} \left(1 - \frac{1}{p}\right) = \frac{1}{\frac{e^c}{2} \ln n} = \frac{1}{0.89053620\dots \ln n} \tag{1.1}$$

式中， c 为欧拉常数， e 为自然对数的底。

逻辑的结论是： n 邻域素数平均密度 D 必定在 \hat{D} 基础上，由“筛余截首”决定。“筛余截首”以概率1固定筛除 p^2 而不是以概率 $\frac{1}{p}$ 筛除 p^2 ， p^2 点的筛除概率比其它点增加了 $1 - \frac{1}{p}$ ，由此必然减小筛余周期前端密度，使区间 (p_0^2, p^2) 内素数平均密度低于 $C = \prod_{1 < p \leq \sqrt{n}} p$ 筛除剩余周期内剩余数的平均密度。

为讨论“筛余截首”对素数密度的影响，正整数 m 的密度分布成为必须直面的问题，首先对此进行概念研究。

2. 正整数在自然数列中的密度分布

定义 2.1 (正整数在自然数列中的密度分布): 正整数 m 在自然数列 $\{N\} = 0, 1, 2, \dots, m, \dots, n, \dots$ 中， m 存

在或被删除对 m 邻域和 $\{N\}$ 全域密度影响程度在 $\{N\}$ 全部元素上的分配函数 $f(n; m)$ (n 为自变量, m 为参变量), 称 m 在 $\{N\}$ 上的密度分布函数。正整数 m 在自然数列 $\{N\}$ 中的密度分布函数是以 n 为自变量, 以 m 为参变量的伽玛(Gama)函数

$$f(n; m) = \frac{n^m}{m!e^n} \tag{2.1}$$

对应的累积密度函数是

$$F(n; m) = 1 - \left(1 + n + \frac{n^2}{2!} + \dots + \frac{n^m}{m!} \right) e^{-n} \tag{2.2}$$

称对 $f(n; m)$ 在其主要作用域上对应的均匀分布函数为正整数 m 在自然数列中的等效均匀密度分布。

定义说明: $f(n; m)$ 满足以下基本要求

1) m 的密度函数 $f(n; m)$ 在 $\{N\}$ 全域上的总和

$$\sum_{n=0}^{\infty} f(n; m) = 1 \tag{2.3}$$

与概率密度函数同形异义。

2) m 为 $f(n; m)$ 均值与密度最大值, 对应于概率密度函数的数学期望和众值, 邻域在概率密度函数中应起主要作用。 $f(n; m)$ 应为一个以 m 为均值和最大值的双向平滑的曲线, 向 0 方向应以较快速率下降并交于 0 点, 向 $+\infty$ 方向应以较快速率下降并一直拖尾到 $+\infty$, 因此, $f(n; m)$ 应符合伽玛分布。

伽玛分布的一般表达式为[2]

$$f(n; \alpha, \beta) = \frac{1}{\alpha! \beta^{\alpha+1}} \cdot n^\alpha \cdot e^{-\frac{n}{\beta}} \tag{2.4}$$

对应的积累密度函数为

$$F(n; m) = 1 - \left(1 + \frac{n}{\beta} + \frac{1}{2!} \left(\frac{n}{\beta} \right)^2 + \dots + \frac{1}{\alpha!} \left(\frac{n}{\beta} \right)^\alpha \right) \cdot e^{-\frac{n}{\beta}} \tag{2.5}$$

使 n 满足众数的充要条件是 $\frac{df}{dn} = 0$, 即

$$\alpha n^{\alpha-1} \cdot e^{-\frac{n}{\beta}} - \frac{n^\alpha}{\beta} \cdot e^{-\frac{n}{\beta}} = 0$$

解得 $\alpha\beta = n$, 设 $n = m$ 时满足众数条件, 即

$$\alpha\beta = m \tag{2.6}$$

$n = m$ 同时应为 $f(n)$ 的均值, 即

$$\frac{\alpha}{\beta} = m \tag{2.7}$$

将(2.6) (2.7)联立, 解得

$$\begin{cases} \alpha = m \\ \beta = 1 \end{cases}$$

将此结果代入(2.4)即得(2.1), 代入(2.5)即得(2.2)。

3) 随着 m 的加大, $f(n; m)$ 应逐步逼近以 m 为中心、左右对称的正态分布, 但向 $+\infty$ 方向的无限拖尾永远不会消失, $f(n; m)$ 永不等同正态分布。因此, $f(n; m)$ 在区间 $(0, m]$ 的累级分布 $f(n; m)|_0^m$ 应逐步趋于 0.5^- , 在区间 $(m, +\infty)$ 的累级分布 $f(n; m)|_m^{+\infty}$ 应逐步趋于 0.5^+ 。

