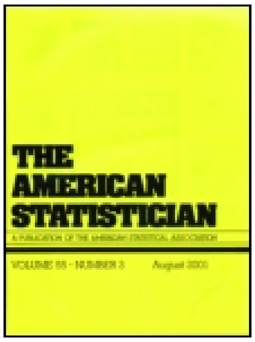

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utas20

Statistics of Primes (and Probably Twin Primes) Satisfy Taylor's Law from Ecology

Joel E. Cohen

To cite this article: Joel E. Cohen (2016) Statistics of Primes (and Probably Twin Primes) Satisfy Taylor's Law from Ecology, The American Statistician, 70:4, 399-404, DOI: 10.1080/00031305.2016.1173591

To link to this article: http://dx.doi.org/10.1080/00031305.2016.1173591

THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 4, 399–404
http://dx.doi.org/10.1080/00031305.2016.1173591

Taylor & Francis
Taylor & Francis Group

INTERDISCIPLINARY

Statistics of Primes (and Probably Twin Primes) Satisfy Taylor's Law from Ecology

Joel E. Cohen

The Rockefeller University and Columbia University, New York, NY, USA; Department of Statistics, Columbia University, New York, NY, USA; Department of Statistics, University of Chicago, Chicago, IL, USA

ASBTRACT
Taylor's law, which originated in ecology, states that, in sets of measurements of population density, the sample variance is approximately proportional to a power of the sample mean. Taylor's law has been verified for many species ranging from bacterial to human. Here, we show that the variance $V(x)$ and the mean $M(x)$ of the primes not exceeding a real number x obey Taylor's law asymptotically for large x. Specifically, $V(x) \sim (1/3)(M(x))^2$ as $x \to \infty$. This apparently new fact about primes shows that Taylor's law may arise in the absence of biological processes, and that patterns discovered in biological data can suggest novel questions in number theory. If the Hardy-Littlewood twin primes conjecture is true, then the identical Taylor's law holds also for twin primes. Taylor's law holds in both instances because the primes (and the twin primes, given the conjecture) not exceeding x are asymptotically uniformly distributed on the integers in $[2, x]$. Hence, asymptotically $M(x) \sim x/2$, $V(x) \sim x^2/12$. Higher-order moments of the primes (twin primes) not exceeding x satisfy a generalized Taylor's law. The 11,078,937 primes and 813,371 twin primes not exceeding 2×10^8 illustrate these results.

ARTICLE HISTORY
Received November 2015
Revised March 2016

KEYWORDS
Generalized Taylor's law; Power law; Prime number theorem; Twin prime conjecture; Uniform distribution; Variance function

1. Introduction

The empirical pattern that has come to be called Taylor's law (henceforth TL) originated from ecological studies of the abundance of insect and plant populations (Bliss 1941; Fracker and Brischle 1944; Hayman and Lowe 1961; Taylor 1961, 1986). TL states that, in sets of measurements of population density, the sample variance is usually nearly proportional to a power of the sample mean. TL has been verified empirically and modeled theoretically for population densities of many species ranging from bacterial to human. TL also has been confirmed and modeled for many other nonnegative quantities in fields beyond ecology (review by Eisler, Bartos, and Kertész 2008), such as agriculture, developmental biology, meteorology, genetics, cancer epidemiology, HIV/AIDS epidemiology, demography, computer science, stock market analysis, currency trading, and physics (where TL is called variously fluctuation scaling or "big" or "giant" or "large" number fluctuations). TL has been observed numerically in mathematical structures such as integer partitions and integer compositions (Xiao, Locey, and White 2015) and the absolute value of the Mertens function (Kendal and Jørgensen 2011). To our knowledge, TL has not previously been demonstrated, numerically or theoretically, for the mean and variance of the primes or twin primes.

Here, we show that the variance $V(x)$ and the mean $M(x)$ of the primes not exceeding a real number x obey Taylor's law asymptotically for large x. Specifically,

$$\begin{gathered} V(x) \sim a(M(x))^b \text{ or equivalently} \\ \lim_{x\to\infty} V(x)\,(M(x))^{-b} = a \\ \text{with } a = 1/3,\ b = 2. \end{gathered} \tag{1}$$

Taking the square root of both sides, the coefficient of variation of the primes not exceeding x, $cv(x) \equiv V(x)^{\frac{1}{2}}/M(x)$, therefore satisfies

$$\lim_{x\to\infty} cv(x) = \sqrt{a} = \sqrt{3}/3 \approx 0.5774. \tag{2}$$

Why might this matter? To biologists, this apparently new finding confirms again, as the two previous citations have shown in other cases, that TL may arise for reasons that have nothing to do with ecological processes such as birth, death, migration, or competition. To mathematicians, this finding confirms again, as physics has long done, the possible payoff of looking in mathematics for novel patterns suggested by biology. To statisticians, this finding confirms again the power of probabilistic thinking and statistical analysis in bridging diverse disciplines and providing essential tools of discovery.

Taylor's law holds because, as we shall show, as $x \to \infty$, the primes not exceeding a real number x are asymptotically uniformly distributed on the integers in $[2, x]$. Hence asymptotically as $x \to \infty$,

$$M(x) \sim \frac{x}{2},\ V(x) \sim \frac{x^2}{12}. \tag{3}$$

Higher-order moments of the primes not exceeding x satisfy a generalization (7) of Taylor's law to higher-order moments.

