Additive walks on $\mathbb{N}$ and Proof of Twin- and d-Primes Conjecture

Gregory M. Sobko

Abstract

An additive model of random walks on set of natural numbers is applied to analyze probability distribution of gaps, that is differences $d = p' - p$, between consecutive prime numbers p' and p. The well known fact is that gaps between consecutive primes can be as small as 2 (for twin primes) and arbitrary large. This work is concerns with sets of primes $D\mathbb{P}_d$ with gaps d (called d-primes), where d is any even number.
For $D\mathbb{P}_2$ we have set of twin primes, with unproved conjecture that $D\mathbb{P}_2$ is infinite set.
We provide some statistical analysis for frequency distribution of d-primes.
The main result of this work is the proof that $D\mathbb{P}_d$ is infinite set for every even d.
The proof is based on modified Cramér's probabilistic model for distribution of prime numbers. This method has been discussed in detail in the author's previous publication[1].

Consider an additive rule to generate stochastic or deterministic sequences of positive integers:

$$\begin{aligned} &v_0 = 0, v_{k+1} = v_k + \Delta v_{k+1} \\ &\text{where } \Delta v_{k+1} = \begin{cases} p_{k+1} \text{ if } v_{k+1} \in \mathbb{P} \\ 0, \text{ otherwise} \end{cases} \text{ for } k = 0,1,2,3,\ldots \end{aligned} \tag{1}$$

This approach leads to 'additive model' of random walks on $\mathbb{N}$ in the study of prime numbers distribution. Though the sequence $\{v_k\}_{k\in\mathbb{N}}$ generated recurrently, is deterministic, each step of the 'walk' (1) can result either in $\Delta v_{k+1} \in \mathbb{P}$, or in 0 (if Δv_{k+1} is a composite number). Differences ('gaps') $\Delta v_{k+1} = p_{k+1} - p_k$ between consecutive two primes look very sporadic and hard to predict.

It is well known that gaps between two consecutive number p and $p+2$ s p_{k+1} and $p_k \geq 3$ can be as small as 2 (for twin primes) or arbitrary big. Indeed, in the sequence of $n-1$ consecutive integers $\{n!+k | 2 \leq k \leq n\}$ each integer $n!+k$ is divisible by k, and therefore this sequence does not include prime numbers. This means that there exist consecutive primes p_i and p_{i+1} such that $p_i < n!+2$ and $p_{i+1} > n!+n$, which implies that $\Delta p_{i+1} = p_{i+1} - p_i \geq n$.

The next definition is a generalization of the notion of twin primes.

Definition 1.

We call numbers $p < p'$ *consecutive* if there is no prime q between them (that is there is no prime q such that $p < q < p'$). A prime number p we call d-*prime* if there exist consecutive primes p, p' such that $p' = p + d$.

Notice that $\Delta p' = p' - p = d$ for d-prime p.

For example, p is 2-prime if and only if p and p' are twin primes, since for primes $p \geq 3$ twin primes are automatically consecutive.

Let us denote $D\mathbb{P}_d = \{ p | p \text{ and } p + d \text{ are consecutive primes} \}$ the set of d-primes.

For example, $D\mathbb{P}_1 = \{2\}$; the set of twin primes is $D\mathbb{P}_2 = \{3,5,11,17,29,41,\ldots\}$.

One of the famous unproved conjectures is that the set $D\mathbb{P}_2$ is infinite.

Table 6.1. d-primes for $d = 2,4,6$ among all primes $p < 200$

$D\mathbb{P}_2$	3 5 11 17 29 41 59 71 101 107 137 149 179 191 197
$D\mathbb{P}_4$	3 7 13 19 37 43 67 79 97 103 109 127 163 193
$D\mathbb{P}_6$	23 31 47 53 61 73 83 131 151 157 167 173

Lemma 1.

$\{D\mathbb{P}_d\}_{d \in 2\cdot\mathbb{N}}$ makes a partition of the set of primes $\mathbb{P}$.

Proof.

Notice that $D\mathbb{P}_d = \varnothing$ for all odd $d > 1$. Obviously, $D\mathbb{P}_d \cap D\mathbb{P}_{d'}$ for all $d \neq d'$.

Then, $D\mathbb{P}_1 \cup \left[\bigcup_{d \in 2\cdot\mathbb{N}} D\mathbb{P}_d \right] = \mathbb{P}$, where $2 \cdot \mathbb{N}$ is set of all even natural numbers.

