Diffusion Approximations for Counting Number of Primes

Gregory M. Sobko

Abstract

Quite deterministic nature of prime numbers, due to the complexity of the recurrent generating algorithms, is mimicking 'randomness' and stimulates to apply some of probabilistic instruments to analyze number-theoretic problems. The key issue in the probabilistic analysis in a number-theoretic framework remains an enigmatic connection between deterministic nature of integer sequences related to prime numbers and their apparent complicated ('unpredictable' or 'chaotic') behavior interpreted as 'randomness'. We derive multiplicative and additive models with recurrent equations for generating sequences of prime numbers based on the reduced Sieve of Eratosthenes Algorithm and analyze their asymptotic behavior with the help of Riemann Zeta probability distribution. This allows to interpret such sequences as realizations of random walks on set $\mathbb{N}$ of natural numbers and on multiplicative semigroups $S(\mathbb{P})$ generated by set of prime numbers $\mathbb{P}$, representing paths of stochastic dynamical systems. We analyze in this work an additive continuous-time probabilistic model of counting function of primes $\pi(n)$ in terms of diffusion approximation of non-Markov random walks. We assume that 'updating' terms η in the recurrent equation $\pi(n(k+1)) - \pi(n(k)) = \eta(n(k+1)$ follow Zeta probability distribution and calculate infinitesimal characteristics of the random walk, which approximate coefficients of the corresponding stochastic differential equation. Computer modeling illustrates graphically an impressive fitting of trajectories for the original counting function, the calculated trend function, and the Brownian approximation.

"Using randomness to study certainty may seem somewhat surprising It is, however, one of the deepest contributions of our century to mathematics in general and to the theory of numbers in particular." (Gérald Tenenbaum, Michel Mendès France, *The Prime Numbers and Their Distribution*. AMS, 2000)

In this paper we consider the sequence $\{\pi(n)\}_{n\in\mathbb{N}}$ as a realization of a random walks $\{\pi(n,\omega)\}_{n\in\mathbb{W}}$ generated by the recurrent equation

$$\pi(n_{k+1})-\pi(n_k)=\eta(n_{k+1}) \text{ where } \eta(n_k)=h\big(\min(\vec{r}(n_k)\big),\ n_k=\nu_k(\omega) \qquad (1)$$

Here $\{\nu_k\}_{k\in\mathbb{N}}$ are assumed to be random variables with Zeta probability distribution. Recall that to define a stochastic process $\xi(t,\omega)$ with a discrete(or a continuous set $\mathcal{X}$ of values we need to have a measurable space $(\mathcal{X},\mathcal{B})$, where often $\mathcal{X}\subseteq\mathbb{N}$ or $\mathcal{X}\subseteq\mathbb{R}^d$, a Borel σ-algebra $\mathcal{B}$ of subsets on $\mathcal{X}$, and a set T of parameters $t\in T$ such that for each $t\in T$, $\xi(t,\cdot):\Omega\to\mathcal{X}$ is a random variable on a probability space $(\Omega,\mathcal{F},P)$. Then, the family $\{\xi(t,\cdot)\}_{t\in T}$ of random variables is called a *stochastic process* in the *phase space* $(\mathcal{X},\mathcal{B})$. The parameter $t\in T$ is usually interpreted as 'discrete time' for a countable set $T\subseteq\mathbb{W}=\mathbb{N}\cup\{0\}$ or as 'continuous time' for the continuous interval $T=\big[t_0,t_f\big)\subseteq\mathbb{R}^+=\big[0,\infty\big)$.

Then, a $\xi(t,\omega)$ is called a *stochastic process* with a discrete or a continuous time, respectively. For any given elementary event $\omega\in\Omega$, a function $x(\cdot):T\to\mathcal{X}$ such that $x(t)=\xi(t,\omega)$ is called a *path* (or a *trajectory*) of the random process. Alternatively, a stochastic process can be defined as a collection of paths (*random* *elements)* $x(\cdot)=\xi(\cdot,\omega)$in a function space $\mathcal{X}^T=\{x(t)\,|\,t\in T\}$ where ω that identifies each path is an elementary event in probability space $(\Omega,\mathcal{F},P)$. Elements (or points) $x\in\mathcal{X}$ are called '*states*' of the process, and $\xi(t,\omega)$ itself is called a process with a discrete or continuous phase space $(\mathcal{X},\mathcal{B})$.

Following the historical traditions of the classical Probability Theory (and the development of Calculus, in general), we try to apply limit theorem approach to analyze behavior of infinite discrete random sequences in terms of continuous-time stochastic processes.