To illustrate, the primes not exceeding $x = 10$ are $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$. For $x = 10$, the number $\pi(x)$ of primes not exceeding x is $\pi(x) = 4$, the mean $M(x)$ of those 4 primes is $M(x) = 4.25 = (2 + 3 + 5 + 7)/4$, and the variance $V(x)$ of those 4 primes is $V(x) = 3.6875 = (2^2 + 3^2 + 5^2 + 7^2)/4 - 4.25^2$. This $V(x)$ is the exact variance (the average squared deviation of each prime from the mean), not the unbiased estimate

CONTACT Joel E. Cohen cohen@rockefeller.edu The Rockefeller University and Columbia University, New York, NY 10065.

of sample variance. The coefficient of variation of the 4 primes is $\sqrt{59}/17$ (≈ 0.4518).

A natural number p is defined to be a twin prime if and only if p is a prime and $p+2$ is a prime. For example, 3 and 5 are all the twin primes not exceeding 10. (By this definition, 7 is not a twin prime because $7+2=9$ is not a prime.) The Hardy-Littlewood twin prime conjecture (see Equation (6)) specifies a counting function for the twin primes not exceeding x. Conditional on the Hardy-Littlewood twin prime conjecture, the twin primes not exceeding x are asymptotically uniformly distributed on the integers in $[2, x]$ and, if $M_2(x),\ V_2(x),\ cv_2(x)$ are the mean, variance, and coefficient of variation of the twin primes not exceeding x, then (1), (2), and (3) remain valid with M replaced by M_2 and V replaced by V_2, and a generalized TL also holds asymptotically.

2. Definitions and Prior Results

If $f(x)$ and $g(x)$ are real-valued functions of a real x, then by definition $f(x)$ is asymptotic to $g(x)$ and we write $f(x) \sim g(x)$ if and only if, as $x \to \infty$, we have $f(x) \to \infty$ and $g(x) \to \infty$ and $f(x)/g(x) \to 1$. The relation "$\sim$" is transitive. We write $f(x) = O(g(x))$ if and only if, for some real $C > 0$ and all real x (possibly in a specified subset), $|f(x)| \le C|g(x)|$. The symbol "$\equiv$" means "is defined as."

We now define TL formally, in three variants: exact ($=$), approximate ($\approx$), and asymptotic ($\sim$). Suppose a nonnegative real-valued random variable $X(x)$ whose distribution depends on a parameter x has mean $M(x)$ and variance $V(x)$. Let $\mathbf{X}$ be the set (finite or infinite) of possible values of x. Then, by definition, the family $\{X(x)\}_{x\in\mathbf{X}}$ satisfies TL if and only if there exist two constants independent of x, namely, $a > 0$ and real b, such that $V(x) = a(M(x))^b$ exactly for all x, or $V(x) \approx a(M(x))^b$ approximately (with error term unstated but small), or $V(x) \sim a(M(x))^b$ asymptotically as $x \to \infty$.

By a natural extension (Giometto et al. 2015), a family $\{X(x)\}_{x\in\mathbf{X}}$ satisfies a generalized TL in three variants ($=$, $\approx$, $\sim$) if and only if there exists a set of real pairs (j, k) including $j = 1$, $k = 2$, and for each such pair (j, k) there exist real constants $a_{jk} > 0$ and b_{jk} independent of x such that $E[(X(x))^k]$ and $a_{jk}(E[(X(x))^j])^{b_{jk}}$ are equal exactly ($=$) or approximately ($\approx$) or asymptotically ($\sim$). TL is the special case $j = 1$, $k = 2$ of the generalized TL. Giometto et al. (2015) gave theoretical and ecological examples of a generalized TL.

Now we turn to primes and twin primes. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ be the natural numbers, $\mathbb{O} = \mathbb{N} \cup \{0\}$ be the nonnegative integers, $\mathbb{R}_+ = [0,\infty)$ and $\mathbb{R}_2 = [2, \infty)$. Let $\mathbb{P} = \{ p_1 = 2,\ p_2 = 3,\ p_3 = 5,\ p_4 = 7, \ldots\}$ be the primes and $\mathbb{P}_2 = \{3, 5, 11, 17, \ldots\} = \{p \in \mathbb{P} | p + 2 \in \mathbb{P}\}$ be the twin primes. For $x \in \mathbb{R}_+$, define $\mathbb{P}(x) \equiv \{p \in \mathbb{P} | p \le x\}$, $\mathbb{P}_2(x) \equiv \{p \in \mathbb{P}_2 | p \le x\}$. For any finite set S, define #S as the number of elements in S. By definition, $\pi(x) \equiv \#\mathbb{P}(x)$ is the number of primes not exceeding x, and $\pi_2(x) \equiv \#\mathbb{P}_2(x)$ is the number of twin primes not exceeding x. If $0 \le x < 2$, then $\pi(x) = \pi_2(x) = 0$.