$f(n; m)$ 是自然数全域上的函数, 与正整数及其累积即阶乘密切相关, 这个定义反映了邻域与全域的关系。

猜想 2.2: 全体正整数的密度分布对同一自然数的总和为 1, 记作

$$\sum_{m=1}^{\infty} f(n; m) = 1 \tag{2.8}$$

由此使数轴上全体正整数布满时, 数轴上任何位置的密度为 1, 删去某一正整数对数轴上任何位置密度均有影响, 但影响的分布服从被删去正整数的密度分布, 全体正整数被删去后, 数轴上任何位置密度均为 0。由此建立数轴上正整数密度分布的加性运算规则, 表为

$$f(n) = \sum_m f(n; m) \tag{2.9}$$

式中, n 为任意选择的一个自然数, m 为全体正整数。

设想此猜想可通过对式(2.1)的函数分析证明。

推论 2.3 (正整数在自然数列中的等效均匀密度分布命题):

正整数 m 在自然数列 $\{N\}$ 中的等效均匀密度分布函数表为以 n 为自变量, 以 m 为参变量的函数

$$f(n; m) = \frac{1}{2m+k} \tag{2.10}$$

式中 $2m+k$ 为 m 密度分布的等效作用域, k 为位于区间

$$k_{\min} = 3 < k < k_{\max} = 4 \tag{2.11}$$

的确定常数, 数的性质和具体值待进一步研究仿真确定, 本文称 k 为素数常数。

证明: 先分析等效均匀分布的前提条件。 $m \rightarrow \infty$ 时, 伽玛分布无限趋近但永不同左右完全对称的正态分布, $m \rightarrow 1$ 时具有与 $m \rightarrow \infty$ 不同的拖尾效应。 m 点是相应伽玛分布的众数点和均值点, 因此, $m \rightarrow \infty$ 时, $F(n; m)|_0^m$ 将无限趋近但不会等于 $F(n; m)|_m^{2m}$, 差别主要在拖尾部分。所以 $m \rightarrow \infty$ 时, 必然存在常数 k 满足 $F(n; m)|_0^m \sim F(n; m)|_m^{2m+k}$, 且拖尾积累密度 $F(n; m)|_{2m+k}^{\infty} \rightarrow 0$ 。正整数的等效均匀分布以此为前提。

1) 由于 $m \rightarrow \infty$ 时伽玛分布趋向左右对称的正态分布, k 应是个小常数; 2) 对密度敏感的是小素数, 但大素数与小素数对渐近趋势的影响是一致的; 3) k 的选择要兼顾参变量 m 左右积累密度差距小和截尾影响小这一相互矛盾的因素 ($m \rightarrow \infty$ 时趋同)。

为更有针对性, 对 m 为小素数平方数的几种情况(任意情况趋势相同), 令 $m = p^2 = 2^2, 3^2, 5^2, 7^2, 11^2$, $k = 2, 3, 4, 5$, 计算结果如表 1。

计算仿真表明: k 由小变大时, $\frac{F(n; p^2)|_{p^2}^{2p^2+k}}{F(n; p^2)|_0^{p^2}}$ 由小变大, 而 $F(n; p^2)|_{2p^2+k}^{\infty}$ 则由大变小, 取 $k_{\min} = 3$ 或

$k_{\max} = 4$ 能较好兼顾要求: 满足在区间 $(0, p^2]$ 的积累密度与区间 $(p^2, 2p^2+k]$ 积累密度无限接近, 即

$F(n; p^2)|_0^{p^2} \sim F(n; p^2)|_{p^2}^{2p^2+k}$, 亦即 $\frac{F(n; p^2)|_{p^2}^{2p^2+k}}{F(n; p^2)|_0^{p^2}} \rightarrow 1$, 且拖尾积累密度 $F(n; p^2)|_{2p^2+k}^{\infty} \rightarrow 0$ 。