Let $\vec{p}(n)=(p_1,p_2,\ldots,p_k)$ be a vector of consecutive prime numbers such that $p_1=2,\ p_k\le n$ and $p_{k+1}>n$. Index k determines here the value of function $\pi(n)=k$ that is the number of primes less than or equal to n so that $\vec{p}(n)=\left(p_1,p_2,\ldots,p_{\pi(n)}\right)$. For each coordinate p_i of vector $\vec{p}(n)$ we determine the residual value $r_i=\mathrm{mod}(n,p_i),\ i=1,2,\ldots,\pi(n),$ and consider the corresponding vector of residuals $(r_1,r_2,\ldots,r_{\pi(n)})$. Notice that, due to the Sieve Algorithm, for an integer $n>2$ to be prime it is necessary and sufficient that all coordinates r_i $\left(1\le i\le\sqrt{\pi(n)}\right)$ of the 'reduced' vector of residuals $\vec{r}(n)=\left(r_1,r_2,\ldots r_{\sqrt{\pi(n)}}\right)$ be different from zero. Meanwhile, if a random integer ν follows Zeta distribution, then, the events that $\nu=n$ does not divide each of consecutive primes $p_1,p_2,\ldots,p_{\pi(\sqrt{n})}$ are independent and can be expressed as a condition:

$$\min\left\{r_i\,|\,1\le i\le\pi\left(\sqrt{n}\right)\right\}>0 \text{ or, equivalently, } \prod_{1\le i\le\pi(\sqrt{n})} r_i>0. \qquad (2)$$

By using the Heaviside function $h(x)=\begin{cases}1 & \text{if } x>0\\ 0 & \text{if } x\le 0\end{cases}$, we can write the recurrent equation for $\pi(n)$ in the form:

$$\pi(n+1) = \pi(n) + h\left(\min_{p\le\sqrt{n}}\left\{\mathrm{mod}(n,p) \mid p \in \mathbb{P}\right\}\right)$$

or, equivalently,

$$\pi(n+1) = \pi(n) + h\left(\min_{i\le\sqrt{n+1}}\left\{r_i \mid r_i = \mathrm{mod}(n,p_i)\right\}\right) = \pi(n) + h\left(\min(\vec{r}(n)\right) \qquad (3)$$

which controls the occurrence of prime numbers in the sequence of all integers $n \ge 3$: $h\left(\min(\vec{r}(n)\right) = 1$ if and only if n is a prime number and $h\left(\min(\vec{r}(n)\right) = 0$ otherwise.

Consider a stochastic process approximation of non-Markov random walks $\left\{\pi(n,\omega)\right\}_{n\in\mathbb{W}}$ such that $\pi(n,\omega) = \pi\left(n\right)$, with $\pi : \mathbb{N}\times\Omega \to \mathbb{N}\cup\left\{0\right\}$ restricted to the interval of discrete 'times' $N_{\min} = n_0 < n_1 < \cdots < n_K = N_{\max}$:

$$\left\{\pi^{\Delta}(t_k) = \pi(n_k,\omega) \mid N_{\min} \le n_k \le N_{\max}\right\}. \qquad (4)$$

Denote $\Delta = \left(0 = t_0 < t_1 < \ldots < t_K = 1\right)$ a partition of an interval $\left[0,1\right]$ into K subintervals, such that $\dfrac{K}{\ln\left(N_{\max}\right)} \to 0 \ as\ N_{\max} \to \infty$.

We can map the closed interval of real numbers $\left[N_{\min}, N_{\max}\right] \subset \mathbb{R}$ to the interval $\left[0,1\right] \subset \mathbb{R}$ with an increasing continuously differentiable function $\tau(x)$ such that $\tau(N_{\min}) = 0,\ \tau(N_{\max}) = 1$.

In the context of our study, a suitable choice of function τ takes the form:

$$\tau(x) = \frac{\displaystyle\int_{N_{\min}}^{x} \frac{dt}{\ln t}}{\displaystyle\int_{N_{\min}}^{N_{\max}} \frac{dt}{\ln t}} = \frac{Li(x) - Li(N_{\min})}{Li(N_{\max}) - Li(N_{\min})} \qquad (5)$$

where $Li(x)$ stands for the *Eulerian logarithmic integral* $Li(x) = \displaystyle\int_2^x \frac{dt}{\ln t}$.

Then, $t_k = \tau(n_k)$ and for τ^{-1} (the inverse of τ) we have $n_k = \tau^{-1}(t_k)\ \ (k = 1,2,\ldots,K)$.