For $x \in \mathbb{R}_2$, the logarithmic integrals are defined by

$$\mathrm{li}(x) \equiv \int_2^x \frac{dt}{\log t},\ \mathrm{li}_2(x) \equiv \int_2^x \frac{dt}{(\log t)^2}. \tag{4}$$

The prime number theorem (e.g., Crandall and Pomerance 2005; Montgomery and Vaughan 2007) states that as nonnegative real $x \to \infty$,

$$\pi(x) \sim \mathrm{li}(x). \tag{5}$$

The Hardy-Littlewood twin prime conjecture (e.g., Crandall and Pomerance 2005, p. 15, their (1.7); Sebah and Gourdon 2002, p. 5, their (2)) states that, for some constant $C_2 > 0$, as nonnegative real $x \to \infty$,

$$\pi_2(x) \sim 2C_2\mathrm{li}_2(x). \tag{6}$$

Approximately $C_2 \approx 0.6601618158\ldots$ (Sebah and Gourdon 2002, p. 5). Although $\mathrm{li}_2(x) \to \infty$ as $x \to \infty$, it has not yet been proved that $\pi_2(x) \to \infty$ as $x \to \infty$.

We review some elementary information about the uniform distribution. By definition, a real-valued scalar random variable X has the uniform distribution on $[0, x]$, where $x > 0$, if and only if, for every $r \in [0, 1]$, $\Pr\{X \le rx\} = r$. If X has the uniform distribution on $[0, x]$, then X has mean $E(X) = x/2$, second raw moment $E(X^2) = x^2/3$, variance $\mathrm{var}(X) = E(X^2) - (E(X))^2 = x^2/12$, and coefficient of variation $\sqrt{3}/3$. For $m \ge 0$, the uniform distribution on $[0, x]$ has mth raw moment $E(X^m) = x^m/(m+1)$ and mth central moment $E((X - x/2)^m) = 0$ if m is an odd integer (since the uniform distribution is symmetric about its mean $x/2$) and $E((X - x/2)^m) = x^m/c_m$ if m is an even integer, where $c_m = 2^m(m+1)$ (which is sequence A058962 in the Online Encyclopedia of Integer Sequences).

Given these moments, it follows that if X is uniform on $[0, x]$, then for all $x > 0$, TL holds exactly, that is, $V(x) = (\frac{1}{3})(M(x))^2$, and if $j > 0$, $k \ge 0$, then the generalized TL holds exactly, that is,

$$E(X^k) = \frac{(j+1)^{\frac{k}{j}}}{k+1}\,(E(X^j))^{\frac{k}{j}}. \tag{7}$$

Suppose Y has the uniform distribution on $[2, x]$, that is, for fixed $x > 2$ and every $r \in [0, 1]$, $\Pr\{Y - 2 \le r(x-2)\} = r$. If X has the uniform distribution on $[0, x]$, then, as $x \to \infty$, every moment of X asymptotically equals the corresponding moment of Y because $\Pr\{X \in (0, 2)\} \to 0$. For example, $E(Y) = 2 + (x-2)/2 = 1 + x/2 \sim E(X) = x/2$.

3. Analytical Results for Primes and Twin Primes

We now interpret the primes and twin primes not exceeding a real number $x \in \mathbb{R}_2$ as families of random variables. For every $x \in \mathbb{R}_2$ and for every $p \in \mathbb{P}(x)$, define $X(x)$ to be the random variable that takes the value $p \in \mathbb{P}(x)$ with probability $1/\pi(x)$. (Recall that $\pi(x)$ is the number of primes not exceeding x.) Then for any $r \in [0, 1]$, the cumulative distribution function of $X(x)$ is, by definition, $F_{X(x)}(rx) = \Pr\{X(x) \le rx\}$ and the mean and variance of $X(x)$ are $M(x) \equiv E(X(x)) = (p_1 + \cdots + p_{\pi(x)})/\pi(x)$ and $V(x) \equiv E(X(x)^2) - (E(X(x)))^2 = (p_1^2 + \cdots + p_{\pi(x)}^2)/\pi(x) - (M(x))^2$.

Similarly, for every $x \in \mathbb{R}_2$ and for every $p \in \mathbb{P}_2(x)$, define $X_2(x)$ to be the random variable that takes the value $p \in \mathbb{P}_2(x)$ with probability $1/\pi_2(x)$. Then for any $r \in [0, 1]$, the cumulative distribution function of $X_2(x)$ is defined to be

$F_{X_2(x)}(rx) = \Pr\{X_2(x) \le rx\}$ and the mean $M_2(x)$ and variance $V_2(x)$ of $X_2(x)$ are defined similarly.

Theorem 1. The primes not exceeding x are asymptotically uniform on $[2, x]$, that is, for any $r \in [0, 1]$, as $x \to \infty$, $F_{X(x)}(rx) \sim r$.

Proof. For any $x \in \mathbb{R}_2$, the number of primes not exceeding x is $\pi(x)$, by definition. Therefore, for any $r \in [0, 1]$, the number of primes not exceeding rx is $\pi(rx)$. Therefore the fraction of primes not exceeding x that do not exceed rx is

$$F_{X(x)}(rx) = \frac{\#\mathbb{P}(rx)}{\#\mathbb{P}(x)} = \frac{\pi(rx)}{\pi(x)}.$$

By the prime number theorem and Lemma A.1 (see the Appendix), as $x \to \infty$, $\pi(x) \sim \mathrm{li}(x) \sim x/\log x$. Therefore $\pi(rx) \sim (rx)/\log(rx)$ and

$$\frac{\pi(rx)}{\pi(x)} \sim \frac{\left(\frac{rx}{\log rx}\right)}{\left(\frac{x}{\log x}\right)} = \frac{r \log x}{\log x + \log r} \to r \text{ as } x \to \infty.$$

□

Theorem 2. If the Hardy-Littlewood twin prime conjecture (6) holds, the twin primes not exceeding x are asymptotically uniform on $[2, x]$.