Table 1. Part of the gamma cumulative distribution calculation results
表 1. 伽玛累积分布部分计算结果

	k	$F(n; p^2) \Big _0^{p^2}$	$F(n; p^2) \Big _{p^2}^{2p^2+k}$	$\frac{F(n; p^2) \Big _{p^2}^{2p^2+k}}{F(n; p^2) \Big _0^{p^2}}$	$F(n; p^2) \Big _{2p^2+k}^{\infty}$
$p^2 = 4$	2	0.37116306	0.59958426	1.61542008	0.02925268
	3		0.61373234	1.65353831	0.01510460
	4		0.62123655	1.67375640	0.00760039
	5		0.62509676	1.68415671	0.00374018
	2		0.58241284	1.41159594	0.00499541
$p^2 = 9$	3	0.41259175	0.58464267	1.41700038	0.00276558
	4		0.58590332	1.42005583	0.00150493
	5		0.58734804	1.42355742	8.06020279E-4
	2		0.55289643	1.23668736	2.49895233E-5
	3		0.55290684	1.23671066	1.45743969E-5
$p^2 = 25$	4	0.44707858	0.55291298	1.23672440	8.43101530E-6
	5		0.55291658	1.23673243	4.83888054E-6
	2		0.53789559	1.16401317	1.17845007E-8
	3		0.46210439		6.99721276E-9
	4		0.53789560	1.16401318	4.13577076E-9
$p^2 = 49$	5	0.46210439			2.43355878E-9
	2				1.96004431E-18
	3		0.52415281	1.101514987	1.17817781E-18
	4		0.47584719		7.06805819E-19
	5				4.23194937E-19

已可看出 $m \rightarrow \infty$ 的基本趋势，即一定存在常数 k

$$k_{\min} = 3 < k < k_{\max} = 4$$

满足要求。 $m \rightarrow \infty$ 时，满足其在区间 $(0, m]$ 的积累密度与区间 $(m, 2m + k]$ 积累密度无限接近，即 $F(n; m) \Big|_0^m \sim F(n; m) \Big|_m^{2m+k}$ ，且拖尾积累密度 $F(n; m) \Big|_{2m+k}^{\infty} \rightarrow 0$ 。

忽略拖尾积累密度 $F(n; m) \Big|_{2m+k}^{\infty}$ ，将 m 的伽玛密度分布看作在区间 $(0, 2m + k]$ 的均匀分布，并称 $(0, 2m + k]$ 为均匀分布的等效作用域， m 在自然数列 $\{N\}$ 中的等效均匀密度分布函数可表为

$$f(n; m) = \frac{1}{2m + k}$$

推论 2.3 证毕。

在一定条件下，正整数的伽玛密度分布函数可用等效均匀分布来等效。

3. 素数平均密度

猜想 3.1 (素数平均密度): n 邻域素数平均密度

$$D = \rho \hat{D} = \frac{e^c}{2} \cdot \left(\frac{1}{\frac{e^c}{2} \ln n} \right) = \frac{1}{\ln n} \tag{3.1}$$

式中, $\rho = \frac{e^c}{2} = 0.89053620\dots$ 称素数特征数, 是一个由自然常数和欧拉常数表示的数。

式(3.1) n 小时 D 不准确; n 较大时 D 比较准确; $n \rightarrow \infty$ 时 D 趋于精确。

证明: 文献[1]已证明, $2, 3, \dots, p_n \leq \sqrt{n}$ 筛除剩余在其周期 $C = \prod_{1 < p \leq \sqrt{n}} p$ 内平均密度

$$\hat{D} = \prod_{1 < p \leq \sqrt{n}} \left(1 - \frac{1}{p} \right) = \frac{1}{\frac{e^c}{2} \ln n} \tag{3.2}$$

当 n 很大时($n \rightarrow \infty$)可等效表为

$$\hat{D} = \prod_p \left(1 - \frac{1}{p} \right) \tag{3.3}$$

式中 p 表示全部素数。

分析“筛余截首”作用。“筛余截首”是在 p_0 筛除剩余周期前端以概率 1 固定筛除相邻素数 p^2 , 由此确认 (p_0^2, p^2) 区间的素数。这必将降低 p_0 筛余周期前端的密度。由正整数密度的加性运算规则, 不失一般性, 可用降密因子 $\lambda = 1 - g(p^2)$ 表示。