Denote $\Delta t_k = t_k - t_{k-1}$. Assume that $N_{\min} \to \infty$ and for each choice of $N_{\min}$ a positive integer K can be taken such that $|\Delta| = \max_{1 \le k \le K} \Delta t_k \to 0$. Here a sequence of random variables $\pi^\Delta(t_k) = \pi(n_k)$ is interpreted as a path of a walking point $\pi^\Delta(t_k)$ that belongs to a measurable space $(\mathcal{X}_k, \mathcal{B}_k)$ at each 'instant of registration' t_k. Probability distribution on the probability space $(\Omega, \mathcal{F}, P)$ generated by the path space $(\mathcal{X}^\infty, \mathcal{F}^\infty) = \left(\prod_{k \in \mathbb{N}} \mathcal{X}_k, \otimes_{k \in \mathbb{N}} \mathcal{B}_k \right)$ of random walks $\{\pi(k, \omega)\}_{k \in \mathbb{W}}$ is determined by transition probabilities $P\left\{ \pi^\Delta(t_{k+1}) \in E \mid \pi^\Delta(\vec{t}_0^{\,k}) = \vec{x}_0^{\,k} \right\}$ where $\vec{t}_0^{\,k} = (t_0, t_1, \ldots, t_k)$, $\vec{x}_0^{\,k} = (x_0, x_1, \ldots, x_k) \in \mathcal{X}_0^k = \prod_{i=0}^{k} \mathcal{X}_i$, $x_i \in \mathcal{X}_i \ (i = 0, 1, \ldots)$, $E \in \mathcal{B}_{k+1}$. Existence and uniqueness of the probability path space $(\Omega, \mathcal{F}, P)$ follows from the theorem of Ionescu Tulcea [41]. Notice that $\pi : \mathbb{N} \to \mathbb{W} = \mathbb{N} \cup \{0\}$ and therefore, we set $\mathcal{X}_k = \mathbb{W}$ for all $k \in \mathbb{N}^*$. To prove the weak convergence of transition probabilities for the sequence of random walks to the diffusion process $\hat{\pi}(t)$ on the time interval $[0,1]$, consider so called *infinitesimal characteristics* of the random walks:

$$m^\Delta(t_k, \vec{x}_1^{\,k}) = \frac{1}{\Delta t_k} E\left\{ \Delta\pi(t_{k+1}) \mid \pi^\Delta(\vec{t}_1^{\,k}) = \vec{x}_1^{\,k} \right\},$$

$$\left[\sigma^\Delta(t_k, \vec{x}_1^{\,k}) \right]^2 = \frac{1}{2 \cdot \Delta t_k} E\left\{ \left[\Delta\pi(t_{k+1}) \right]^2 \mid \pi^\Delta(\vec{t}_1^{\,k}) = \vec{x}_1^{\,k} \right\} \tag{6}$$

$$g^\Delta\left(t_k, \vec{x}_1^{\,k}; \Gamma_{k+1} \right) = \frac{1}{\Delta t_k} E\left\{ I_{\Gamma_{k+1}}(\Delta\pi^\Delta(t_{k+1}) \mid \pi^\Delta(\vec{t}_1^{\,k}) = \vec{x}_1^{\,k} \right\}.$$

Here $\Delta\pi(t_{k+1}) = \pi^\Delta(t_{k+1}) - \pi^\Delta(t_k) = \eta^\Delta(t_{k+1}) = \pi(n_{k+1}) - \pi(n_k) = \eta(n_{k+1})$

$\Delta\pi(t_{k+1}) = \pi^\Delta(t_{k+1}) - \pi^\Delta(t_k) = \pi(n_{k+1}) - \pi(n_k) = \eta^\Delta(t_{k+1}) = \eta(n_{k+1})$, $\vec{t}_1^{\,k} = (t_1, t_2, \ldots, t_k)$,

$$\pi^{\Delta}(\vec{t}_1^{\,k}) = \left(\pi^{\Delta}(t_1), \pi^{\Delta}(t_2), \ldots, \pi^{\Delta}(t_k)\right) = \left(x_1, x_2, \ldots, x_k\right) = \vec{x}_1^{\,k}, \;\; I_{\Gamma}(x) = \begin{cases} 1 \text{ if } x \in \Gamma \\ 0 \text{ otherwise} \end{cases},$$

$\Gamma_{k+1} \subset \mathcal{X}_{k+1} \setminus \{x_k\} \in \mathcal{B}_{k+1}$.