Proof. For any $x \in \mathbb{R}_2$, the number of twin primes not exceeding x is $\pi_2(x)$, by definition. Therefore, for any $r \in [0, 1]$, the number of twin primes not exceeding rx is $\pi_2(rx)$. Therefore, the fraction of twin primes not exceeding x that do not exceed rx is $\frac{\#\mathbb{P}_2(rx)}{\#\mathbb{P}_2(x)} = \frac{\pi_2(rx)}{\pi_2(x)}$. If (6) holds, then, as $x \to \infty$, $\pi_2(x) \sim 2C_2 \mathrm{li}_2(x) \sim 2C_2 \frac{x}{(\log x)^2}$ by Lemma A.1 (the Appendix), and likewise $\pi_2(rx) \sim 2C_2 \frac{rx}{(\log rx)^2}$. Therefore,

$$\frac{\pi_2(rx)}{\pi_2(x)} \sim \frac{\left(\frac{rx}{(\log rx)^2}\right)}{\left(\frac{x}{(\log x)^2}\right)} = r\left(\frac{\log x}{\log x + \log r}\right)^2 \to r.$$

□

This proof remains valid regardless of the value of $C_2 > 0$.

Corollary 1. For any $x \in \mathbb{R}$, (1), (2), and (3) hold and (7) holds asymptotically (not exactly) for the number $X(x)$ of primes not exceeding x. If the Hardy-Littlewood twin prime conjecture (6) holds, the same is true for the number $X_2(x)$ of twin primes not exceeding x.

4. Logarithmic Integral Approximations

Crandall and Pomerance (2005, pp. 10–11) "note that one useful, albeit heuristic, interpretation of [the logarithmic integral (4) in the prime number theorem (5)] is that for random large integers x the 'probability' that x is prime is about $1/\ln x$." This perspective suggests, and rigorous proof provided below will confirm, an alternative asymptotic expression for the mean and the variance of the primes (and twin primes) not exceeding x. Define

$$\mu(x) \equiv \int_2^x \frac{t \cdot dt}{\log(t)} \Big/ \int_2^x \frac{dt}{\log(t)}, \tag{8}$$

$$\sigma^2(x) \equiv \int_2^x \frac{t^2 \cdot dt}{\log(t)} \Big/ \int_2^x \frac{dt}{\log(t)} - (\mu(x))^2. \tag{9}$$

Similarly, for the twin primes, define

$$\mu_2(x) \equiv \int_2^x \frac{t \cdot dt}{(\log t)^2} \Big/ \int_2^x \frac{dt}{(\log t)^2}, \tag{10}$$

$$\sigma_2^2(x) \equiv \int_2^x \frac{t^2 \cdot dt}{(\log t)^2} \Big/ \int_2^x \frac{dt}{(\log t)^2} - (\mu_2(x))^2. \tag{11}$$

Corollary 2. As $x \to \infty$,

$$M(x) \sim \mu(x) \sim \frac{\pi(x^2)}{\pi(x)} \sim \frac{x}{2}, \tag{12}$$

$$V(x) \sim \sigma^2(x). \tag{13}$$

Moreover, if the Hardy-Littlewood twin prime conjecture (6) holds, then

$$M_2(x) \sim \mu_2(x) \sim \frac{C_2(\pi(x))^2}{\pi_2(x)} \sim \frac{x}{2}, \tag{14}$$

$$V_2(x) \sim \sigma_2^2(x), \tag{15}$$

$$\frac{C_2(\pi(x))^3}{\pi_2(x)\,\pi(x^2)} \sim 1. \tag{16}$$

Proof. As $M(x)$ and $\mu(x)$ are both asymptotically $x/2$, they are asymptotic to one another, and similarly for the other moments. From the definition (8) and Lemma A.1,

$$\mu(x) \sim \frac{\mathrm{li}(x^2)}{\mathrm{li}(x)} \sim \frac{\frac{x^2}{2\log x}}{\frac{x}{\log x}} = \frac{x}{2} \sim M(x).$$

The prime number theorem also gives $\frac{\mathrm{li}(x^2)}{\mathrm{li}(x)} \sim \frac{\pi(x^2)}{\pi(x)}$. This proves (12).

Similarly, from the definition (9) and Lemma A.1,

$$\sigma^2(x) \sim \frac{\mathrm{li}(x^3)}{\mathrm{li}(x)} - \left(\frac{\mathrm{li}(x^2)}{\mathrm{li}(x)}\right)^2 \sim \frac{\frac{x^3}{3\log x}\cdot}{\left(\frac{x}{\log x}\right)} - \left(\frac{x}{2}\right)^2$$

$$= \frac{x^2}{12} \sim V(x).$$

This proves (13). The proofs of (14) and (15) follow the same procedure. In the proof of (14), after proving that $M_2(x) \sim \mu_2(x)$ because both are asymptotic to $x/2$, one notes that

$$\frac{x}{2} = \frac{\frac{x^2}{2(\log x)^2}}{\frac{x}{(\log x)^2}} \sim \frac{(\mathrm{li}(x))^2}{2\mathrm{li}_2(x)} \sim \frac{C_2(\pi(x))^2}{\pi_2(x)}.$$

Finally, (16) follows from dividing (14) by (12). □

5. Numerical Comparisons of Asymptotic Formulas with Exact Counts

For 31 values of x selected from 10 through 2×10^8, I computed (starting from the "primes" function of Matlab, Release 2015a) the number, mean, variance, and cv of the primes (twin primes) not exceeding x, the asymptotic approximations to these quantities based on the ratios of logarithmic integrals, and $x/2$ and $x^2/12$. The 31 values of x examined were integral rounded approximations to increments by one-quarter on a $\log_{10}$ scale from 10 to 10^8, plus the geometric mean between 10^8 and

Table 1. For selected x, values of the number, mean, variance, and cv of primes not exceeding x, the asymptotic number, mean, variance, and cv based on the logarithmic integral in the prime number theorem, and the limiting expressions given by the right sides of (3). The limiting cv, 0.57735, is the same for all x. The notation E+nn means multiply by 10^{+nn}.