对以概率 1 固定筛除 p^2 , 其中概率 $\frac{1}{p}$ 已在表示 p 筛降密作用的式(3.3) $\hat{D} = \prod_p \left(1 - \frac{1}{p} \right)$ 中考虑, 因此,

λ 应修正为 $\lambda = 1 - g(p^2) \left(1 - \frac{1}{p} \right)$ 。

由式(2.1), p^2 的密度分布函数是以 n 为自变量, 以 p^2 为参变量的伽玛函数

$$f(n; p^2) = \frac{n^m}{m! e^n} = \frac{n^{(p^2)}}{(p^2)! e^n} \tag{3.4}$$

考虑到: 1) p 相邻素数间平均间隔为 $\ln p$, 故 p^2 邻域素数间平均间隔为 $2 \ln p$, $2 \ln p = o(p^2)$, 即在间隔很大的被删除相邻 p^2 之间, 存在很多素数, 被筛除的 p^2 点相对于保留下的素数点是“稀疏”的。2) 相邻被筛素数平方数可分别表为 p^2 和 $(p + \ln p)^2$, p^2 的密度分布函数的根均方差为 $\sigma = p$, 且 $(p + \ln p)^2 \gg p$, 即 $(p + \ln p)^2$ 远在 $f(n; p^2)$ 的方差之外, 即被筛素数平方数远在相邻密度曲线的方差之外。3) 相邻伽玛密度函数的中心逐步移位和相似性。4) 被筛素数平方数 p^2 在筛除剩余数中随机分布。

由此使相邻被删除 p^2 的伽玛密度分布函数适用于等效均匀分布, 引式(2.10), 可用 $m = p^2$ 对应的均匀分布来等效。即

$$g(p^2) = \frac{1}{2p^2 + k} \tag{3.5}$$

再将这一降密作用嵌入到表示 p 筛降密作用的式(3.3)中, 随之一起递推, 形成加性降密因子 $\lambda = 1 - g(p^2) \left(1 - \frac{1}{p} \right)$ 之间的乘性递推。由此实现小素数降密因子向大素数的逐步传递, 并形成素数平均密度

$$D = \prod_p \left\{ \left[1 - g(p^2) \left(1 - \frac{1}{p} \right) \right] \left(1 - \frac{1}{p} \right) \right\}$$

等效表为

$$D = \prod_p \left[1 - g(p^2) \left(1 - \frac{1}{p} \right) \right] \cdot \prod_p \left(1 - \frac{1}{p} \right) \tag{3.6}$$

令素数特征数

$$\rho = \prod_p \left[1 - g(p^2) \left(1 - \frac{1}{p} \right) \right] \tag{3.7}$$

即

$$D = \rho \cdot \hat{D}$$

由此得

$$D = \prod_p \left[1 - \frac{1}{2p^2 + k} \left(1 - \frac{1}{p} \right) \right] \hat{D} = \rho \hat{D} \tag{3.8}$$

相应地

$$\rho = \prod_p \left[1 - \frac{1}{2p^2 + k} \left(1 - \frac{1}{p} \right) \right] \tag{3.9}$$

引式(2.11)

$$k_{\min} = 3 < k < k_{\max} = 4$$

式(3.8)对应表为

$$\prod_p \left[1 - \frac{1}{2p^2 + 3} \left(1 - \frac{1}{p} \right) \right] \hat{D} < D < \prod_p \left[1 - \frac{1}{2p^2 + 4} \left(1 - \frac{1}{p} \right) \right] \hat{D} \tag{3.10}$$

对 ρ 判敛, 取

$$\ln \left(\frac{1}{\rho} \right) = \sum_p \left[\ln \left(1 + \frac{(p-1)}{p(2p^2 + k) - (p-1)} \right) \right]$$

随 p 增大, 序列 $\left(1 + \frac{(p-1)}{p(2p^2 + k) - (p-1)} \right)$ 单调递减且有下界 1, $\ln \left(\frac{1}{\rho} \right)$ 收敛于一个大于 0 的常数, 因此, ρ 必收敛于一个小于 1 的常数。