This implies that any prime number p is a d-prime for an appropriate d.

Indeed, due to the Euclid Theorem, there are infinitely many prime numbers, Therefore, for any prime p there exist the next (that is consecutive) prime p', so that $p \in D\mathbb{P}_d$ where $d = p' - p$.

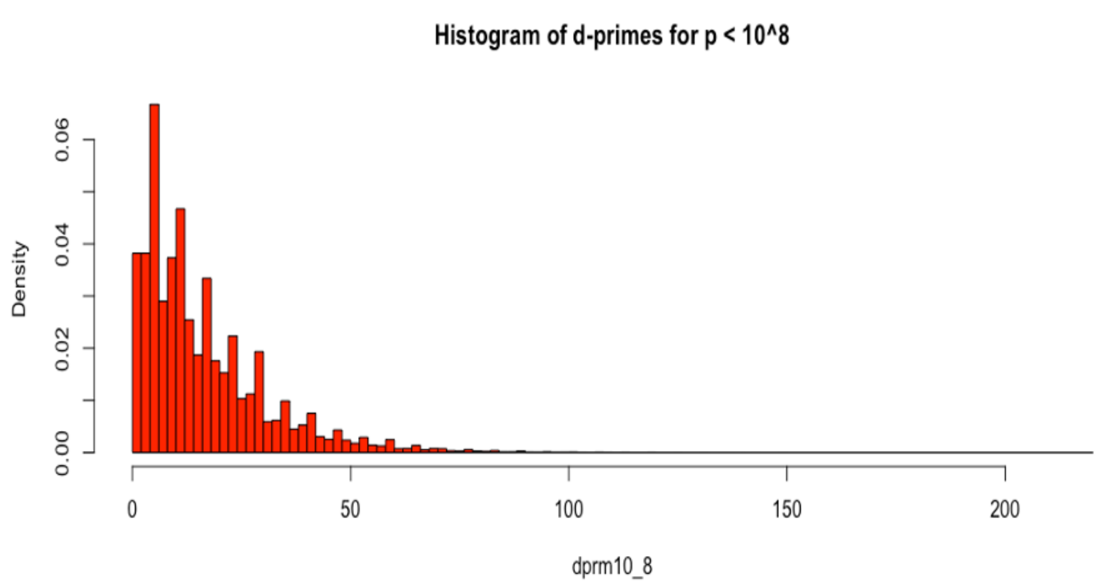
Q.E.D.

The first *conjecture* is that $D\mathbb{P}_d \neq \varnothing$ for all even values of $d \geq 2$.

The second conjecture is that every $D\mathbb{P}_d$ is an infinite set for all even values of $d \geq 2$.

Remark1.

For each even number d consider a partition of $\mathbb{N}$ by a finite number of congruence classes $\{C_{dr} | 0 \leq r \leq d-1\}$. Then, the intersection of the partition sets with $\mathbb{P}$ $\{C_{dp} \cap \mathbb{P} | 1 < r < d-1\}$ makes a partition of the infinite sets of all primes , such that at least one of classes C_{dr} must contain infinitely many prime numbers. Primes populate sets $D\mathbb{P}_{d,N} = D\mathbb{P}_d \cap [2,N]$ not evenly as illustrated by the histogram below for $N = 10^9$.



Computer calculations show so far that the most frequent value of consecutive primes gaps is $d = 6$.
According to the Prime Number Theorem, the counting function of primes on $\mathbb{N}$ is given by the asymptotic formula:

$$\pi(x) = \sum_{p \in [2,x] \cap \mathbb{P}} 1 \sim Li(x) = \int_2^x \frac{dt}{\ln t}$$

This leads to the heuristic assumption about the probability of prime distribution on $\mathbb{N}$. According to the Cramér's model, occurrences of primes in $\mathbb{N}$ are controlled by the sequence of independent variables $\{v_k\}_{k\in\mathbb{N}}$, $v_k = k$,

and associated independent Bernoulli variables $\xi_k = \begin{cases} 1 \text{ if } \ v_k = k \in \mathbb{P} \\ 0, \text{ otherwise} \end{cases}$

such that

$$P\{\xi_k = 1\} = \frac{1}{\ln k}, P\{\xi_k = 0\} = 1 - \frac{1}{\ln k} \text{ for all } k \geq 3 \tag{3}$$

As we know, the sequence of primes $\{p_n\}_{n\in\mathbb{N}}$ is deterministic and is recurrently controlled by the corresponding vector of residuals

$\vec{r}_n = \left(r_1, r_2, \ldots, r_{\pi(\sqrt{n})}\right)$, where $r_i = \text{mod}(n, p_i), i = 1, 2, \ldots, \pi(\sqrt{n})$.