Column 1 x	2 $\pi(x)$	3 $M(x)$	4 $V(x)$	5 cv(x)	6 li(x)	7 $\mu(x)$	8 $\sigma^2(x)$	9 $\sigma(x)/\mu(x)$	10 $x/2$	11 $x^2/12$
1000	168	453.1369	88389.44	0.656101	176.5645	445.3023	89698.21	0.672569	500	83333.33
10,000	1229	4667.531	8764508	0.634273	1245.092	4627.936	8787164	0.640526	5000	8333333
100,000	9592	47372.45	8.65E+08	0.620836	9628.764	47260.02	8.66E+08	0.622685	50,000	8.33E+08
1,000,000	78,498	478361.3	8.57E+10	0.612129	78626.5	478311.4	8.59E+10	0.612646	500,000	8.33E+10
10,000,000	664,579	4,820,082	8.54E+12	0.606185	664917.4	4,820,061	8.54E+12	0.606330	5,000,000	8.33E+12
10^8	5,761,455	48,461,680	8.51E+14	0.601936	5,762,208	48,460,300	8.51E+14	0.601974	50,000,000	8.33E+14
2×10^8	11,078,937	97,049,672	3.4E+15	0.600883	11,079,974	97,048,315	3.4E+15	0.600913	1E+08	3.33E+15

Table 2. For selected x, values of the number, mean, variance, and cv of twin primes not exceeding x, the asymptotic number, mean, variance, and cv based on the logarithmic integral in the twin primes conjecture, and the limiting expressions given by the right sides of (3). The limiting cv, 0.57735, is the same for all x.

Column 1 x	2 $\pi_2(x)$	3 $M_2(x)$	4 $V_2(x)$	5 cv$_2(x)$	6 li$_2(x)$	7 $\mu_2(x)$	8 $\sigma^2{}_2(x)$	9 $\sigma_2(x)/\mu_2(x)$	10 $x/2$	11 $x^2/12$
1000	35	345.2286	78968.63	0.813993	34.68506	360.1005	94714.55	0.854643	500	83333.33
10,000	205	4209.761	8,686,191	0.700095	162.2412	4111.51	9,292,462	0.741419	5000	8,333,333
100,000	1224	44571.94	9.16E+08	0.679003	945.7596	43903.62	9.01E+08	0.683579	50,000	8.33E+08
1,000,000	8169	454422.2	8.84E+10	0.654203	6246.976	453577.4	8.84E+10	0.655510	500,000	8.33E+10
10,000,000	58,980	4,625,962	8.74E+12	0.639191	44499.56	4,622,007	8.74E+12	0.639766	5,000,000	8.33E+12
10^8	440,312	46,791,563	8.69E+14	0.630139	333530.2	46,798,118	8.68E+14	0.629587	50,000,000	8.33E+14
2×10^8	813,371	93,846,098	3.47E+15	0.627457	616347.8	93,875,262	3.47E+15	0.627171	10^8	3.33E+15

2×10^8, namely, 10, 18, 32, 56, 100, 178, 316, 562, 1000, 1778, 3162, 5623, 10,000, 17,783, 31,623, 56,234, 100,000, 177,828, 316,228, 562,341, 1,000,000, 1,778,279, 3,162,278, 5,623,413, 10,000,000, 17,782,794, 31,622,777, 56,234,133, 100,000,000, 141,421,356, 200,000,000. For the sake of brevity, only a subset of these values of x and associated results are presented in Table 1 for primes and Table 2 for twin primes. All the selected values of x are represented in Figures 1–3.

For all values of x examined here, $M(x) < x/2$, that is, the mean of the primes not exceeding x is less than the asymptotic mean, which is half of x. The following numerical observations characterize all but the first five of the 31 values of x examined, that is, $x \geq 178$, but all these inequalities fail to hold for some smaller values of x. The mean $M(x)$ of the primes not exceeding x satisfies $\mu(x) < M(x)$ and $|M(x) - \mu(x)| < |x/2 - M(x)|$, that is, the ratio $\mu(x)$ of logarithmic integrals approximates $M(x)$ more closely than the mean of the asymptotic uniform distribution approximates $M(x)$. Moreover, $x/2 - M(x)$ is monotonically increasing in x (for this limited selection and range of x). As for the variance $V(x)$ of the primes, for the selected values of $x \geq 1000$, but not for some smaller values, $x^2/12 < V(x) < \sigma^2(x)$ and $|\sigma^2(x) - V(x)| < |V(x) - x^2/12|$,