By setting $n_k = n_0 + k$ for all $k = 0, 1, \ldots, K$, we have:

$$\Delta\pi^{\Delta}(t_{k+1}) = \pi(n_{k+1}) - \pi(n_k) = \eta(n_k + 1) = \eta^{\Delta}(t_{k+1})$$

$$\Delta\pi^{\Delta}(t_{k+1}) = \pi(n_{k+1}) - \pi(n_k) = \pi(n_k + 1) - \pi(n_k) = \eta(n_{k+1}) = \eta^{\Delta}(t_{k+1})$$

$$\Delta\pi^{\Delta}(t_{k+1}) = \pi(n_{k+1}) - \pi(n_k) = \pi(n_k + 1) - \pi(n_k) = \eta(n_k + 1) = \eta^{\Delta}(t_{k+1}),$$ where

$$P\left\{\eta^{\Delta}(t_{k+1}) = 1\right\} = r_{n_{k+1}} = \prod_{p \le \sqrt{n_{k+1}}} \left(1 - \frac{1}{p}\right),$$

$$P\left\{\eta^{\Delta}(t_{k+1}) = 0\right\} = 1 - r_{n_{k+1}}$$

We have then,

$$m^{\Delta}(t_k, \vec{x}_k) == \frac{1}{\Delta t_k} \cdot E\left\{\eta(n_{k+1})\right\} = \frac{1}{\Delta t_k} \cdot \prod_{p \le \sqrt{n_{k+1}}} \left(1 - \frac{1}{p}\right) \tag{7}$$

Similar, since $\eta(n_{k+1}) = \left[\eta(n_{k+1})\right]^2$, we have

$$\left[\sigma^{\Delta}(t_k, \vec{x}_k)\right]^2 = \frac{1}{\Delta t_k} \cdot E\left\{\left[\eta(n_{k+1})\right]^2\right\} = \frac{1}{\Delta t_k} \cdot \prod_{p \le \sqrt{n_{k+1}}} \left(1 - \frac{1}{p}\right) \tag{8}$$

By applying the first Merten's theorem to (7) and (8), we have

$$\begin{aligned} m^{\Delta}(t_k, x_{n_k}) &= \frac{1}{\Delta t_k} \cdot \frac{c}{\ln(n_k + 1)} \cdot \left[1 + O\left(\frac{1}{\ln(n_k + 1)}\right)\right] \\ \left[\sigma^{\Delta}(t_k, x_{n_k})\right]^2 &= \frac{1}{\Delta t_k} \cdot \frac{c}{\ln(n_k + 1)} \cdot \left[1 + O\left(\frac{1}{\ln(n_k + 1)}\right)\right] \end{aligned} \tag{9}$$

Lemma 1.

For any interval $[a, b]$ with integer a and b such that $0 < a < b$, we have

$$\left|\sum_{a < n \le b} \frac{1}{\ln n} - \int_a^b \frac{dt}{\ln t}\right| \le \int_a^b \frac{dt}{t \cdot (\ln t)^2} \le \frac{b - a}{a \cdot (\ln a)^2} \tag{10}$$

Proof.

Due to the *Euler's summation formula*, for positive integer numbers a and b and a function f with a continuous derivative f' on $[a,b]$, we have

$\sum_{a<n\le b} f(n) = \int_a^b f(t)\,dt + \int_a^b \left(t-[t]\right) f'(t)\,dt$, where $[t]$ denotes an integer part of t.

$$\left|\sum_{a<n\le b} \frac{1}{\ln n} - \int_a^b \frac{dt}{\ln t}\right| = \left|\int_a^b \left(t-[t]\right) f'(t)\,dt\right| \le \int_a^b \left|f'(t)\right| dt \le \frac{b-a}{a\cdot(\ln a)^2}$$

Q.E.D.

Consider the *Eulerian logarithmic integral* $Li(x) = \int_2^x \frac{dt}{\ln t}$ to evaluate $\sum_{i=n_k+1}^{n_{k+1}} r_{ki}$.

Lemma 2.

$$\sum_{i=n_k+1}^{n_{k+1}} r_{ki} = \int_{n_k}^{n_{k+1}} \frac{dt}{\ln t} = Li(n_{k+1}) - Li(n_k) + O\left(\frac{n_{k+1}}{\ln^2(n_{k+1})}\right) = \frac{n_{k+1}}{\ln n_{k+1}} - \frac{n_k}{\ln n_k} + O\left(\frac{n_{k+1}}{\ln^2 n_{k+1}}\right)$$

Proof.

We have

By using approximation [7]: $li(x) = \int_0^x \frac{dt}{\ln t} = \frac{x}{\ln x} + O\left(\frac{1}{\ln^2 x}\right)$,

we have: $Li(x) = li(x) - li(2)$, where $li(x) = \int_0^x \frac{dt}{\ln t} = \frac{x}{\ln x} + O\left(\frac{x}{\ln^2 x}\right)$.