实际上, 相邻 p^2 的等效作用域 $(0, p^2 + k]$ 相互自嵌套并不断扩大, 由此在累乘中形成降密因子的后向传递, 并通过累乘形成极限。

当 n 很大时, 对 $1 < p \leq \sqrt{n}$, 式(3.10)等效表为

$$\prod_{1 < p \leq \sqrt{n}} \left[1 - \frac{1}{2p^2 + 3} \left(1 - \frac{1}{p} \right) \right] \hat{D} < D < \prod_{1 < p \leq \sqrt{n}} \left[1 - \frac{1}{2p^2 + 4} \left(1 - \frac{1}{p} \right) \right] \hat{D} \tag{3.11}$$

即

$$\rho_{\min} \hat{D} < D < \rho_{\max} \hat{D}$$

仿真到 $1 < p < 1000000$ 表明, $\rho_{\min} = 0.88967765\dots$, $\rho_{\max} = 0.89489472\dots$,
即

$$D_{\min} = \rho_{\min} \hat{D} = \frac{\rho_{\min}}{\frac{e^c}{2} \ln n} = \frac{0.88967765\dots}{0.89053620\dots \ln n} = \frac{0.99903591\dots}{\ln n} \quad (3.12)$$

$$D_{\max} = \rho_{\max} \hat{D} = \frac{\rho_{\max}}{\frac{e^c}{2} \ln n} = \frac{0.89489472\dots}{0.89053620\dots \ln n} = \frac{1.00489426\dots}{\ln n} \quad (3.13)$$

由此确定当 $k = 3$ 和 $k = 4$ 时, D 的以 $\frac{1}{\ln n}$ 为中心的非常接近的下限和上限

$$\frac{0.99903591\dots}{\ln n} < D < \frac{1.00489426\dots}{\ln n} \quad (3.14)$$

对 $1 < p < p_n \leq \sqrt{n}$, 通过 p_n “筛余截首”, 确认区间 $(p_{n-1}^2, p_n^2 \leq n)$ 的素数, 猜想存在合适的满足式(2.10)

$$f(n; m) = \frac{1}{2m+k} \text{ 的 } k_{\min} = 3 < k < k_{\max} = 4, \text{ 同时满足 } \rho = \prod_{1 < p \leq \sqrt{n}} \left[1 - \frac{1}{2p^2+k} \left(1 - \frac{1}{p} \right) \right] = \frac{e^c}{2} = 0.89053620\dots,$$

由此使 n 邻域素数密度 $D = \frac{1}{\ln n}$ 成立。

猜想 3.1 证毕。

需要说明: 对由小到大的素数, 等效作用域由小变大, 相同的常数 k 对伽玛分布的截尾效应由大变小, 在等效作用域定义下以均匀分布表示伽玛分布的理想程度由小变大, 都反应了素数由小变大时粒度由大变小的事实。

$D = \frac{1}{\ln n}$ 与高斯(Gauss)建议式相同[3] [4]。

D 的不准确性源于: 在素数特征数 $\rho = \prod_{1 < p \leq \sqrt{n}} \left[1 - \frac{1}{2p^2+k} \left(1 - \frac{1}{p} \right) \right]$ 的渐近积累过程中, 与 \hat{D} 一样, 小素数粒度大、积累速率高。 ρ 与 \hat{D} 相乘后加权因子抵消, 但对小素数的不准确性以素数的固有特征保留下来。

正整数在自然数列的密度分布概念及“筛余截首”中正整数 p^2 对筛除剩余周期前端密度的影响, 是素数密度成立的关键。正整数的伽玛密度分布和等效均匀分布, 以及筛除 p^2 的降密作用, 看来都是合理的。核心的问题, 一是正整数在自然数列的伽玛密度分布 $f(n; m) = \frac{n^m}{m!e^n}$ 概念; 二是存在满足 $k_{\min} = 3 < k < k_{\max} = 4$ 的

常数 k , 将正整数 p^2 的伽玛密度分布 $f(n; p^2) = \frac{n^{(p^2)}}{(p^2)!e^n}$ 等效表为均匀分布 $f(n; p^2) = \frac{1}{2p^2+k}$, 常数 k 能满

足在区间 $(0, p^2]$ 与区间 $(p^2, 2p^2+k]$ 的积累密度无限接近, 即 $F(n; p^2) \Big|_0^{p^2} \sim F(n; p^2) \Big|_{p^2}^{2p^2+k}$, 且拖尾积累密度