The assumption that terms of the sequence $\{\xi_k\}_{k\in\mathbb{N}}$ in Cramér's model are independent random variables is justified by the statement that this sequence is asymptotically independent, as proved in [1].
Validity of the choice of probabilities in Cramér's model is supported by the proof given in [1] that as $k \to \infty$:

$P\left\{\xi_k = 1\right\} \sim \frac{c}{\ln k}$, where $c = \frac{2}{e^{\gamma}}$ *and* $\gamma = \lim\left(\sum_{k=1}^{n} \frac{1}{k} - \ln n\right) \approx 0.577215664$

is Euler's constant

Denote $\pi_d(x) = \sum_{[2,x]\cap D\mathbb{P}_d} 1 = \sum_{p\in D\mathbb{P}_d} I_{[2,x]}(p)$ number d-primes of in the interval

$[2,x]$. Given prime number p, the corresponding vector of residuals

$\vec{r}_n = \left(r_1, r_2, \ldots, r_{\pi(\sqrt{n})}\right)$ must have all non-zero components.

One of quite reasonable questions is how frequently d-primes may occur among all prime numbers. We can evaluate the empirical probability of d-primes by the relative frequency $P\left\{v \in D\mathbb{P}_d \cap [2,x]\right\} \approx \frac{\pi_d(x)}{\pi(x)}$.

Denote $\xi_d(n) = \begin{cases} 1 \text{ if } n \in D\mathbb{P}_d \\ 0, \text{ otherwise} \end{cases}$. Then, $\pi_d(x) = \sum_{n\le x} \xi_d(n)$.

We have ‘randomization’ of $\xi(n)$ in the form. $\xi(v) = \begin{cases} 1 \text{ if } v = p \in D\mathbb{P}_d \\ 0, \text{ otherwise} \end{cases}$.

Assuming the Cramer’s assumption of independence of consecutive primes, we have:

$$
\begin{aligned}
&P\left\{\xi_d(v) = 1\right\} = P\left\{v \text{ and } v + d \text{ are consecutive primes}\right\} \\
&= P\left\{v \text{ and } v + d \text{ are prime numbers with no primes in the open interval } (v, v+d)\right\} \\
&= P\left\{v \in \mathbb{P}\right\} \cdot P\left\{\bigcap_{i=1}^{d-1} \left\{(v+i) \notin \mathbb{P}\right\}\right\} \cdot P\left\{(v+d) \in \mathbb{P}\right\}
\end{aligned}
$$

Then, $P\left\{\bigcap_{i=1}^{d-1}\{(\nu+i)\notin\mathbb{P}\}\right\}=\prod_{i=1}^{d-1}\left[1-P\{(\nu+i)\in\mathbb{P}\}\right]$. Following the Cramér's model

assumption, $P\{\nu=k\in\mathbb{P}\}=\frac{1}{\ln k}$, we obtain:

$$P\left\{\xi_d(\nu)=1\middle|\nu=k\right\}=\frac{1}{(\ln k)\cdot(\ln(k+d))}\cdot\prod_{i=1}^{d-1}\left(1-\frac{1}{\ln(k+i)}\right)=\Psi(k,d)$$

Denoting $\phi(k,d)=\prod_{i=1}^{d-1}\left(1-\frac{1}{\ln(k+i)}\right)$, we write the function $\Psi(k,d)$ as

$$\Psi(k,d)=\frac{\phi(k,d)}{\ln(k)\cdot\ln(k+d)}$$

Thus, mathematical expectation and variance of $\xi_d(\nu)$ given $\nu=k$ can be expressed as $E\left\{\xi_d(\nu)\middle|\nu=k\right\}=\Psi(k,d), Var\left\{\xi_d(\nu)\middle|\nu=k\right\}=\Psi(k,d)\cdot(1-\Psi(k,d)$.