This implies $Li(n_{k+1}) - Li(n_k) = \frac{n_{k+1}}{\ln n_{k+1}} - \frac{n_k}{\ln n_k} + O\left(\frac{n_{k+1}}{\ln^2 n_{k+1}}\right)$.

Q.E.D.

Consider now a diffusion process $\hat{\pi}(t)$ given by stochastic integral:

$$\hat{\pi}(t) = \int_0^t m(s)\,ds + \int_0^t \sigma(s)\,dw(s) \tag{11}$$

where $m(t)=\frac{c}{\ln\left(\tau^{-1}(t)\right)}, \sigma(t)=\frac{1}{2}\cdot m(t)\cdot\left(1-m(t)\right), c=\frac{2}{e^{\gamma}}, 0\le t\le 1; \tau^{-1}(t)=x$.

with the transition probability $u(t,x,A)=P\left\{\hat{\pi}(t)\in A\,|\,\hat{\pi}(t_0)=0\right\}$.

Here $w(t)$ is a process of Brownian motion on $0\le t\le 1$.

The semigroup of linear operators U_t is defined on the space of bounded measurable functions by $\left(U_t f\right)(x)=\int f(y)u(t,x,dy)$.

We have the *infinitesimal generator of the semigroup* U_t given by the formula:

$$\left(Lf\right)(x)=\lim_{\Delta t\to 0}\frac{\left(U_{t+\Delta t}f\right)(x)-f(x)}{\Delta t}.$$

On the set of twice continuously differentiable functions $C^2(\mathbb{R})$ the generator L takes a form of a differential operator $\left(Lf\right)(x)=m(x)\frac{\partial f}{\partial x}+\frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2}$.

The function $V(t,x)=\left(U_t f\right)(x)=E\left[Y(t)\,|\,Y(t_0)=x\right]=\int f(y)u(t,x,dy)$ satisfies the equation

$$\frac{\partial V}{\partial t}=\frac{1}{2}\sigma^2(x)\frac{\partial^2 V}{\partial x^2}+m(x)\frac{\partial V}{\partial x} \qquad (12)$$

with the initial condition $V(t_0,x)=f(x)$. Taking as an initial condition δ-function, we have $V(t,x)=u(t,x,y)$, called a *fundamental solution* to (12).

This means that the transition probability has a density $u(t,x,y)$, so that

$$P\left\{Y(t)\in A\,|\,Y_0(t)=x\right\}=\int_A u(t,x,y)\,dy.$$

By applying the generalized limit theorem [22, 23] about convergence of random walks $\pi^{\Delta}\left(t_k\right)$ as $\left|\Delta\right|=\max_{1\le k\le K}\Delta t_k\to 0, N_{\min}\to\infty$ to diffusion processes (6.9), we obtain an approximation of $\left\{\pi(n,\omega)\right\}_{n\in\mathbb{W}}$ in terms of diffusion processes, defined for expanding intervals $[N_{\min},N_{\max}]$ of approximation on $\mathbb{N}$.

Theorem 1.

Transition probabilities

$P\left\{\pi^{\Delta}(t_{k+1}) \in E \mid \pi^{\Delta}(t_k) = x_k, \pi^{\Delta}(t_{k-1}) = x_{k-1}, \ldots, \pi^{\Delta}(t_0) = x_0\right\}$, where $\vec{x}_k = (x_1, \ldots, x_k) \in \mathbb{N}^k$,

of the defined above non-Markov random walks $\left\{\pi^{\Delta}(t_k) \mid N_{\min} < k < N_{\max}\right\}$ *converge weakly* to the transition probabilities of the diffusion process $\hat{\pi}(t)$ given by the stochastic integral

$$\hat{\pi}(t) = \int_0^t \hat{m}(s)\, ds + \int_0^t \hat{\sigma}(s)\, dw(s), \tag{13}$$

where $\hat{m}(t) = \dfrac{c}{\ln\left(\tau^{-1}(t)\right)}, \hat{\sigma}(t) = \dfrac{1}{2} \cdot \hat{m}(t) \cdot \left(1 - \hat{m}(t)\right), c = \dfrac{2}{e^{\gamma}}, 0 \le t \le 1,$

$\tau^{-1}(t) = x,\ N_{\min} \le n \le N_{\max}, \tau(N_{\min}) = 0, \tau(N_{\max}) = 1,\ c = \dfrac{2}{e^{\gamma}} \approx 1.122918968$

with the Euler's constant $\gamma = \sum_{m \le n} \dfrac{1}{m} - \ln n + O\left(\dfrac{1}{n}\right), \gamma \approx 0.577215664$,

as $\left|\Delta\right| = \max_{1 \le k \le K} \Delta t_k \to 0,\ N_{\min} \to \infty$.