$F(n; p^2) \Big|_{2p^2+k}^{\infty} \rightarrow 0$; 三是将这一等效均匀分布通过累乘表示连续筛除 p^2 的降密作用必然满足降密因子

$$\rho = \prod_p \left[1 - \frac{1}{2p^2+k} \left(1 - \frac{1}{p} \right) \right] = \frac{e^c}{2}, \text{ 式中, } \prod_p \text{ 为对全部素数的连乘, 由此导致 } n \text{ 邻域素数密度 } D = \frac{1}{\ln n}。$$

限于作者水平和高精度计算条件, 不能对此深入研究。真诚地希望数学界和各界批评指正, 开展专

题理论研究和高精度仿真计算,完成正整数密度分布的概念研究和猜想的证明。期待数学界的权威论证,向所有提出批评和做出研究的朋友、老师和专家表示衷心感谢。

推论 3.2 ($\hat{D} \rightarrow D$ 映射关系): 由式(3.2) $\hat{D} = \prod_{1 < p \leq \sqrt{n}} \left(1 - \frac{1}{p}\right) = \frac{1}{\frac{e^c}{2} \ln n}$ 及式(3.1) $D = \frac{1}{\ln n}$, 可直接建立由

p_n 筛除剩余周期 $\prod_{1 < p \leq p_n} p$ 内剩余平均密度 \hat{D} 到 n 邻域素数平均密度 D 的 1-1 映射关系[5]

$$\hat{D} = \prod_{1 < p \leq \sqrt{n}} \left(1 - \frac{1}{p}\right) \rightarrow D = \frac{1}{\ln n} \tag{3.15}$$

映射式(3.15)就是由线性筛除剩余通往非线性素数的桥梁,它直接源于埃氏筛法的“纯净筛除”、“周期重构”和“筛余截首”三步骤,素数密度直接源于埃氏筛法的素数生成机制。

推论 3.3 (筛除剩余周期内前端密度): 设有以 p_0 为基础连续 p_0 筛、 p 筛和 p' 筛, p 筛剩余在周期 $C = \prod_p p$ 内保持筛除概率 $d = \frac{1}{p}$ 和筛余密度 $\hat{D} = \prod_p \left(1 - \frac{1}{p}\right)$ 的同时,由于“筛余截首”即 p 筛以概率 1 固定筛除 p^2 的积累作用,使 p_0 筛余周期前端 (p_0^2, p'^2) 的密度低于周期 $C = \prod_{1 < p \leq p_0} p$ 内平均密度,这部分筛除 p^2 外全部为素数。筛除 p^2 的有效作用域是 $(0, 2p^2 + k)$, 而前端 $(p'^2 - p_0^2) \ll C = \prod_{1 < p \leq p_0} p$, p'^2 固定筛除点在 $(0, 2p^2 + k)$ 内非常靠近 p^2 的位置。这样,筛除 p^2 在确认 (p_0^2, p'^2) 内素数的同时,也为筛除 p'^2 确认区间 (p^2, p'^2) 内素数做好了实际准备。

推论 3.4 (素数分布的分形结构与混沌性态): 埃氏筛法的“纯净筛除”、“周期重构”和“筛余截首”,由简单机制生成了复杂性态。1) 在形成 \hat{D} 的过程中,形成周期内平均密度 $\hat{D} = \prod_{1 < p \leq \sqrt{n}} \left(1 - \frac{1}{p}\right)$ 是线性行为,但由于互质素数的“纯净筛除”和“周期重构”,已经产生了筛除剩余数的轨道不可预测性。2) 随着素数由小到大,“筛余截首”中连续 p^2 筛除的自相似的有效作用域形成 $(0, 2 \times 2^2 + k), (0, 2 \times 3^2 + k), \dots, (0, 2p^2 + k), \dots$ 无限自嵌套,并通过等效均匀分布的降密累乘,使素数密度和筛除剩余密度之比快速收敛到 $\rho = \frac{e^c}{2} = 0.89053620\dots$, 其间线性筛除剩余的周期由小依阶乘率变大,这代表了素数快速走向混沌的过程。