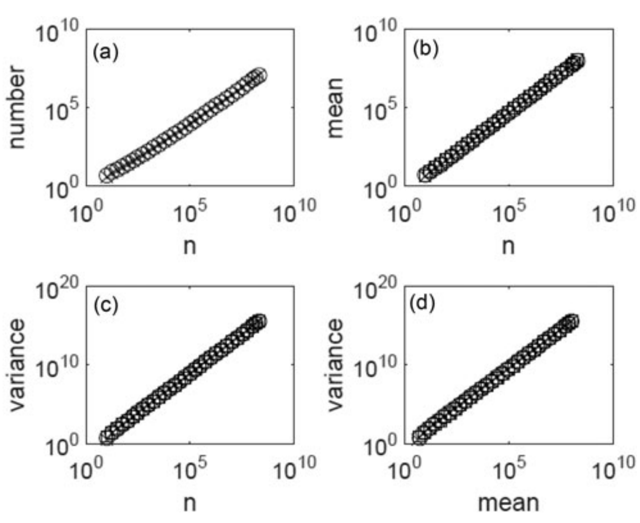

Figure 1. As functions of selected values of x, the (a) number, (b) mean, and (c) variance of primes not exceeding x, according to exact enumeration (marker ×), the formulas based on the logarithmic integral (marker ◯), and (for (b) and (c)) the asymptotic expressions given by (3) (marker □). Panel (d) tests the power law (1) by plotting $\log_{10}$(*variance of primes*) as a function of $\log_{10}$(*mean of primes*), using the same markers. In (d), the markers □ are perfectly linear with slope 2 on log-log coordinates. The superposition of all three markers in (b), (c), and (d) confirms the accuracy of the asymptotic expressions.

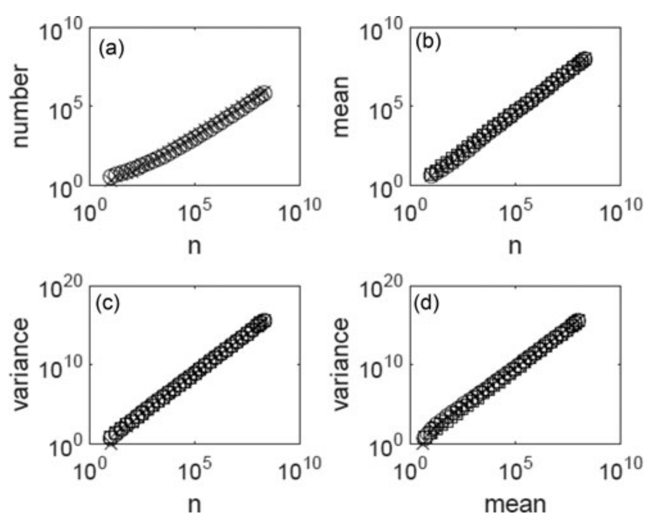

Figure 2. As functions of selected values of x, the (a) number, (b) mean, and (c) variance of twin primes not exceeding x, according to exact enumeration (marker ×), the formulas based on the logarithmic integral (marker ◯), and (for (b) and (c)) the asymptotic expressions given by (3) (marker □). Panel (d) tests Taylor's law (1) by plotting $\log_{10}$(*variance of twin primes*) as a function of $\log_{10}$(*mean of twin primes*), using the same markers. In (d), the markers □ are perfectly linear with slope 2 on log-log coordinates. The superposition of all three markers in (b), (c), and (d) confirms the accuracy of the asymptotic expressions.

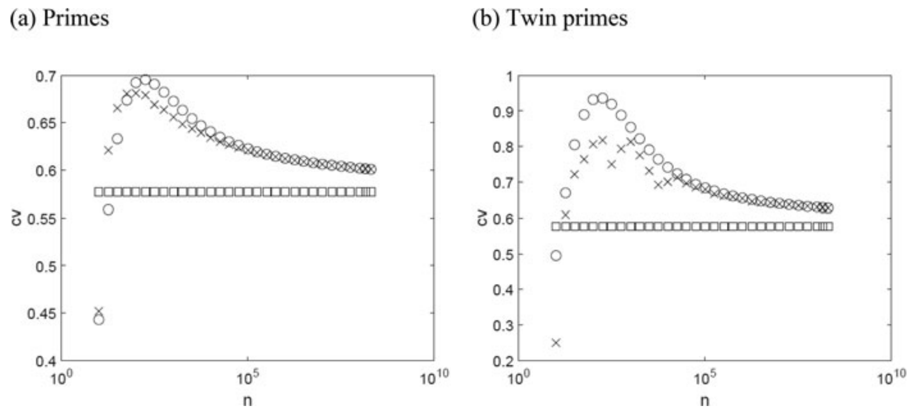

Figure 3. As functions of selected values of x, the coefficient of variation (cv) of the (a) primes and (b) twin primes not exceeding x, according to exact enumeration (marker $\times$), the formulas based on the logarithmic integral (marker ◯), and the limiting value 0.5774 on the right side of (2) (marker □).

that is, the ratio of logarithmic integrals approximates the actual variance $V(x)$ more closely than the variance of the asymptotic uniform distribution approximates $V(x)$. It is tempting to conjecture that there exists an integer x_0 such that for all $x \geq x_0$, the preceding inequalities hold.

Figure 1 plots the number, mean, and variance of the primes not exceeding selected values of x, according to exact enumeration, the formulas based on the logarithmic integral, and the asymptotic expressions given by the right sides of (3). Figure 1(d) tests Taylor's power law (1) by plotting $\log_{10}(\textit{variance of primes})$ as a function of $\log_{10}(\textit{mean of primes})$.