Proof.

Since $\sum_{k=1}^{K} \Delta t_k = 1$, to due to Lemma 3.1.1, we have:

$$\sum_{k=1}^{K} \frac{1}{\ln n_k} \le \frac{K}{\ln(N_{\min})} \cdot \left(1 + O\left(\frac{1}{\ln(N_{\min})}\right)\right) \to 0, \text{ while } \frac{K}{N_{\max}} \to 0$$

Then, formulas (3.1.7), due to the second Merten's theorem (the *Merten's formula*) [3, 19], imply:

$$\sum_{k=1}^{K}\left[\,|\, m^{\Delta}(t_k,\vec{x}_k)-m(t_k)\,|+|\left(\sigma^{\Delta}(t_k,\vec{x}_k)\right)^2-\left(\sigma(t_k,\vec{x}_k)\right)^2\,|\,\right]\cdot\Delta t_k$$

$$=2\cdot\sum_{k=1}^{K}\left[\left|\frac{1}{\ln(n_k)}-\frac{1}{\ln(n_k)}\cdot\Delta t_k+\frac{1}{\ln(n_k)}\cdot O\left(\frac{1}{\ln(n_k)}\right)\right|\right]\le 2\cdot\sum_{k=1}^{K}\left[\frac{1}{\ln(n_k)}\left(1-\Delta t_k+O\left(\frac{1}{\ln(n_k)}\right)\right)\right]$$

$$\le 2\cdot\max_{1\le k\le K}\left|1-\Delta t_k+O\left(\frac{1}{\ln(n_k)}\right)\right|\cdot\sum_{k=1}^{K}\frac{1}{\ln n_k}\to 0,\ \text{as } \mathrm{N}_{\min}\to\infty$$

For $g(k)=\pi(k+1)-\pi(k)\le 1$, we have $P\left\{\eta^{\Delta}(k)>1\right\}=0$ for all k, so that all conditions are satisfied to apply the limit theorems for random walks proved in [22 ,23, 24].

Q.E.D.

The figures below illustrate graphically the diffusion approximation of distribution of primes in terms of $\pi(n)$ on different intervals of the argument.

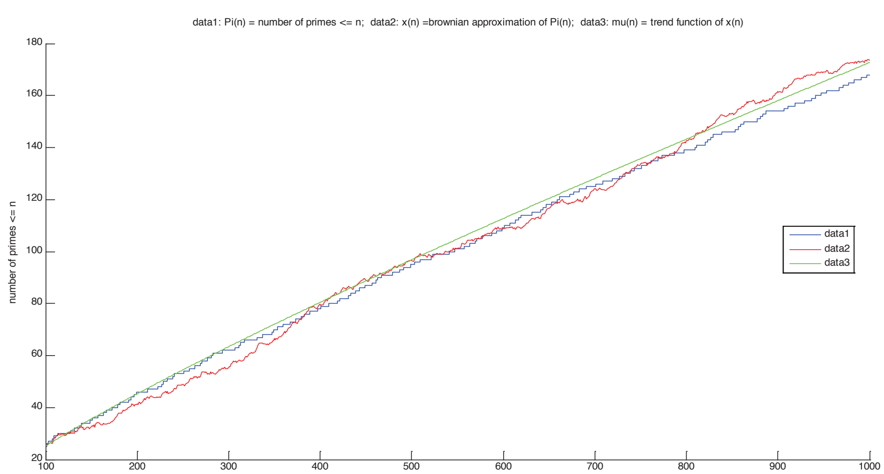
Legend for the graphs on the following figures:

data1: $\pi(n) =$ exact number of primes $\le$ n

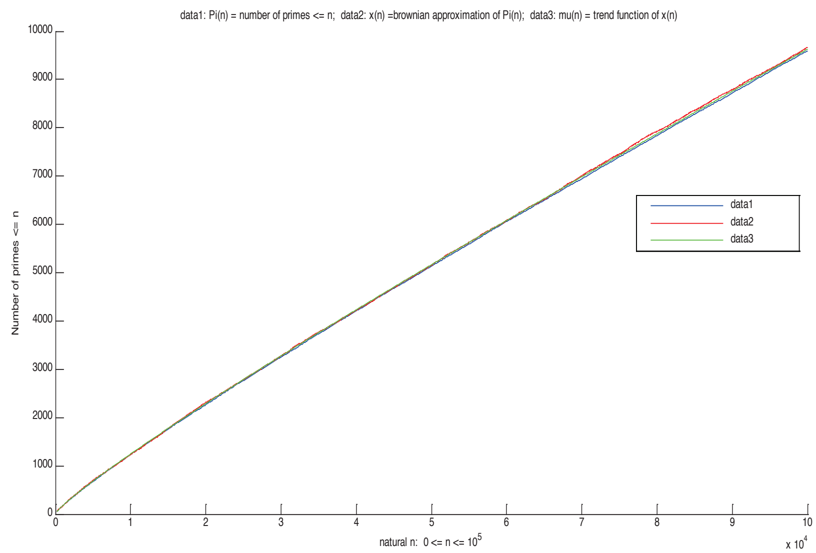
data2: Brownian approximation $X_n=\mu(n)+\xi_n\cdot\sigma(n)$ of $\pi(n)$