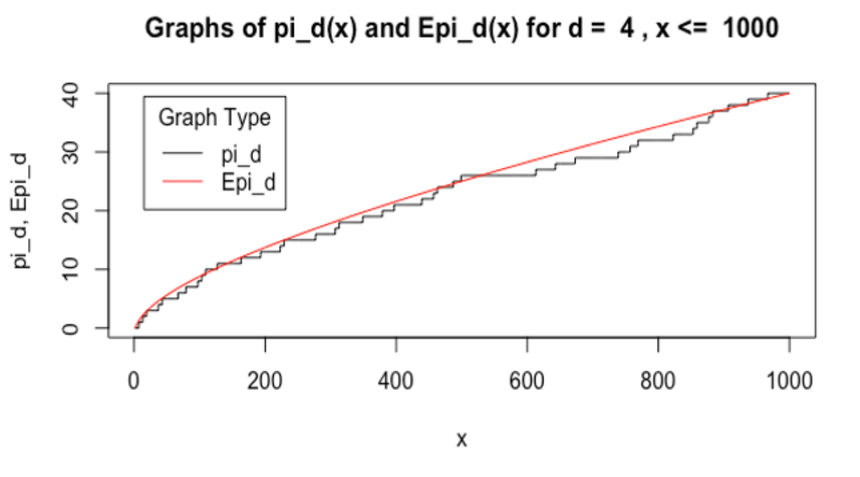
This implies:

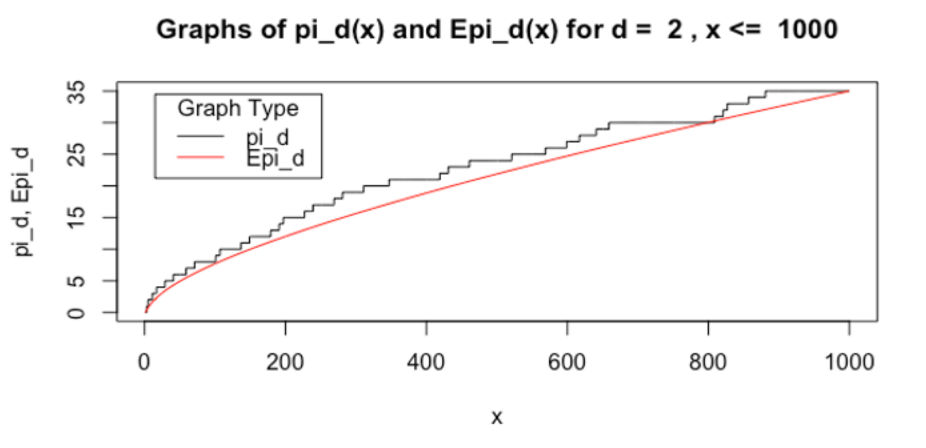
$$\begin{aligned}E\{\pi_d(x)\}&=\sum_{k\le x}E\left\{\xi_d(\nu)\middle|\nu=k\right\}=\sum_{k\le x}\Psi(k,d)\\ Var\{\pi_d(x)\}&=\sum_{k\le x}Var\left\{\xi_d(\nu)\middle|\nu=k\right\}=\sum_{k\le x}\Psi(k,d)\cdot(1-\Psi(k,d))\end{aligned}\tag{2}$$

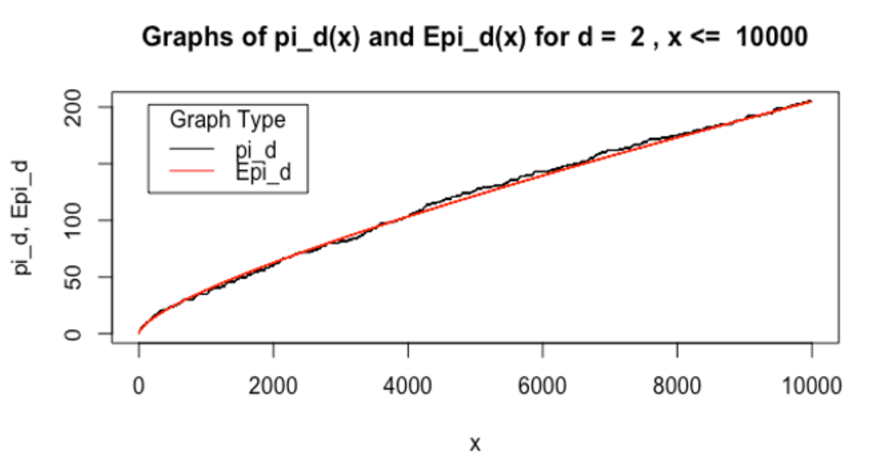
Using (2), we can approximate the mathematical expectation and variance of in the integral form:

$$\begin{aligned}E\{\pi_d(x)\}&=\sum_{k\le x}\Psi(k,d)\sim\int_2^x\frac{\phi(t,d)}{(\ln t)\cdot(\ln(t+d))}dt\\ Var\{\pi_d(x)\}&\sim\int_2^x\frac{\phi(t,d)}{(\ln t)\cdot(\ln(t+d))}\cdot\left(1-\frac{\phi(t,d)}{(\ln t)\cdot(\ln(t+d))}\right)dt\end{aligned}\tag{3}$$