In Table 2 for the twin primes, the only consistent inequality (for this limited selection and range of x) is that, for $x \geq 32$, $M_2(x) < x/2$, and for $x \geq 10^3$, $x/2 - M_2(x)$ increases monotonically with x. Figure 2 for twin primes is analogous to Figure 1 for primes, with similar results.

Figure 3 plots the coefficient of variation (cv) of the (a) primes and (b) twin primes not exceeding the 31 selected values of x, according to exact enumeration, the formulas based on the logarithmic integral, and the asymptotic expressions given by the right side of (2). Evidently, the convergence of the cv to the limiting value 0.5774 is slow, but for $x = 2 \times 10^8$ the enumeration ($\times$) agrees with the logarithmic integrals ($\circ$) to better than one part in 10^4 for primes and agrees to better than one part in 10^3 for twin primes.

Comparing the 31 values for primes and twin primes shows that, for all selected $x \geq 10$, $\mu(x) > \mu_2(x)$ and $\sigma(x)/\mu(x) < \sigma_2(x)/\mu_2(x)$. For each selected $x \geq 32$ (but not for $x = 18$), $M(x) > M_2(x)$, $cv(x) < cv_2(x)$. For all selected $x \geq 316$, $\sigma^2(x) < \sigma_2^2(x)$. For all selected $x \geq 17783$, $V(x) < V_2(x)$. It is tempting to conjecture that all these inequalities hold for all x sufficiently large. If these conjectures hold, it is somewhat surprising that Taylor's law (1) holds with identical parameters for primes and twin primes.

Table 3 gives, for selected x, the numbers of primes and twin primes not exceeding x, three asymptotic estimates of their mean, and the ratio of two such estimates. In Table 3, the numbers of primes not exceeding x are from "How Many Primes Are There?", by Chris K. Caldwell, *https://primes.utm.edu/howmany.html#table*, accessed 2015-10-11, and the numbers of twin primes not exceeding x are from Sebah and Gourdon (2002). All other quantities in Table 3 are my computations using Matlab (Release 2015a). Other counts of primes and twin primes are available at "Tables of values of pi(x) and of pi2(x)", by Tomás Oliveira e Silva, *http://sweet.ua.pt/tos/primes.html*, accessed 2016-02-16.

Table 3. For selected x, values of the number of primes and number of twin primes not exceeding x, their asymptotic mean $x/2$, and the asymptotic estimates of their mean $\pi(x^2)/\pi(x)$ from (12) and $C_2(\pi(x))^2/\pi_2(x)$ from (14). The last column is asymptotic to 1 from (16).

x	$\pi(x)$	$\pi_2(x)$	$x/2$	$\pi(x^2)/\pi(x)$	$C_2(\pi(x))^2/\pi_2(x)$	$C_2(\pi(x))^3/(\pi_2(x)\pi(x^2))$
1.E+01	4	2	5.E+00	6	5	0.845
1.E+02	25	8	5.E+01	49	52	1.049
1.E+03	168	35	5.E+02	467	532	1.139
1.E+04	1,229	205	5.E+03	4,688	4,864	1.038
1.E+05	9,592	1,224	5.E+04	47,441	49,623	1.046
1.E+06	78,498	8,169	5.E+05	479,094	497,965	1.039
1.E+07	664,579	58,980	5.E+06	4,822,514	4,943,549	1.025
1.E+08	5,761,455	440,312	5.E+07	48,466,636	49,768,463	1.027
1.E+09	50,847,534	3,424,506	5.E+08		498,416,327	
1.E+10	455,052,511	27,412,679	5.E+09		4,986,799,998	
1.E+11	4,118,054,813	224,376,048	5.E+10		49,895,129,297	
1.E+12	37,607,912,018	1,870,585,220	5.E+11		499,150,311,715	
1.E+13	346,065,536,839	15,834,664,872	5.E+12		4,992,961,628,438	
1.E+14	3,204,941,750,802	135,780,321,665	5.E+13		49,940,610,730,906	
1.E+15	29,844,570,422,669	1,177,209,242,304	5.E+14		499,490,694,800,737	
1.E+16	279,238,341,033,925	10,304,195,697,298	5.E+15		4,995,585,553,449,950	

6. Conclusions and Open Questions

By viewing a classic area of number theory in the light of an ecological pattern, Taylor's law, we have been led to apparently new facts about primes and twin primes. This stimulation of mathematics by biology is not a fluke, but a productive strategy (Cohen 2004). This discovery demonstrates, once again, that the bread-and-butter tools of probability and statistics (e.g., random variables, the cumulative distribution function, expectation, variance) matter in seemingly unrelated areas of pure mathematics. It also gives one more confirmation that Taylor's law can show up where it is least expected.

Many further mathematical questions arise from these elementary results, and the tools of probability and statistics may prove useful in addressing these questions as well. How fast does the mean of primes (twin primes) not exceeding x converge to $x/2$? How fast does the variance of primes (twin primes) not exceeding x converge to $x^2/12$? How fast do the mean and variance of primes and twin primes converge to the asymptotic formulas based on ratios of logarithmic integrals for the mean and variance of primes and twin primes? Answers to these questions may follow from a more precise version of the prime number theorem, which bounds the error of the logarithmic integral approximation (5): for some $C > 0$, for all $x \in \mathbb{R}_2$, $\pi(x) = \mathrm{li}(x) + O(f(x))$, where $f(x) = x\exp(-C(\log x)^{3/5}(\log\log x)^{-1/5})$ (Montgomery and Vaughan 2007, p. 194, their eq. (6.28)).