data 3: Trend function $\mu(n)$ of X_n

Approximation of $\pi(n)$ for $n: 100\le n\le 1000$



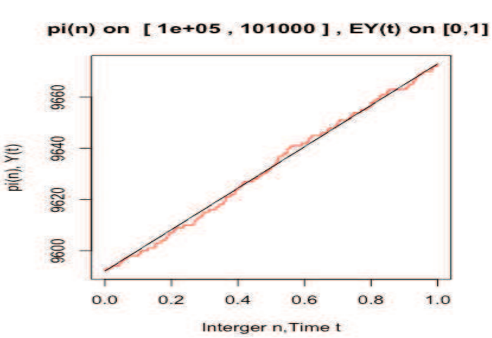
Approximation of $\pi(n)$ for $n: 0 \le n \le 10^5$

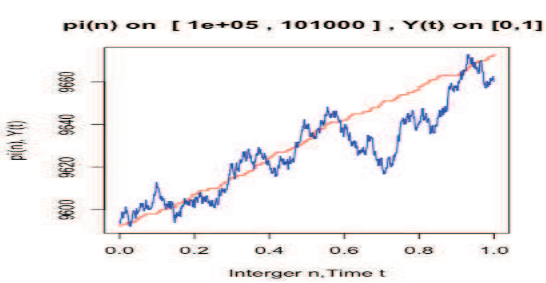


Approximation of $\pi(n)$ for $n: 5\times 10^5 \le n \le 10^6$

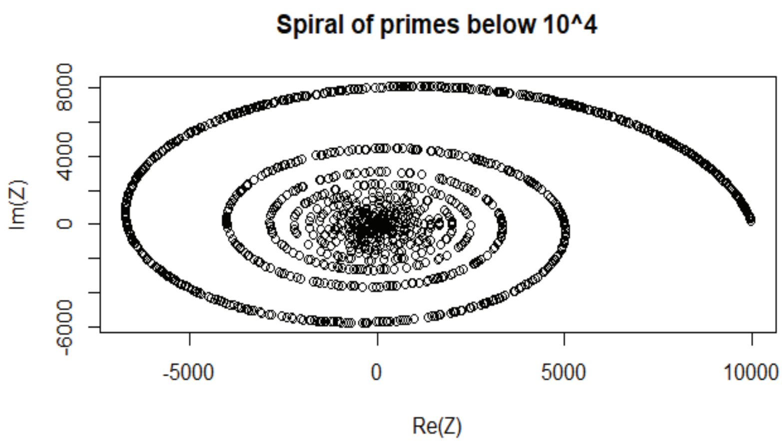
On figures below there are the graphs of paths described evolution of the 'walk' of a counts $\{\pi(n) \mid n \in \mathbb{N}\}$ of consecutive primes restricted to the intervals $[N_{\min}, N_{\max}] \subset \mathbb{N}$ and approximating diffusion processes :

$$Y(t) = \hat{\pi}^{\Delta}(t) \text{ and their expectations } \mathrm{E}Y(t) \text{ for } t \in [0,1] \subset \mathbb{R}.$$