不过,由于 $p \rightarrow \infty$ 空间无限,这种混沌永远不会形成由混沌常数表征的倍周期分岔。3) 这就决定了,能用代数式给出表达的,除个别情况外,只能是素数的统计特性,包括密度、数量,以及对素数序列和初始正整数生成背景加不同约束条件的素数组的数量。一般而言,不方便以代数式对素数的轨道特性和结果进行确切描述。

4. 由素数密度到素数分布经典问题

文献[1]证明了 p 筛对 p_0 筛除剩余的筛除率和 p 筛对 p_0 筛除剩余中各筛除率的筛除率均为 $\frac{1}{p}$, 记为

$$d = \frac{1}{p} \tag{4.1}$$

文献[1]为确定不超过正整数 x 的素数数量,将素数密度 $D = \frac{1}{\ln n}$ 进行由代表最小素数的正整数 2 到最大正整数 x 的积分 $\pi(x) = \int_2^x \frac{dt}{\ln t}$, 并将积分展开为渐近级数

$$l_{sx} = \sum_{k=1}^{r-1} \left(\frac{(k-1)!x}{(\ln x)^k} - \frac{(k-1)!2}{(\ln 2)^k} \right) + \left(\left(\frac{(r-1)!x}{(\ln x)^r} - \frac{(r-1)!2}{(\ln 2)^r} \right) (1+o(1)) \right) \quad (4.2)$$

式中 r 为渐近级数的最佳截断

$$r = \left\lceil \frac{\ln x - \ln 2}{\ln \ln x - \ln \ln 2} \right\rceil - 1 \quad (4.3)$$

从而得出了与素数定理一致的更加精细的表示。

借助筛除概率式(4.1)、素数密度式(3.1)、 $\hat{D} \rightarrow D$ 映射式(3.15)和渐近级数概念,对包括孪生素数猜想和哥德巴赫猜想在内的各类约束素数组,如孪生素数猜想、多孪生素数猜想、偶数哥德巴赫猜想、素数差哥德巴赫猜想、 $ap \pm bq = n$ 哥德巴赫猜想等,可将相应约束条件嵌入埃氏筛法的递推步骤,通过素数组生成背景分析和筛除作用分析,在埃氏筛法对筛除剩余数的筛除率的基础上,确定对筛除剩余数组的筛除率,进而借助简单的数学变换和 $\hat{D} \rightarrow D$ 映射关系确定约束素数组密度和数量,并以相应不同形式对数积分的渐近级数表示。

作者在中文杂志《前沿科学》2016年第4期发表的论文《素数分布研究新思路的若干应用》[5]中,研究了这些问题并得出了结论,多数结论与历史上已形成的猜想一致。这说明各类素数分布问题均可纳入基于素数生成与筛除剩余分布演化机制的统一理论框架,也印证了作者通过建模与递推,解析埃氏筛法,研究素数分布的信念[1]: 1) 机制决定性质,素数分布的性质与特点必然植根于埃氏筛法这个“素数生成系统”中。2) 不能静态的就素数论素数,而要从源头上解析埃氏筛法,在动态中考察素数的生成与演化。3) 必须“庖公解牛”,“于关节处切割自然”[6],对素数生成过程和步骤进行客观的准确的切割。4) 各类素数分布问题的基础机制相同,必然存在内在的同一性,因此要整体地综合地相关地而不是孤立地研究这些问题。

致 谢

青年高级工程师王伟华博士、汤宏伟、王远和李军博士完成了初稿英译和仿真计算,崔雁巍完成了文档整理和校对,在此表示衷心感谢。衷心感谢爱妻邢纪荣和家人的理解、支持和关爱。

参考文献

- [1] 崔蕴华. 素数分布研究的一种新思路[J]. 前沿科学, 2016, 10(2): 12-32.
- [2] [美] A.M.穆德, F.A.格雷比尔, 著. 统计学导论[M]. 史定华, 译. 北京: 科学出版社, 1978.
- [3] 王元. 论哥德巴赫猜想[M]. 济南: 山东教育出版社, 1999.
- [4] Rosen, K.H. (2004) Elementary Number Theory and Its Applications. 4th Edition.
- [5] 崔蕴华. 素数分布研究新思路的若干应用[J]. 前沿科学, 2016, 10(4): 26-45.
- [6] 欧阳莹之. 复杂系统理论基础[M]. 上海: 上海科技教育出版社, 2002.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2160-7583，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：pm@hanspub.org