Of the inequalities observed numerically in the comparisons between exact counts and asymptotic formulas in the previous section, which are true for all sufficiently large x?

Can any or all of (1), (3), (2) and (7) for primes and their analogs for twin primes be derived from Hawkins's random sieve or other random models of the primes (Lorch and Ökten 2007; Bui and Keating 2010)?

Which of these results can be extended to prime pairs with a gap $p_{n+1} - p_n$ other than 2 or to other sequences of primes (Lorch and Ökten 2007; Bui and Keating 2010)?

In addition to these new questions for mathematics is a question for further biological research. Do any processes involved in generating the primes (and twin primes) have biological counterparts where TL has been observed empirically?

Appendix

Lemma A.1. For $m \in \{0, 1, 2, \ldots\}$, $n \in \{1, 2, 3, \ldots\}$, $x > 2$,

$$\int_2^x \frac{t^m \cdot dt}{(\log(t))^n} \sim \frac{x^{m+1}}{(m+1)(\log x)^n}.$$

In particular,

$$\int_2^x \frac{t^m \cdot dt}{\log(t)} \sim \frac{x^{m+1}}{(m+1)\log x} \sim \mathrm{li}(x^{m+1}),$$

$$\int_2^x \frac{t \cdot dt}{(\log t)^2} \sim \frac{(\mathrm{li}(x))^2}{2},$$

$$\int_2^x \frac{t^2 \cdot dt}{(\log t)^2} \sim \frac{x(\mathrm{li}(x))^2}{3}.$$

Proof. Integration by parts gives

$$\int_2^x \frac{t^m \cdot dt}{(\log(t))^n} = \frac{x^{m+1}}{(m+1)(\log x)^n} + \frac{n}{m+1}\int_2^x \frac{t^{m+1} \cdot dt}{t(\log(t))^{n+1}} - C,$$

where $C = 2^{m+1}/[(m+1)(\log 2)^n]$. As $x \to \infty$, the ratio of the second term on the right to the first term on the right approaches 0. □

Acknowledgment

The author thanks Lek-Heng Lim, Carl Pomerance, the editor Nicole Lazar, and the associate editor for helpful exchanges and suggestions, and Priscilla K. Rogerson for help.

Funding

This work was partially supported by U.S. National Science Foundation grant DMS-1225529.

References

Bliss, C. I. (1941), "Statistical Problems in Estimating Populations of Japanese Beetle Larvae," *Journal of Economic Entomology*, 34, 221–232. [399]

Bui, H. M., and Keating, J. P. (2010), "On Twin Primes Associated with the Hawkins Random Sieve," available at *http://arxiv.org/abs/math/0607196v3* [404]

Cohen, J. E. (2004), "Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better," *Public Library of Science Biology*, 12, 2017–2023. [404]

Crandall, R., and Pomerance, C. (2005), *Prime Numbers: A Computational Perspective* (2nd ed.), New York: Springer Science+Business Media. [400,401]

Eisler, Z., Bartos, I., and Kertész, J. (2008), "Fluctuation Scaling in Complex Systems: Taylor's Law and Beyond," *Advances in Physics*, 57, 89–142. [399]

Fracker, S. B., and Brischle, H. A. (1944), "Measuring the Local Distribution of *Ribes*," *Ecology*, 25, 283–303. [399]

Giometto, A., Formentin, M., Rinaldo, A., Cohen, J. E., and Maritan, A. (2015), "Sample and Population Exponents of Generalized Taylor's Law," *Proceedings of the National Academy of Sciences*, 112, 7755–7760. [400]

Hayman, B. I., and Lowe, A. D. (1961), "The Transformation of Counts of the Cabbage Aphid (*Brevicoryne brassicae* (L.))," *New Zealand Journal of Science*, 4, 271–278. [399]

Kendal, W. S., and Jørgensen, B. (2011), "Taylor's Power Law and Fluctuation Scaling Explained by a Central-Limit-Like Convergence," *Physical Review E*, 83, 066115, 1–7. [399]

Lorch, J., and Ökten, G. (2007), "Primes and Probability: The Hawkins Random Sieve," *Mathematics Magazine*, 80, 116–123. [404]

Montgomery, H. L., and Vaughan, R. C. (2007), *Multiplicative Number Theory I. Classical Theory*, Cambridge, UK: Cambridge University Press. [400,404]

Sebah, P., and Gourdon, X. (2002), "Introduction to Twin Primes and Brun's Constant Computation," available at *http://numbers.computation.free.fr/Constants/Primes/twin.html* [400,403]

Taylor, L. R. (1961), "Aggregation, Variance and the Mean," *Nature*, 189, 732–735. [399]

Taylor, L. R. (1986), "Synoptic Dynamics, Migration and the Rothamsted Insect Survey: Presidential Address to the British Ecological Society, December 1984," *Journal of Animal Ecology*, 55, 1–38. [399]

Xiao, X., Locey, K. J., and White, E. P. (2015), "A Process-Independent Explanation for the General Form of Taylor's Law," *The American Naturalist*, 186, E51–E60. [399]