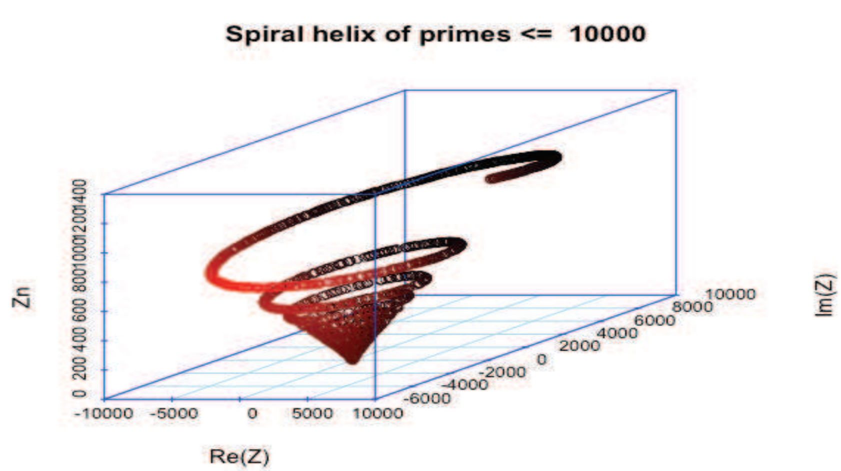




The sequence of vectors $(\vec{p}(n),\vec{r}(n))$, $(n=2,3,\ldots)$ created by consecutive n primes and the residual values $\vec{r}=\text{mod}(n,\vec{p})$, allows an interesting 3D presentation. In each pair $(\vec{p}(n),\vec{r}(n))$ vector of primes $\vec{p}(n)$ represents a 'radial' component, while the vector of residuals $\vec{r}(n)$, due to its natural periodicity, represents a 'circular' component.



Denote $z_k = p_k \cdot \exp\left(2\pi i \cdot \frac{r_k}{p_k - 1}\right)$, $r_k = \text{mod}(n, p_k)$, $(k=1,2,3,\ldots)$ - a sequence of complex numbers and the vector $\vec{z}(n)=(z_1,\ldots,z_n)$. Then for any $n>2$ vector $\vec{z}(n)$ takes a shape of a spiral helix as in the pictures below.



REFERENCES

1. Granville, A., *Harald Cramér and the distribution of prime numbers*, Scandinavian Actuarial Journal, 1:12-28. 1995]
2. Mark Kac, *Statistical Independence in Probability, Analysis and Number Theory*, The Mathematical Association of America, John Wiley and Sons, Inc.,1972.
3. Tom M. Apostol, *Introduction to Analytic Number Theory*, Springer, 2000.
4. Edward Beltrami, *What is Random? Chance and Order in Mathematics and Life.* Springer-Verlag New York, Inc. , 1999.
5. Song Y. Yan, *Number Theory for Computing*, Springer, Springer-Verlag, 2000.
6. A.N. Shiryaev, *Probability*, 2nd edition, Springer, 1996.
7. Gérald Tenenbaum, Introduction to Analytic and Probabilistic Number Theory, 3rd Edition, AMS, 2015.
8. Gérald Tenenbaum and Michael Mendès , *The Prime Numbers and Their Distribution,* AMS, 2000.
9. Gwo Dong Lin and Chin-Yuan Hu, *The Riemann zeta distribution*, Bernoulli 7(5), 2001, pp. 817-828.
10. Solomon W. Golomb, *A Class of Probability Distributions on the Integers*. Journal of Number Theory 2, 1970, p. 189-192.
11. Michel Loèv, Probability Theory, Third Edition. Dover Publications, Inc., Mineola, New York, 2017.
12. G.M. Sobko, *The First Diffusion Problem on Differentiable Manifolds*, Theory of Probability and its Applications, v. XVII, n.3, 1972, pp. 521 – 528.

13. G. M. Sobko, *A Diffusion Approximation of non-Markov Random Walks on Differentiable Manifolds*, Theory of Probability and its Applications, v. XVIII, n.1, 1973, pp. 41 - 53

14. Г.М. Собко, *Предельные теоремы для случайных блужданий на дифференцируемых многообразиях*. Механико-математический факультет, МГУ, 1973. (G. M. Sobko, *Limit Theorems for Random Walks on Differentiable Manifolds*. Dissertation. Department of Mathematics and Mechanics. Moscow State University, 1973, pp. 1-148.

15. А.Н. Колмогоров, *Основные понятия теории вероятностей*. Москва, “Наука”, 1974.

16. A.N. Shiryaev, *Probability*, 2nd edition, Springer, 1996.

17. S.R.S Varadhan, *Probability Theory*. Courant Institute of Mathematical Studies, American Mathematical Society. 2001.

18. S.R.S Varadhan, *Stochastic Processes*. Courant Institute of Mathematical Studies, American Mathematical Society. 2007.

19. Mark B. Villarino Depto, *Mertens' Proof of Mertens' Theorem. Depto de Matemática, Universidad de Costa Rica, 2060 San Jose',* Costa Rica April 28, 2005

20. Marek Wolf, *Random walk on the prime numbers*, Physica A 250 (1998). Institute of Theoretical Physics, University of Wroclaw, Poland.

21. Jacque Neveu, *Bases Mathématiques du Calcul des Probabilités*, Masson et Cie, Paris, 1964.

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), Number Theory.

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)