Comparison of $\pi_d(x)$ distribution with its mathematical expectation $E\{\pi_d(x)\}$ is given in the pictures below computed for $d=2,4$ and $x=1000,\ x=10{,}000$


Graphs of pi_d(x) and Epi_d(x) for d = 4 , x <= 1000
Graph Type
pi_d
Epi_d
pi_d, Epi_d
0
10
20
30
40
0
200
400
600
800
1000
x


Graphs of pi_d(x) and Epi_d(x) for d = 2 , x <= 1000
Graph Type
pi_d
Epi_d
pi_d, Epi_d
0
5
15
25
35
0
200
400
600
800
1000
x


Graphs of pi_d(x) and Epi_d(x) for d = 2 , x <= 10000
Graph Type
pi_d
Epi_d
pi_d, Epi_d
0
50
100
200
0
2000
4000
6000
8000
10000
x

Theorem 6.1

For each even value of $d \geq 2$ there are infinitely many consecutive prime numbers with gap equal to d, so that every $D\mathbb{P}_d$ is an infinite set for all even values of $d \geq 2$.

Proof.

This statement is proved by using the equivalence:

$$E\{\pi_d(x)\} = \sum_{k \leq x} \Psi(k,d) \sim F(x) = \int_2^x \frac{\psi(t,d)}{(\ln t)\cdot(\ln(t+d))} dt \text{ as } x \to \infty$$

Indeed, if we assume that for some even $d \geq 2$ there exists $x_{\max}$ such that $\pi_d(x) = \pi_d(x_{\max})$ for all $x > x_{\max}$, then $\pi_d(x)$ becomes constant for sufficiently large values of x . But this contradicts the above equivalence since function $F(x)$ is strictly increasing for all $x > 2$ because it has positive derivative

$F'(x) = \dfrac{\psi(x,d)}{\ln(x)\cdot\ln(x+d)} > 0$ for all even $d \geq 2$ and $x > 2$.

Q.E.D.

REFERENCES

1. Gregory M. Sobko, Probabilistic Approach to some Additive and Multiplicative Problems of Number Theory, https://vixra.org/abs/2110.0022.
2. Granville, A., Harald Cramér and the distribution of prime numbers, Scandinavian Actuarial Journal, 1:12-28. 1995]
3. Cramér Harald, On the order of magnitude of the difference between consecutive prime numbers, Acta Arithmetica, 2: 23-46.1936.
4. Mark Kac, *Statistical Independence in Probability, Analysis and Number Theory*, The Mathematical Association of America, John Wiley and Sons, Inc., 1972.
5. Tom M. Apostol, Introduction to Analytic Number Theory, Springer, 2000.
6. Song Y. Yan, *Number Theory for Computing*, Springer, Springer-Verlag, 2000.
7. A.N. Shiryaev, *Probability*, 2nd edition, Springer, 1996.
8. V.S. Varadarajan, Euler Trough Time: A new Look at Old Themes, AMS, 2006.
9. Gérald Tenenbaum and Michael Mendès , *The Prime Numbers and* Their Distribution, AMS, 2000.
10. Michel Loèv, Probability Theory, Third Edition. Dover Publications, Inc., Mineola, New York, 2017.
11. Richard K. Guy, Unsolved Problems in Number Theory. Second Edition. Springer-Verlag, New York, 1994.

2010 Mathematics Subject Classification
Primary 11Nxx, 11N05, 11A41,11A07, 11A25, 11A51,
11N60, 11N37, 11P32, 11Mxx, 11M06, 11Y05;
Secondary: 11N36, 11Y16, 60-xx, 60Fxx, 60J65, 60G50,
60Bxx, 60B15

email: gbeholder@gmail.com (Gregory Sobko)

About the author

Gregory M. Sobko is a retired professor of mathematics at National University, San Diego, and at UC San Diego Extension. He obtained his MS degree in 1967 and PhD in Mathematics in 1973 from Moscow State University, Russia.
His works are in Probability Theory (Limit Theorems for random walks on Lie groups and differentiable manifolds) and applications (Random Fields, Stochastic Differential Equations) and in Number Theory.